Yuxuan (Wayne) Wang

https://waynewangyuxuan.github.io/personal_site/ || waw009@ucsd.edu

EDUCATION SUMMARY:

University of California, San Diego, Jacobs School of Engineering, San Diego, CA — September 2025 - June 2027
Master of Science, Computer Science

New York University, Tandon School of Engineering, Brooklyn, NY — September 2021 - May 2025
Bachelor of Science**,** Computer Science & Math

TECHNICAL SKILLS:

Python || Java || C/C++ || SQL || Go || Kafka || Redis || PostgreSQL || Hive || Spark || Flink || AWS (Lambda, S3, DynamoDB, Step Functions, EventBridge) || Docker || Kubernetes || Linux (Shell, IPC, Threads)

EXPERIENCES:

Software Engineering Intern - ByteDance — June 2025 – September 2025
Global Monetization Product and Technology - TikTok Ads Diagnosis
(Java, Python, SQL, Kafka, Hive, Spark, Big Data, TikTok Ads)

- Engineered and scaled backend data pipelines for TikTok Ads diagnostics, supporting multi-DC (SG, VA) deployments.
- Productionized entry points & plugins: exposed Copilot via AdDebug Web Chat and Lark Bot through an RPC→HTTP gateway; added a plugin framework to call HTTP/RPC/X-Table/Auto Diagnoser with prompt logging & monitoring.
- Built internal Copilot backend: Feishu-doc ETL → Elasticsearch hybrid recall (text + vector); integrated secure conversation storage, access tokens, and RPC→HTTP gateway for service unification.
- Designed and standardized Intent-vs-Actual Budget Delta metrics, re-architecting 2 core and 13 downstream pipelines (~450 TB backfill) for transparent, auditable financial reporting.
- Contributed to performance tuning and alerting with Redis caching, data deduplication, and anomaly analysis in periodic pipelines.

Research Assistant - New York University — June 2024 – May 2025
Evaluation of Graph-Based Vocabulary Mismatch Solution in Information Retrieval
Supervised by Professor Torsten Suel @ New York University, Tandon School of Engineering
(Information Retrieval, Search Engine, Query Processing, Graph Algorithms, Linux, HPC, Python)

- Investigated graph-based retrieval algorithms to improve large-scale search efficiency; designed a modular reranking pipeline combining sparse and dense models.
- Applied KNN/HNSW indexing and vector masking (PRFDIME) to improve recall and reduce compute cost by 40–50%.

Data Engineering Intern - CITIC Poly Fund (Guangzhou) — June 2023 - August 2023
(Python, AWS, NoSQL, Kafka, DynamoDB, NLP)

- Built an automated investment-intelligence system for chip/EDA sectors: Wind + web-crawled data → Kafka→Lambda→DynamoDB pipeline.
- Extracted key structured fields (investors, round, amount) and applied BERT-based sentiment + summarization for financial trend tracking.
- Deployed Streamlit dashboard visualizing sentiment, keyword clusters, and trends, replacing a manual spreadsheet workflow across the department.

PROJECTS:

Notate (Cross-App Capture & Automation Engine) — September 2025 - October 2025

- Built a lightweight macOS/mobile app that listens across apps for trigger-based input (e.g., !! buy milk !!), clears text automatically, and centralizes scattered notes into a single, encrypted database.
- Integrated Claude Sonnet 4.5 to parse and classify snippets — automatically converting notes into structured actions such as contacts, calendar events, reminders, or map locations.
- Designed the workflow for frictionless task execution, where AI suggestions (e.g., “add to calendar,” “search for shopping guide”) can be confirmed or triggered directly within the app, turning passive notes into active outcomes.

ShadowDash (C++ Build System Optimization)

- Built a C++ DSL to construct build DAGs directly in memory, bypassing Ninja’s text parsing.
- Validated on zlib and LLVM, achieving 1.7×–2× faster compile times via optimized dependency graph execution and macro-based manifests.

TripPlanner (AWS Serverless Routing Platform) — March 2025 - May 2025

- Designed route-optimization and notification backend using AWS Lambda, Step Functions, and DynamoDB; integrated Amazon Location Service and OSRM for dynamic routing.
- Implemented Redis caching and alert aggregation for low-latency, cost-efficient serverless workflows; delivered API Gateway interface for real-time updates.