

# Kevin Fu

[kevinfu@uchicago.edu](mailto:kevinfu@uchicago.edu) | [linkedin.com/in/kevinzefu](https://www.linkedin.com/in/kevinzefu) | [github.com/kzfubar](https://github.com/kzfubar) | [kzfubar.com](https://kzfubar.com)

## EDUCATION

---

**The University of Chicago**, M.S. in Computer Science (AI Systems Specialization) December 2026 (Expected)  
**Washington University in St. Louis**, B.S. in Computer Science May 2020

## SKILLS

---

**AI Systems:** Agent orchestration (*LangGraph, LangChain*), RAG (*FAISS, Elasticsearch*), MCP tooling, OpenAI APIs  
**Data Science & ML:** Python (*PyTorch, scikit-learn, TensorFlow, NLTK*), SQL, NoSQL (*MongoDB, Redis*), Spark  
**Infrastructure:** Distributed Systems (*Kubernetes, Kafka*), Cloud Computing, Data Pipelines & Analytics, DevOps

## PROFESSIONAL EXPERIENCE

---

**The University of Chicago**, *Graduate Research Assistant* November 2025 - Present

- Led technical strategy for a DOE-funded research project, establishing engineering workflows (GitHub, Slack), advising cross-functional teams on software best practices, and designing **AI systems for an autonomous experimentation workflow** in physical sciences at the Pritzker School of Molecular Engineering.
- Developed data and control pipelines for **experiment execution** (Bluesky) and **real-time data access** (Tiled), enabling coordination between robot arm actuation and data collection instruments for thin-film synthesis.
- Designed an **LLM-based reasoning and optimization** layer that maintains structured experimental state, encodes domain constraints, and runs subsequent thin-film trials with updated synthesis parameters.

**Independent Study, Pre-Graduate Preparation** June 2023 – August 2025

- Co-authored a manuscript (under review) **evaluating AI-driven systems** for sales training, contributing analysis of system architecture and deployment tradeoffs in regulated environments (see Previous Research Experience).
- Built **Python-based AI tools** including a task adherence agent integrating Todoist and Google Calendar APIs for personalized coaching feedback, and a RAG-based ROI calculator grounded in the ROI Institute's methodology.

**IMC Financial Markets** August 2020 – March 2023

*Software Developer, Equity Options Trading Team*

- Improved routing logic and **system design** by collaborating with traders to implement higher fidelity decisions for an autonomous order routing system using Java, MySQL, and Kafka that increased realized performance by **10–20%**.
- Diagnosed **peak-load bottlenecks** by profiling JVM components and instrumenting runtime **observability** (Prometheus, OpenTelemetry), enabling targeted scaling mechanisms and up to **3× throughput gains**.
- Improved analytical query performance by **~30%** and expanded historical lookback from weeks to months at constant latency by **optimizing data layout** in a distributed datastore (HDFS), supporting **up to 100M records** daily.

*Software Developer Intern, DevOps Team* June 2019 – August 2019

- Created **standardized orchestration patterns** with Go and Helm in Kubernetes and drove a company-wide migration from legacy bare-metal and Docker deployments across teams in Sydney, Amsterdam, and the US.
- Executed phased migration to Kubernetes in coordination with multiple service owners, transitioning **~80% of production trading deployments** from legacy infrastructure while minimizing deployment failures.

## SELECTED PROJECT

---

**Episodic Memory for Persistent AI NPCs**, *MPCS Course Project* February 2026 – Present

- Developed a **multi-agent architecture** for persistent in-character NPC interaction in tabletop RPG sessions.
- Built an **episodic memory retrieval system** using Elasticsearch with hybrid keyword and vector search, ranking memories by emotional importance and recency and benchmarking against **static RAG and GraphRAG baselines**.
- Designed a **tool-routed agent pipeline** separating rules grounding using the Open5e SRD, game state management in SQLite, memory retrieval, and **two-phase reasoning** from mechanics to in-character dialog across four MCP servers.
- **Finetuned GPT 4.1 Nano and Mini** on 17K crowdsourced fantasy dialog exchanges from the Meta AI fantasy corpus (LIGHT dataset) to improve and evaluate NPC dialog quality and stylistic consistency.
- Evaluated the system across 90+ test queries and 34 structured recall questions measuring identity consistency, retrieval grounding, and dialog quality.

## PREVIOUS RESEARCH EXPERIENCE

---

**AI Sales Training: Balancing Performance and Risk in a Global Pharmaceutical Company**, *Co-author* 2025  
**Sentiment Analysis for Student Feedback in Computing Education**, *Research Assistant* 2020  
**Evaluating Big Data Applications on Heterogeneous Architectures**, *NSF REU Undergraduate Researcher* 2017