SHLOK TOMAR

(312)-375-7583 | shloktomar65@gmail.com | linkedin.com/in/shlok-tomar-us/ | https://shlok-crypto.github.io/

EDUCATION

Washington State University Pullman, USA

Master of Computer Science (Thesis) ***Prof (Dr.) Janardhan Rao Doppa*** ***GPA****:* 3.9/4.0 *Graduated Jan 2025*

Relevant Coursework: Neural Network Design, Big Data Analysis, Reinforcement Learning and Advanced Algorithms

Amity University Noida, India

Bachelor of Science in Computer Application ***GPA***: 3.8/4.0 *Graduated May 2022*

SKILLS/TECHINICAL/TOOLS

Languages: Python, C++, C, R, Java, JavaScript, PHP, SQL, .NET Core, HTML, CSS, Node.js

Frameworks: TensorFlow, PyTorch, OpenCV, Scikit-learn, Keras, Flask, React, Azure, AWS, MongoDB, GitHub, Hadoop, Spark

Skills: OOP, Tableau, Power BI, Excel, LLM, Transformer (Large Language Models), NLP (Natural Language Processing), Linux

WORK EXPERIENCE

Production Engineer/ DevOps Intern Seattle, WA

Meta (Facebook-MLH) *June 2025 – September 2025*

- Worked with Production Engineers to design reliable, scalable, production-ready system, achieving 97.6% uptime and serving users globally
- Reduced manual deployment time by 95% by designing and building a complete CI/CD pipeline with Internal Git Actions that automated all building, testing, and release processes to a live server.
- Utilized Docker and managed scalable cloud infrastructure to support high-volume data processing and real-time model serving, helping decrease setup time by over 87% and ensure consistency across development and production.
- Implemented a Prometheus and Grafana monitoring stack to improve operational efficiency, creating custom dashboards to analyze container-level CPU, memory, and network performance during 10,000-request load tests.

Project Lead AI/ML - LLM Engineer Pullman, WA

Washington State University *August 2023 –Present*

- Developed BODE-GEN, a Bayesian Optimization framework for LLM-driven program synthesis, achieving 96% pass@1 accuracy on HumanEval+ benchmarks with models like ChatGPT-3.5, DeepSeek-Coder-33B, and CodeLlama-7B.
- Integrated an auxiliary LLM with Gaussian Processes, and dimensionality-scaled priors, enhancing surrogate modeling in sparse data settings with a 23% reduction in inference variance.
- Optimized sampling efficiency by 13× and reducing token consumption from 113K to 8.7K per evaluation, leading to a 6.2× decrease in API costs, while boosting test-case pass rates by 37% over Chain-of-Thought (CoT) and OPRO

Lead Software Engineer AI/ML Intern Berkeley, CA

UC Berkeley & United States Department of Agriculture (USDA) *May 2023 – August 2023*

- Developed computer vision pipeline using CNNs (YOLO, Faster R-CNN) to detect cervids in low-resolution satellite imagery.
- Designed and trained deep learning architectures to differentiate Chronic Wasting Disease (CWD) infected vs. healthy animals based on physiological markers. achieving a 96.4% accuracy
- Analyzed 2M+ GPS collar data points, mapping elk movement trajectories, and herd structures. Applied geospatial clustering, time-series forecasting (ARIMA, LSTMs), and epidemiological modeling to predict infection hotspots.
- Worked directly with US Government's *SCI-Net High-Performance Computing* (HPC) unit to optimize computational workflows and improve model efficiency by 67.3%

Software/ Internal Technology Intern Roorkee, India

Microsoft *September 2021 – February 2022*

- Gained expertise in Azure, ML, AI, and Computer Vision through application-based projects with Microsoft engineers.
- Developed an AI-powered Windows application to detect improper body posture during exercise.
- Performed data analysis on system performance, identifying key usability trends and optimization strategies.

PROJECTS

FlavorBlend – AI-Powered Fusion Recipe Generation app

- Developed FlavorBlend, an AI-powered fusion recipe generation app within 24 hours, using Meta Llama 3-8B Instruct for text generation and Stable Diffusion XL Base 1.0 for image generation, enabling real-time recipe creation.
- Built a Node.js and Express.js backend to process user inputs, generate fusion recipes, and store data in MongoDB, integrating Firebase authentication for secure, real-time sharing
- Created a Flask web app with HTML, CSS, and JavaScript, ensuring accessibility across both web and mobile platforms.
- Won "Best Gen AI App on Cloudflare" and Swire Coca-Cola Track Winner ($2,000 scholarship) at CrimsonCode'25, competing against 50+ teams.

AgAID Digital Hackathon – Snowpack Predication Challenge

- Developed a Transformer based Neural Network model to predict Snow Water Equivalent (SWE) with 87.8% accuracy, aiding water resource management decisions in the Western U.S.
- Processed and integrated 9M+ rows of meteorological data from 8 unstructured datasets, using chunked loading, spatial joins, and KNN imputation to improve data consistency and model performance.
- Built and deployed a full-stack web application with React and Flask, enabling real-time SWE predictions and interactive data visualizations from user-uploaded CSV files.
- Earned an Honorable Mention for ranking in the top 3, recognized for innovative AI-driven solutions and real-time prediction capabilities.

FBI Hate Crime Data Analysis *Apr 2022*

- Analyzed FBI's annual hate crime dataset to draw insights on racial hate crime trends by factors such as bias criteria, geography, and frequency of crime
- Pre-processed and cleaned 21,000+ data entries with Pandas, decreasing processing time by 34 %
- Visualized data trends using Pandas and matplotlib to help improve analysis interpretation with users

LEADERSHIP/ VOLUNTEERING

Teaching Assistant |Algorithm Design| Neural Networks | Advance Computer Vision August *2023 - 2024*

- Assisted in the delivery of Algorithms Design, Advanced Computer Vision, and Neural Networks courses, aiding in the development of course content and providing technical support for programming assignments and projects.
- Led review sessions, helping students debug and optimize algorithms, implement advanced computer vision techniques, and apply neural network architectures for various problem-solving tasks.

Delegate, Harvard Project for Asian and International Relations (HPAIR) *June 2024*

- Accepted for the prestigious international conference focused on addressing key global issues in Asia and beyond.
- Engaged in high-level discussions and workshops on international relations, economic development, and innovation
- Networked with global leaders, policymakers, and distinguished academics to foster cross-cultural collaboration.

PUBLICATIONS

Shlok Tomar, Aryan Deshwal, Ethan Villalovoz, Haipeng Cai, Janardhan Rao Doppa. *Test-Driven Code Generation using LLMs via Bayesian Optimization*. In Proceedings of the 2025 AAAI Conference on Innovative Application of Artificial Intelligence (IAAI 2025) (Under Review)

Shlok Tomar, Aryan Deshwal, Ethan Villalovoz, Mattia Fazzini, Haipeng Cai, Janardhan Rao Doppa. *Sample Efficient LLM-Driven Program Synthesis: A Novel Bayesian Optimization Approach*. In Proceedings of the 2025 IEEE/ACM International Conference on Software Engineering (ICSE 2025) (Under Review)