Anany K. Singh

+1 (716) 573-2402 | ananykumarsingh25@gmail.com | linkedin.com/in/anany-singh | github.com/Anany25

Summary

AI Engineer with hands-on experience designing and deploying scalable systems in natural language processing, computer vision, and deep learning. Skilled in LLMs, Retrieval-Augmented Generation (RAG), and MLOps, with proven ability to build and optimize end-to-end AI pipelines from data preprocessing and model training to deployment. Passionate about advancing Generative AI, Reinforcement Learning, and Prompt Engineering to drive real-world innovation.

Skills

Roles	AI Engineer, Machine Learning Engineer, Applied Scientist
Languages	Python, C++, SQL, R
Frameworks & Libraries	PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, Hugging Face, LangChain, Kornia, Flask
Cloud & DevOps	Google Cloud Platform (GCP), AWS, Microsoft Azure, Docker, Kubernetes, CI/CD
Machine Learning Concepts	Machine Learning Models, Data Pipelines, Model Evaluation, Feature Engineering, Model Optimization
Developer Tools	Git, GitHub, Linux, Jupyter Notebook, MySQL, Tableau, Power BI

Professional Experience

UGenome AI Oct 2025 - Present

AI Software Engineer *Buffalo, NY*

- Developing a conversational AI platform for personalized genomics, enabling users to interpret and explore genomic data through an LLM-powered chatbot.
- Implementing a Retrieval-Augmented Generation (RAG) pipeline leveraging Knowledge Graphs and vector databases to deliver precise, context-aware insights.
- Building an automated extraction and structuring workflow for bioinformatics datasets to populate a reasoning-ready Knowledge Graph.
- Deploying the production-scale system on Microsoft Azure using PyTorch, LangChain, and Azure AI Foundry, ensuring reliability and scalability.

Projects

Advanced Image Colorization GAN *(PyTorch, Vision Transformer, GANs)* Aug 2025 - Sep 2025

Personal Project (Sole Contributor)

- Developed a hybrid GAN combining U-Net and Vision Transformer (ViT) to enhance global and local contextual understanding in grayscale image colorization.
- Trained on over 40,000 COCO images, achieving a 12% improvement in LPIPS perceptual similarity over a baseline U-Net.
- Applied Spectral Normalization and Two Time-Scale Update Rule (TTUR) to stabilize adversarial training and ensure convergence.
- Designed a composite loss function blending adversarial, Charbonnier, and perceptual (LPIPS) losses to maximize realism and color fidelity.

Large-Scale Search Portal *(PyTorch, GCP, Flask, FAISS, LLM)* Sep 2024 - Nov 2024

Academic Project at University at Buffalo

- Built a semantic search engine for 50,000+ Wikipedia articles using sentence-transformer embeddings and FAISS vector indexing.
- Reduced query latency by over 65% (to under 8 seconds) through FAISS optimization and Maximum Marginal Relevance (MMR) retrieval.
- Integrated Llama 3 with Flask API to generate concise, context-aware summaries from retrieved results.
- Deployed the scalable system on Google Cloud Platform using Compute Engine for production-level performance.

Education

University at Buffalo, The State University of New York

Master of Science in Engineering Sciences (Artificial Intelligence) *Buffalo, NY*

- GPA: 3.53/4.00

Vellore Institute of Technology

Bachelor of Technology in Computer Science & Engineering *Vellore, India*

- GPA: 3.72/4.00 (WES Evaluated)

Certifications & Awards

- Mathematics for Machine Learning Specialization – Imperial College London (Coursera)
- Machine Learning on Google Cloud Specialization – Google Cloud (Coursera)