

Wenxuan (Vivian) Cao

vivian.cao@emory.edu | 605-5951-942 | www.linkedin.com/in/wenxuan-cao123

EDUCATION

Emory University

Bachelor of Science in Computer Science

GPA: 3.94/4.00

Atlanta, GA

expected December 2027

SKILLS

Languages: Java, Python, C, C++, JavaScript

Technical Skills: SQL, AWS (EC2, S3, DynamoDB, Lambda), Git, React, Node.js, PostgreSQL, MongoDB, Docker, RESTful APIs

Courses Taken: Machine learning, Systems Programming, NLP, Database Systems, Machine Level Programming, Data Mining, DS&A

PROFESSIONAL EXPERIENCE

HKBA (Hong Kong Blockchain Association)

Hong Kong

Programming Intern

May 2026 – Present

- Wrote 40+ Hardhat unit tests achieving 78% coverage on core Ethereum smart contract functions (mint, transfer, pause, access control); performed manual security review and Slither static analysis, remediating 3 flagged vulnerabilities pre-deployment
- Implemented Ethereum smart contracts for token issuance, membership verification, and escrow logic supporting HKBA's Web3 membership platform
- Optimized frontend performance via lazy loading, code splitting, and Cloudflare CDN integration; reduced bundle size 54% (2.4MB → 1.1MB) and improved mobile Lighthouse Performance score from 62 to 84
- Took initiative cut time to first byte from ~900ms to ~200ms by configuring CDN caching and static asset delivery
- Deployed full-stack application on AWS (EC2, RDS, S3) with Docker containerization and IAM-based access control; manually load-tested at 50 concurrent connections via Postman with no timeouts

Prompt Regenerator for the "Car Wash" Reasoning Failure in LLM

Atlanta, GA

Research & Chrome Extension Developer

Jan 2026 – June 2026

- Designed a rule-based diagnostic system (regex, keyword/lexicon, and tokenization-based analysis) that classifies prompts into 6 failure categories grounded in computational linguistics coursework, deployed live across 4 LLM providers (ChatGPT, Claude, Qwen, Llama)
- Architected a normalized analyzer contract decoupling diagnosis logic from the UI, enabling the system to run as a Manifest V3 Chrome extension with real-time DOM tracking (MutationObserver) and positioning the codebase for a planned upgrade to embedding-/LLM-based analysis without any UI rewrite
- Delivered an end-to-end tool from prompt diagnosis to rewrite to one-click apply, giving users actionable, explainable feedback on prompt quality before submission, reducing the trial-and-error cycle of prompt iteration

BridgeAn

Hong Kong

Backend & AI Agent Engineering Intern

Sep 2025 – Dec 2025

- Refactored a 1,000+ line monolithic document-parsing module into two dedicated agent classes and rapidly ramped up on an unfamiliar AI infrastructure stack like local LLM serving (Ollama), vector databases (ChromaDB), and embedding/reranking models (BGE-M3, BGE-reranker-v2-m3) to ship a working retrieval pipeline against a fast-moving internal roadmap
- Solved the "chunking breaks tables" problem common to naive Excel-to-LLM pipelines by generating LLM-based table descriptions for embedding while preserving full table structure in metadata, then built two-stage retrieval (dense recall + BGE rerank) to improve financial-document lookup accuracy from 72% to 91% on a 200-query evaluation set, a 26% relative improvement over the baseline
- Built a two-stage LLM drafting agent (DeepSeek V3 for initial generation, DeepSeek R1 for refinement) that auto-generates sections of an M&A CIM from a defined data schema, deployed and tested in a local Unix/Linux development environment
- Diagnosed systemic failure modes in generated output (cross-section content duplication, section misclassification) through hands-on debugging, directly informing the next architecture iteration (a planned LangGraph-based agent split)

AiFinSphere

Remote

Data Engineer Intern

Jan 2025 - March 2025

- Analyzed 50+ reports and market research papers, building automated data ingestion scripts (Python + SQL) to extract and structure 20+ critical data points on industry trends and business performance.
- Processed 1.2M+ user reviews and transaction logs by designing ETL pipelines (Pandas / PySpark), enabling large-scale sentiment analysis and pricing model evaluation.
- Created a 15-slide analytics report by warehouse queries and visualization tools; Researched and codified fin-tech compliance guidelines (CFTC, FINRA, SEC, 200+ pages), designing a structured data model to map regulatory requirements to product features for downstream analytics
- Integrated pipelines with AWS services (S3, Lambda, Redshift) for data storage, transformation, and analytics, improving scalability and reducing manual processing time by 30%

LEADERSHIP EXPERIENCE AND COMMUNITY INVOLVEMENT

AI Data Lab – Center for AI Learning

Atlanta, GA

Researcher & Project Analyst Leader

Jan 2025 – May 2025

- Analyzed 10M+ donation records from Feeding America using ARIMA time series models to identify long-term trends
- Organized team members and cleaned 500K+ records and built 8+ data visualizations to communicate insights in an academic research poster and presented findings to 200+ students and sponsors, and maintained a GitHub repository (1.5K+ lines of documented code) for open-source collaboration