

Yuyang Wang

Santa Clara, CA | yw8396@gmail.com | 914-830-8396 | yuyangwang.org | github.com/echo-cool | linkedin.com/in/yuyangwang-yw

SUMMARY

Software engineer with 2+ years building production **LLM agent platforms** and large-scale data systems. Shipped an **AI-powered data analysis platform** end-to-end, sped up billion-record analysis **5x**, and own monitoring in two countries.

EDUCATION

Cornell University <i>M.Eng. in Computer Science Noel Croucher Graduate Fellowship</i>	Aug 2023 – May 2024 New York, NY
University College Dublin (dual degree with Beijing University of Technology) <i>B.Sc. in Software Engineering</i>	Sep 2019 – Jul 2023 Dublin, Ireland

EXPERIENCE

Polaris Wireless <i>Software Engineer</i>	Jun 2024 – Present Santa Clara, CA
- Shipped an AI-powered data analysis platform end-to-end (FastAPI, React, LangGraph, PostgreSQL): a streaming LLM agent runtime with disconnection-safe resumable runs, checkpointed state, MCP tools, self-hosted vLLM, JWT/RBAC access controls, human-in-the-loop approval, and an nsjail sandbox for model-generated code; agents run governed text-to-SQL across five database engines (ClickHouse, DuckDB, SQL Server, MySQL, PostgreSQL). - Built a 133-case end-to-end evaluation harness that grades each agent answer in four layers (tool routing, arguments, results, rendered charts/maps) against a ClickHouse ground-truth oracle; drove accuracy to 97% (129/133) and cut suite runtime 57% by fixing the tool-call and grounding defects it exposed. - Cut co-presence (meeting detection) analysis from 2 minutes to under 30 seconds (5x) across billions of mobility records by redesigning the ClickHouse pipeline; derived and documented its statistical confidence model. - Built and own production KPI monitoring for customer deployments in two countries (Flask, Grafana, Prometheus, Loki, Docker): 100+ KPIs across 48 dashboards, 15 alert rules, 15-minute analysis jobs, web command portal, monthly reports. - Developed workflow agents that turn natural-language requests into validated AIS API calls, resolving entity and geofence names, parsing dates, and checking parameters. Mentored several interns and delegated project tasks to colleagues in India while owning core engineering work.	
Cornell Tech / National Zoning Atlas <i>Graduate Research Assistant (Software Research Engineer)</i>	Aug 2023 – Dec 2023 New York, NY
Built an LLM pipeline for structured extraction and RAG over U.S. zoning-code documents (OpenAI API, AWS Textract OCR), with DVC-tracked data and automated evaluation, cutting manual review time for the Atlas team.	
University College Dublin <i>Teaching Assistant, School of Computer Science</i>	Sep 2020 – Mar 2023 Beijing–Dublin International College
Taught labs and tutorials, graded coursework, and held office hours for six undergraduate computer science modules.	
MapGIS <i>Software Engineer Intern</i>	Jun 2022 – Sep 2022 Beijing, China
Scaled a Django/MySQL service with read/write splitting and Redis caching; built its CI/CD pipeline and AWS Lambda jobs.	
Lenovo <i>Software Engineer Intern</i>	Jul 2021 Beijing, China
Built a real-time engine-sound diagnostics API (FastAPI, Kafka, KFServing); raised throughput from 10 to 150+ queries/s.	

SELECTED PROJECTS

- Compile N Run** (compilenrun.com, 2024 – present): Built an open-source programming tutorial platform covering **70+ topics**; reached **265K active users in 2026 YTD** and **300+ GitHub stars** across its English and Chinese editions.
- MiniTorch**: Implemented a PyTorch-like deep-learning framework with autograd, Numba/CUDA kernels, and conv layers.

SKILLS

LLM / Agents	Agent runtimes, tool use, MCP, vLLM, streaming (SSE), text-to-SQL, RAG, evaluation pipelines, PyTorch
Languages	Python, SQL, TypeScript/JavaScript, Java, C/C++ , Swift, Bash
Data	ClickHouse, DuckDB, PostgreSQL, MySQL, SQL Server, Redis, Hadoop/Spark, Kafka, geospatial analytics
Platform	Docker, Kubernetes, Grafana, Prometheus, Loki, Flask, FastAPI, React/Next.js, AWS, CI/CD, Linux

PUBLICATIONS

- Y. Wang, Y. Hao, A. Cong**. “Harnessing Machine Learning for Discerning AI-Generated Synthetic Images.” *arXiv* 2401.07358, 2024.
- Y. Wang**. “Heterogeneous Seismic Waves Pattern Recognition in Oil Exploration with Spectrum Imaging.” *IEEE ICCIA*, 2022.