Anshumaan Singh

+1 (934)-219-4495 | Stony Brook, NY | anshumaanvsingh@gmail.com

www.linkedin.com/in/anshumaanvsingh | Web Portfolio: krosskinetic.github.io | GitHub: github.com/KrossKinetic

---

EDUCATION

Stony Brook University | Stony Brook, NY   *August 2023 - May 2027 (Anticipated)*

Bachelor of Science in Computer Science

GPA: 3.93/4.0

Honors: Computer Science Honors, University Scholars, Dean's List 2023 - 2025

Relevant Coursework: Object-Oriented Programming, Data Structures, System Fundamentals, Linear Algebra, Probability and Statistics, Discrete Math, Calculus I & II

TECHNICAL SKILLS

Programming Languages: MIPS, Python, Java, C, Kotlin (for Android Jetpack Compose), OCaml, JavaScript, HTML/CSS

Frameworks & Libraries: TensorFlow, LangChain, HuggingFace, OpenCV, Pandas, NumPy, Flask, Bootstrap

Developer Tools & Platforms: Docker, AWS, Git, GitHub, Linux/Unix, HPC Clusters (Slurm), MySQL, SQLite3

PROFESSIONAL WORK EXPERIENCE

Machine Learning Researcher   *June 2025 - Present*

Stony Brook University's LUNR AI Lab | Stony Brook, NY

- Developed a Python script to programmatically download a 100,000-sample dataset, balanced across specified languages (e.g., Python, C, C++, based on input parameters), from the public bigcode/StackV2 S3 repository.
- Adapted the REPLUG LSR framework for a code-based RAG baseline by implementing a Python driver script for execution and replacing its API dependency with a local vLLM server, reducing direct inference costs by over 99%.
- Developed a unified Python client to benchmark and query commercial APIs (GPT-4o, Claude 3.5) and local models with vLLM, implementing SQLite3 caching to eliminate over 85% of redundant API calls during testing.

Volunteer Machine Learning Engineer   *November 2024 – February 2025*

Humanity Unleashed (Humun.org) | Remote

- Built pipelines processing 700k rows of economic data, publishing 5 datasets to HuggingFace with 1,400+ downloads.
- Collaborated with 2 other peers to develop a policy generation and summarization pipeline utilizing LangChain to translate complex economic data into nuanced policy explanations.

Lead Web Developer & Machine Learning Researcher   *February 2024 – October 2024*

International Genetically Engineered Machine | Stony Brook, NY

- Engineered a Q&A chatbot using LangChain and HuggingFace that creates vector embeddings of custom research documents and ingests them to generate context-aware answers, ensuring specialized responses with high accuracy.
- Led and mentored a 3-person wiki team within a 15-person research group, developing a project wiki (HTML, CSS, JavaScript, and Flask) & attracting 15 stakeholders and investors.

Full Stack Web Developer Intern   *April 2023 – June 2023*

Omniscience Corporation | Remote

- Developed a full-stack life insurance web application using Flask, MySQL, JavaScript, and Bootstrap.
- Containerized with Docker and deployed on AWS for a reliable and scalable production environment.

PROJECTS

NotiSentry | Android Application & Machine Learning Project   *June 2025 - Present*

- Architecting a smart DND Android app using Kotlin and Jetpack Compose that leverages on-device Gemini Nano for privacy-centric summaries and critical alert filtering while prioritizing UX and accessibility using Material 3 Design.

Windows CenterStage | Computer Vision Project   *November 2023 – December 2023*

- Built a Python program that uses a custom-trained facial recognition model to track and center a user's face on camera, leveraging TensorFlow and OpenCV on a dataset of over 6,000 images.