Connor Gag

La Jolla, CA | gagconnor@gmail.com | (507) 766-6008 | connor-gag-2317b6176 | connorgag

Summary

Data Scientist currently studying Computer Science with an emphasis on Artificial Intelligence. Passionate about helping people and doing work that matters. Skilled in machine learning, software engineering, and pipeline development in the healthcare industry. Bias for action with demonstrated success in creating and deploying ML models and collaborating with team members to achieve business objectives.

Skills

- Languages: Python (TensorFlow, PyTorch, Pandas, NumPy, Sci-kit learn, Anaconda, Jupyter, Selenium, Matplotlib, PySpark) • SQL • Java • REST APIs • Bash
- Technologies: Git • Google Cloud Platform • Azure • Snowflake • Docker • PowerBI • Unity • RAG • PostgreSQL • OracleDB • Jira

Education

MS **UC San Diego**, Computer Science, Artificial Intelligence specialization — Sept 2024 – Dec 2025

- GPA: 3.76/4.0
- Relevant coursework: NLP, Deep Neural Networks, Probabilistic Reasoning, Bayesian Networks, Algorithms

BS **Gustavus Adolphus College**, Computer Science, Minor in Math — Sept 2018 – May 2022

- GPA: 3.96/4.0
- Relevant Coursework: Algorithms, Theory of Computation, Database Systems, Compiler Design, Principles of Programming Languages, Machine Learning, Linear Algebra, Multivariable Calculus, Mathematics and Computer Science Proofs, Statistics

Experience

UnitedHealth Group, Data Scientist — Remote, May 2022 – June 2024

- Developed and deployed a machine learning model to predict housing instability among Medicaid members. Prioritized recall to identify at-risk individuals and prevent homelessness.
- Collaborated with cross-functional teams to clean and process claims and publicly available data for over 8 million Medicaid members.
- Designed and implemented a program to collect geospatial SDOH (social determinants of health) data for individuals within a specified radius, supporting data-driven insights.
- Parallelized geospatial data-collection processes, accelerating location data aggregation for integration into ML models and predictive analytics.
- Developed and validated a predictive model to identify Medicaid members at risk of transitioning to long-term care, enhancing resource allocation.
- Improved security by remediating over 200 security vulnerabilities in the team's GitHub repositories.
- Migrated 20 GitHub repositories to GitHub Enterprise Cloud.
- Built and automated data pipelines using Google Cloud Functions and BigQuery, enabling scalable and efficient data workflows.
- Spearheaded the team's transition to Agile workflows, achieving measurable gains

in project delivery speed and team collaboration.
- Designed a Retrieval-Augmented Generation (RAG) system enabling users to query company documents, providing quick, domain-agnostic access to insights via GPT.

UnitedHealth Group, Data Science Intern — Remote
June 2021 – Aug 2021

- Achieved 86% test coverage for a large-scale ML data pipeline using Pytest.
- Refactored codebase to enhance modularity and readability.
- Organized and presented testing procedures to the team to improve team practices.

Gustavus Adolphus College, Teaching Assistant — St. Peter, MN
Feb 2021 – June 2021

- Assisted professor in managing lab sessions and projects.
- Graded assignments for quality and completeness.
- Provided individual support to students to troubleshoot and understand concepts.

Projects

BERTopic Podcast Topic Modeling — GitHub Repo

- Fine-tuned transformer model BERTopic to accurately extract topics from podcasts and YouTube videos.
- Pulled and cleaned high quality transcripts from Youtube videos.
- Analyzed topic trends of different podcasts using almost 3,000 videos.

Expectation Maximization on Movie Reviews — GitHub Repo

- Classified individuals into 4 types of movie watchers, which can be used to predict each person's reviews on future movies.
- Accomplished this through applying 256 iterations of Expectation Maximization on a dataset of movie reviews.

Transformer — GitHub Repo

- Built a transformer-architecture encoder and decoder from scratch.
- Adjusted various architecture parameters such as positional embeddings to maximize accuracy.

Machine Learning Engineer Nanodegree — GitHub Repo

- Trained and deployed model to predict inidividuals' general health given survey data, resulting in an R-Squared value of .5.
- Created projects utilizing model deployment and endpoint consumption.

N-Gram Log Liklihood — GitHub Repo

- Built a program that computes the unigram and bigram log likelihood of a sequence of characters, tokenizing by word.
- Combined the unigram and bigram models to compute the log likelihood of a sentence.