

Ray Chu

646-925-0205 | ray.chu29@stu-mail.hunter.cuny.edu | linkedin.com/in/RayChu83 | github.com/RayChu83 | raychu.info

EDUCATION

Hunter College

New York, NY

B.A. in Computer Science, Minor in Math

Aug. 2025 – Present

- 1 of 41 Computer Science Scholars selected in Class of 2029
- Relevant Coursework: Data Structures and Algorithms, Software Analysis and Design, Discrete Structures, Intro to Computer Science

EXPERIENCE

Software Engineer Intern

Mar. 2026 – Present

Blitz (Backed by Antler S26)

New York, NY

- Shipped a unified role-based access control which authorizing all authenticated routes server-side against 3 platform account types (Admin, Business, or Workforce) plus per-resource access checks for payments.
- Reduced failing CI/CD tests by 15% by developing 12 custom AI agent skills that automatically route agent workflows to the right documentation, eliminating messy repository searches and enforcing clean coding standards.
- Led the front-end user interface development during a Next.js 16 migration which reduced bounce rates by 16%.

Software Engineer Intern

Jul. 2024 – Jan. 2025

Unlevered

New York, NY

- Decreased AI token consumption by 23% and slashed response times from 30s to 200ms by implementing a prompt caching system for previously fetched financial summary queries.
- Accelerated AI-summary processing speeds by 60% by decoupling a single prompt workload into 7 parallel chunked LLM calls.
- Improved fetch response time by 42% by implementing server-side pagination for filing search results which shrunk payload sizes by 85% and limited results to 10 per page.
- Optimized the front-end load performance by 15% by implementing UI streaming for chunked LLM responses which delivered real-time summary updates as chunks arrived.
- Decreased server query traffic by 72% by developing a client-side debouncing system that delays search requests until user input pauses.
- Eliminated 700ms of fetch latency by implementing client-side caching for already fetched financial data which reduced redundant API calls by 37% across page navigation.

PROJECTS

Syllabus to Calendar | *Next.js, TypeScript, Supabase, LangChain*

- Developed a full-stack web application that automates manual calendar entry for students by using an LLM (OpenAI/LangChain) to parse uploaded PDF syllabi and auto generate calendar events that sync to students Google Calendar.
- Reduced AI API costs and cut response time by 90%+ for repeat uploads by hashing syllabus content and caching parsed results in Postgres, serving duplicate syllabi instantly instead of re-calling the LLM.
- Cut AI parsing costs by 20–30% by preprocessing raw PDF-extracted text (stripping repeated headers/footers/whitespace) before sending it to the LLM, directly reducing input tokens on every uncached request

TECHNICAL SKILLS

Languages: JavaScript, TypeScript, Python, C++, HTML, CSS

Frameworks: React.js, Next.js, Redux, MobX, TailwindCSS, Material UI, Framer Motion, GSAP

Developer Tools: MongoDB, DynamoDB, Firebase, Supabase, RESTful APIs

Libraries: AWS, Azure, Docker, Doppler, GitHub, Figma, Adobe Photoshop