JIAJUN (NICK) HUO

✉ jiajunhuo726@gmail.com · 📞 267-777-8518 · ⊙ github.com/nickhuo · 🔗 nickhuo.com

EDUCATION

University of Illinois Urbana Champaign   Aug 2024 - May 2026
Master of Science in Information Science (STEM)   *Urbana, IL*
- GPA 3.9/4.0 · Applied Machine Learning & Database Systems (both A) · Program #1 in Info Systems (U.S. News)

Shenzhen University   Sep 2019 - Jun 2023
Bachelor of Science in Mathematics   *Shenzhen*
- GPA: 3.79/4.50 (Top 10%) · Multiple scholarships · 4 ML modeling wins in supply chain, finance, and computer vision

WORK EXPERIENCE

Donut Labs   Jun 2025 - Aug 2025
Software Engineer Intern   *Agentic Browser Backed by Sequoia Capital*
- Engineered GPT-seeded ANN sentiment pipeline parsing 50 k tweets/day, slashing inference cost 19× to 0.003¢ per tweet
- Implemented prompt-tagged Pinecone vector DB, labeled 82 % tweets and cut GPT calls 80 % with equal accuracy
- Architected 18-endpoint RESTful FastAPI microservice for LLM inference & content ops: QPS ↑3× with P95 220 ms; auto-Swagger cut frontend integration 66 %, and Pydantic validation trimmed 4xx/5xx errors 40%
- Rolled out LangSmith observability across agent stack; engineered a closed-loop LLM evaluation harness with automated BLEU/BERTScore gates plus human adjudication, achieving 100 % PR coverage and surfacing regressions within 24 h.

Sonic SVM   Mar 2023 - Jul 2024
Software Engineer - Data Solutions   *Bitkraft-backed SaaS for 50+ Game Studios, $100M valuation*
- Led the development of data governance system, drove 3 C-level strategic decisions with quant insights, report to CPO
- Scaled data sources by 2x and architected ETL pipelines on GCP, ensuring 99.9% accuracy with optimized SQL queries
- Automated API log extraction workflow from PostgreSQL to BigQuery using Airflow, reducing 43.2% processing time
- Built a Pub/Sub → Dataflow streaming pipeline, surfacing in-game KPIs to BI dashboards with <5 s latency
- Launched Looker Studio dashboards consumed by 7 members; cut insight turnaround from 2 days → 30 min

PROJECTS

ReplicaGenie
- Architected 4-service LLM platform; modular microservices allow independent scaling, isolated failures & fast rollouts
- Automated LinkedIn/GitHub scraping via Selenium on AWS Lambda; parsed & stored structured data in MongoDB
- Built MongoDB → RabbitMQ CDC stream pushing changes to feature store in <1 s, enabling real-time personalization
- Delivered 4-bit LoRA fine-tuning on SageMaker, trimming GPU training cost 77 % and guaranteeing reproducibility
- Integrated Opik tracing and auto-eval, detecting hallucination & relevance and lifting LangChain-SageMaker reliability
- Orchestrated 6-service Docker stack-MongoDB replica, RabbitMQ, Qdrant—delivering production-ready RAG

BiteMatch
- Led a 4-member Agile squad, grooming backlog and delivering on 2-week sprints with 100% sprint goal hit rate
- Top 10% in course. Developed smart recipe app to personalize meal choices using React, Node.js, and MySQL
- Created RESTful API with 12+ endpoints handling user auth, recipe CRUD, video search, and review system
- Reduced YouTube API calls by caching videos in DB and using a quota-aware fetch logic to avoid exceeding limits
- Containerized services with Docker & GitHub Actions; push-to-prod cycle shrank from 20 min to <5 min

SKILLS

- **Programming**: Python, TypeScript, Go, JavaScript, SQL, Zsh & Bash, Node.js, React
- **Frameworks**: React, Next.js, Jenkins, Tailwind CSS, Vite, Node.js, Django, Express.js, MongoDB, Redis
- **DevOps** : AWS, Docker, Terraform, GitHub Actions, Vercel, Airflow, Linux/Unix, Git