

Gavin Abrigo

(408)-888-8732 | abrigogavin05@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Computer Science undergraduate with hands-on experience developing scalable software solutions and automated workflows using Python, Go and JavaScript. Built ML-driven pipelines for knowledge graph retrofitting, retrieval-augmented generation and image classification, optimizing accuracy and efficiency. Skilled in data structures, algorithms, SQL and REST APIs, collaborating cross-functionally to deliver robust systems. Seeking a Software Engineering Intern role to contribute to complex, scalable projects.

EDUCATION

University of California Merced

Bachelor of Science, Computer Science & Engineering

May 2027

Merced, CA

TECHNICAL SKILLS

- **Programming Languages:** TypeScript, JavaScript, GoLang, Java, Python, C/C++, HTML/CSS, LaTeX, MIPS Assembly, SQL, MATLAB, SPARQL
- **Data & Databases:** PostgreSQL, SQLite, MongoDB, Excel
- **Principles & Tools:** Shell, React.js, FastAPI, Flask, REST APIs, Git, GitHub, Node.js, Next.js, Vite, StreamLit, Docker, CI/CD, JSON, System Architecture, Software Development, Data Structures, Algorithms, Infrastructure Development, Web Application Development, Unix/Linux Environments, Mobile Application Development, Distributed Systems, Parallel Systems
- **AI & ML Tools:** Nvidia NIM, Tesseract, OpenCV, PyTorch, Pandas, LLM API, Claude API, OpenAI API, Prompt Design, Large Language Models, Foundational Models, Retrieval Augmented Generation, Machine Learning, Information Retrieval, Natural Language Processing

EXPERIENCE

LLM Knowledge Graph Retrofitter (KGR) - Prof. Pandey

Sep 2025 - Present

Undergraduate Research Intern

Merced, CA

- Improve factual accuracy to reduce hallucinations in NLP models by building a Python-based Knowledge Graph Retrofitting (KGR) and Retrieval-Augmented Generation (RAG) pipeline using Wikidata and GraphRAG, leveraging machine learning techniques and best practices in software development.
- Fine-tune entity linking systems and evaluate LLM verification performance with NLP metrics such as exact match and F1 using scikit learn on benchmark datasets, applying data structures and algorithmic approaches to optimize throughput.
- Perform statistical modeling of knowledge graph data with R to identify entity relations and improve retrofitting efficiency.
- Experiment with deep learning architectures in PyTorch to optimize retrieval-augmented generation performance, analyzing results to choose the most effective models.
- Analyze information and evaluate pipeline outputs to select optimal strategies for knowledge graph retrofitting and verification.

BART - Software Engineering Capstone

Jan 2026 - May 2026

Software Engineering Intern

Merced, CA

- Develop scripts to automate classification of infrastructure incident images submitted via web and mobile app, leveraging JavaScript and Python automation frameworks along with structured metadata to route issues to appropriate maintenance teams.
- Developing image embedding and similarity detection workflows (Google Gemini) to de-duplicate submissions with postgres vector embeddings, extract contextual location data, and reduce manual review overhead.
- Collaborate with product management and engineering teams to refine feature requirements and choose optimal solutions in an agile environment.

PROJECTS

SMART BART

May 2026

Software Engineering Capstone

- Developed an image-based incident reporting system in Python using SentenceTransformers for deep learning-driven auto-classification of BART maintenance photos, clustering duplicates with cosine similarity to improve processing speed
- Designed a unified workflow as a Flask web app on a Unix/Linux environment that aggregates scattered photo reports into organized maintenance intelligence
- Used vector embeddings, SentenceTransformers, and cosine similarity to cluster duplicate incident photos, so BART staff see one grouped report

NYC Taxi Trip Duration Regression and Digit Classification

Oct 2025 - Dec 2025

NYC Taxi Trip Duration Regression and Digit Classification

- Engineered end-to-end scikit-learn pipelines in Python to predict NYC taxi trip durations and classify multi-digit MNIST images, demonstrating expertise in statistical modeling and machine learning
- Built and evaluated classification models for MNIST digit recognition and regression models for taxi trip duration using scikit-learn pipelines, applying feature engineering and algorithmic optimizations
- Implemented logistic regression with hyperparameter tuning and an XGBoost model with 5-fold stratified cross-validation on large-scale datasets, leveraging Pandas for data preprocessing and outlier removal