


# BENJAMIN HOLDERBEIN

✉ [benjamin.holderbein@gmail.com](mailto:benjamin.holderbein@gmail.com)  [linkedin.com/in/benjamin-holderbein](https://www.linkedin.com/in/benjamin-holderbein)  [github.com/BenjaminHolderbein](https://github.com/BenjaminHolderbein)

AI Engineer & Master's Student in AI and Data Science based in SF, building and deploying LLM solutions.

## Experience

---

### AI Engineer

San Francisco, CA

*Asurion, Intern*

October 2025 - Present

- Built modular RAG system to replace 3-year-old semantic search serving 10,000+ daily customer queries.
- Architected ingestion pipeline supporting 4 embedding models (BGE, Qwen3, Nemotron, Gemini) for evaluation.
- Led development of Redis vector storage with automated preprocessing, chunking, and embedding workflows.
- Integrated DeepEval framework to benchmark embedding models and retrieval strategies across evaluation metrics.

### Software Engineer

San Francisco, CA

*StudyStudio.ai, Intern*

January 2025 - May 2025

- Developed and deployed cross-platform mobile app using React Native, TypeScript, and Clerk for iOS and Android.
- Contributed to GitHub workflows including pull requests, issue tracking, and collaborative code reviews.
- Delivered a polished prototype that mirrored core web functionality and enabled future app store deployment.

### AI Engineer

San Francisco, CA

*Data Institute, Intern*

May 2024 - August 2024

- Designed and implemented a natural language processing algorithm to automate internship/student matching.
- Utilized large language models to cluster student qualifications and match them with employer-defined requirements.
- Identified performance bottlenecks and developed a preprocessing pipeline that reduced runtime and API costs by 40%.
- Defined success metrics, evaluated methodologies, and delivered a fully functional prototype optimized for performance.
- Solution actively used in University of San Francisco's intern placement process and scalable for broader application.

### Data Scientist

Davis, CA

*Frontier Energy, Inc., Intern*

May 2023 - August 2023 and May 2022 - October 2022

- Supported analysis of 40+ data channels monitoring building performance across residential pilot sites in California.
- Built interactive Azure Data Explorer dashboard to visualize real-time performance data for engineering teams.
- Engineered Python pipeline to clean, validate, and resample time-series data from 1-second to 5-minute intervals.
- Optimized data aggregation workflows across diverse datasets, improving processing efficiency for ongoing analysis.

## Projects

---

### Connected Vehicle Data Pipeline

- Built connected vehicle data pipeline generating synthetic transactions and loading into PostgreSQL database.
- Implemented ETL workflows, feature engineering, and fraud detection using scikit-learn, NumPy, and pandas.
- Developed interactive Streamlit dashboard for real-time fraud monitoring with dynamic filters and visual analytics.
- Dashboard: <https://connected-vehicle-data-pipeline.streamlit.app>
- Repository: <https://github.com/BenjaminHolderbein/connected-vehicle-data-pipeline>

### Developing a Large Language Model from Scratch

- Designed and built an LLM in Python using PyTorch following Sebastian Raschka's "Build an LLM."
- Repository: [https://github.com/BenjaminHolderbein/LLM\\_from\\_scratch](https://github.com/BenjaminHolderbein/LLM_from_scratch)

### Creating a Cellular Microscopy Counting Algorithm

- Developed automated cell counting solution using U-Net CNN for image segmentation, replacing manual counting.
- Achieved mean error of 1.4 cells, significantly outperforming the target accuracy of 3 cells for biomedical research.
- Report: [docs.google.com](https://docs.google.com). Code: [colab.research.google.com](https://colab.research.google.com).

## Skills

---

**Languages:** Python, SQL, Bash, R

**ML/AI Frameworks:** PyTorch, Hugging Face, LangChain, Scikit-learn, Pandas, NumPy, Streamlit, DeepEval, Gemini

**Tools & Technologies:** Git, GitHub, Docker, PostgreSQL, Redis, Azure, GCP, VS Code

**Leadership:** President of USF's 100+ member Rock Climbing Club for two years, managing operations and compliance.

## Education

---

### University of San Francisco

San Francisco, CA

*M.S. in Data Science and Artificial Intelligence*

July 2025 - June 2026

### University of San Francisco

San Francisco, CA

*B.S. in Data Science, 3.96 GPA*

August 2021 - May 2025