George Gu

georgu@umich.edu | (989) 948-8968 | www.linkedin.com/in/george-gu-146bb0251 | github.com/georgu28

Education

University of Michigan May 2028
Bachelor of Science in Engineering, Computer Science *Ann Arbor, MI*

- **GPA**: 3.81/4.00
- **Coursework**: Operating Systems, Web Systems, Computer Security, Machine Learning, Artificial Intelligence, Reinforcement Learning, Theory of Computation, Computer Organization

Experience

Capital One June – August 2026
Technology Intern *Richmond, VA*

- Architected data pipelines via scheduled AWS Lambda workers to ingest daily snapshots from internal asset-management systems, yielding clean tabular datasets to train predictive models on application lifecycle and business criticality.
- Engineered Python REST APIs orchestrating data-attestation and asset lifecycle compliance workflows, securing them with parameterized PostgreSQL queries and role-based access control across multi-level corporate domains and org hierarchies.
- Authored 20+ automated integration tests covering mock database transactions, cross-region failures, and network-timeout edge cases, lifting repository-wide coverage to 91% (100% on all new files) for the compliance service.

BoilerVault January 2026 – Present
Founding Engineer (Contract) *West Lafayette, IN*

- Architected a multi-tenant operations platform with FastAPI, PostgreSQL, and Next.js/TypeScript for a 3-campus student storage business, implementing JWT auth, role-based access control, and 20+ REST endpoints validated by 140+ tests.
- Engineered idempotent webhook pipelines across 4 Stripe accounts and 3 WordPress booking sites, reconciling out-of-order payment events to automate customer, order, and billing creation with 0 manual entry across $130K+ in gross revenue.
- Replaced a failing Zapier/spreadsheet workflow by migrating 2,000+ legacy records via three-tier matching (exact email, rapidfuzz fuzzy name, Stripe charge fallback) and deploying on Railway/Vercel as the business's production system of record.

Nexteer Automotive May – August 2025
Software Engineering Intern *Saginaw, MI*

- Built an LLM-powered IDE extension in Python and TypeScript that parsed 300+ internal engineering guidelines and automated compliance checks across C and H files, integrating Azure AI and Copilot APIs into the developer workflow.
- Engineered prompt pipelines and few-shot strategies to extract and interpret rules from unstructured documentation, tuning the system to 95% violation-detection accuracy and saving engineers an estimated 8 hours of review per week.
- Deployed the tool globally across 26 sites and integrated it with internal dev pipelines, cutting manual code-review time by 87.5% and enabling reliable, real-time automation screening for safety-critical embedded steering systems.

Villanova University June – September 2023
Data Engineer & First Author *Villanova, PA*

- Engineered end-to-end data pipelines in R, using PCA and clustering to process and visualize 28,000+ pathogen isolates across 10+ years, integrating bacterial species, antimicrobial agents, and AMR resistance genes to uncover essential trends.

Projects

Scalable Search Engine January – April 2026
Systems Developer *Ann Arbor, MI*

- Built a multi-stage MapReduce pipeline in Python to construct an inverted index over 3,000+ documents, computing TF-IDF scores and document-normalization factors across a parallel, multi-job processing architecture for scalable retrieval.
- Implemented a REST API index server in Flask serving 3 partitioned inverted index segments, integrating PageRank-weighted cosine similarity scoring to balance TF-IDF with link authority when ranking query results by relevance.
- Designed a service-oriented search frontend that dispatches concurrent REST requests across index segment servers using threads, then aggregates and re-ranks the results to return the top 10 hits with details from a SQLite database.

Resume Screener January – May 2025
Machine Learning Developer *Ann Arbor, MI*

- Built an NLP pipeline over a labeled dataset of 1,000+ resumes and job postings, applying tokenization, TF-IDF vectorization, and feature engineering to convert unstructured text into structured features for resume-to-posting matching.
- Designed scikit-learn and TensorFlow models to rank resumes against job postings, using sentence-transformer embeddings and RAG retrieval over job descriptions to surface a user's best-matching role, deployed via Streamlit to 100+ users.

Skills/Additional

Languages: Python, SQL, C/C++, Java, JavaScript, TypeScript, HTML, CSS
Frameworks/Developer Tools: Git, Docker, Kubernetes, Unix, Node.js, React.js, Flask, PyTorch, TensorFlow, PostgreSQL
Interests: Philosophy and Metaphysics, Game Design, Bouldering, Skateboarding, Golf, Guitar, Calisthenics

Publication

Gu G, Pei H, Zhou A, Fan B, Zhou H, Choi A, Huang Z, "A Comprehensive Study of Historical Detection Data for Pathogen Isolates from U. S. Cattle", 2023, *Antibiotics*, 12, 1509.
doi: https://doi.org/10.3390/antibiotics12101509