Rohan Praveen Chavan

Blacksburg, VA | +1-540-824-9961 | rohanchavan0701@gmail.com
LinkedIn | GitHub | Portfolio

Education

Virginia Polytechnic Institute and State University Aug 2024 – May 2026
M.S. in Computer Engineering, GPA: 3.68/4.0 *Blacksburg, VA*

K.J. Somaiya College of Engineering Aug 2020 – May 2024
B.Tech in Information Technology (Distinction; Honors in AI, GPA ˜3.8/4.0) *Mumbai, India*

Experience

AI/ML Intern May 2025 – August 2025
AutoUnify (San Francisco, California)

- Optimized and fine-tuned LLMs on Google Vertex AI for worker and QA pipelines, boosting task accuracy by 30%+.
- Built a scalable evaluation framework to benchmark multiple LLMs on software specification and code-generation tasks, enabling data-driven deployment decisions.
- Engineered prompt-tuning strategies that improved model consistency on complex API specs (100+ models, 500+ endpoints), reducing manual corrections by 25%.
- Collaborated on the **AgentMo system**, defining templates, refining prompts, and integrating LangGraph-based AgentModels to strengthen multi-agent orchestration and HITL workflows.
- Developed Vertex AI + Gemini pipeline to auto-generate and validate schemas, cutting manual validation by 25+ hours/week across 10+ services.
- Implemented GitOps with GitHub Actions and Pytest, increasing CI/CD merge velocity by 20% and deployment reliability.
- Skills: Vertex AI, Gemini, Prompt Engineering, AgentMo, LangGraph, TypeSpec, Azure, GitHub Actions, Pytest

Intelligent Process Automation Intern Jan 2024 – Jun 2024
Colgate-Palmolive GBS *Mumbai, India*

- Designed and deployed 3 scalable AI-powered chatbot solutions for O2I and S2P workflows using Python, ChatGPT, and Kore.ai, automating 70% of manual queries and enhancing workflow efficiency by 40%.
- Engineered and integrated a RESTful ChatGPT API, reducing average response latency from 30s to under 10s, while applying Logistic Regression and Decision Trees for fine-tuned query classification.
- Built a BERT-based NLP pipeline with React.js frontend and PyTorch/TensorFlow backend, improving user adoption by 50% and winning 1st place in a departmental hackathon (**+35% accuracy**, **-60% resolution time**).

Team Member — Amazon Nova AI Challenge Jan 2025 – June 2025
Virginia Tech (Team HokieTokie) *Blacksburg, VA*

- Led red teaming for Amazon Bedrock models, creating a taxonomy-guided synthetic data pipeline (117K samples) spanning CWE vulnerabilities and MITRE ATT&CK adversarial behaviors, improving model security generalization.
- Developed adversarial attack frameworks (**PAIR, PEZ, Crescendo**) that generated 3,000+ jailbreak prompts and uncovered 2,480 unique vulnerabilities in LLM code generation.
- Built **ShieldBot** with GPT-4 and Claude, achieving 56s detection latency to block 50+ adversarial prompts daily for 10K+ queries.
- Designed iterative patch-and-retry loops with Amazon CodeGuru and skill-based augmentation (983 atomic skills in 7 attack families) to systematically eliminate vulnerabilities.
- Engineered a two-stage ensemble (Vulnerability + Refusal Experts) via SFT/DPO fusion, balancing refusal robustness with secure code generation.
- Achieved **97% reduction in malicious compliance**, **62% fewer vulnerabilities**, and 46% lower multi-turn jailbreak success; ranked **1st in Tournament 2** and **2nd in Tournament 1**, outperforming Claude-3.7, Gemini-Pro, and CodeLlama-70B.
- Skills: LLM Red Teaming, Synthetic Data Generation, Adversarial ML, Secure Code Gen, Amazon CodeGuru, Claude, GPT-4, DeepSeek

Academic Projects

SentiMint — AI Stock Analyzer (LangChain, GPT-3.5, Selenium, Next.js, Flask) **[GitHub]**

- Built a LangChain multi-agent system for real-time sentiment + fundamentals, achieving **85% trend accuracy** across 500+ Yahoo Finance datasets.
- Designed debate-style personas and token-confidence scoring, **cutting decision latency by 40%** for 100+ users at Agent Hacks 2024.

Technical Skills

Languages: Python, C++, SQL, JavaScript, Go
Frameworks & Libraries: PyTorch, TensorFlow, LangChain, HuggingFace, Django, React.js, OpenCV, FAST API, RAG
Cloud & DevOps: AWS (Lambda, API Gateway, S3, ECS), Google Cloud, Azure, Docker, Kubernetes, Terraform, GitHub Actions, Jenkins, IaC
Data/Databases: MongoDB, MySQL, PostgreSQL, Cosmos DB, Firebase, Supabase
Tools: GitHub, GitLab, Pytest, Hadoop, Apache Spark, Tableau, MATLAB, Jira, Figma
Integration: REST APIs, OpenAI APIs, Chatbot Development, CI/CD Pipelines, MLOps