

Aadil Mallick

703-712-9239 | aadiltj.mallick@gmail.com | [linkedin.com/in/aadil-mallick](https://www.linkedin.com/in/aadil-mallick) | github.com/aadilmallick | aadilmallick.com

PROFESSIONAL SUMMARY

Software Engineer and recent graduate from William & Mary with a B.S. in Computer Science and Applied Mathematics. Developed generative AI and Retrieval-Augmented Generation (RAG) applications in Python using Ollama, Langchain, FastAPI, and Pydantic. Designed testable backend APIs and modular services, then containerized and deployed them using Docker, Kubernetes, AWS and CI/CD. Worked in production software across research, defense, and federal-contracting domains. Proficient in Claude Code, Cursor, Codex, agent skills, and agentic loops.

- **AI & Generative AI:** Built LLM-powered applications using RAG, Langchain, embeddings, and prompt design. Executed model inference on local and cloud infrastructure (Ollama, HuggingFace, OpenAI).
- **ML Math Theory Background:** Developed a strong theoretical foundation through coursework in Probability Theory, Abstract Algebra, Multivariable Calculus, Linear Algebra, Machine Learning, and Neural Networks.
- **ML/AI Coding:** Experienced in Python-based data science and machine learning development using libraries such as NumPy, Pandas, TensorFlow, PyTorch, and Scikit-learn.
- **AI security:** Used cybersecurity, networking, evals, and practical web security to reinforce AI systems and protect against prompt injection and data poisoning.

EDUCATION

College of William & Mary Williamsburg, VA
Bachelor of Science in Computer Science and Applied Mathematics; GPA: 3.70/4.0 May 2026

- **Relevant Coursework:** Neural Networks, Linear Algebra, Multivariable Calculus, Differential Equations, Cybersecurity, Machine Learning, Abstract Algebra, Probability, Data Science.

Thomas Jefferson High School for Science and Technology Alexandria, VA
High School Diploma; GPA: 4.1/4.0 June 2022

- **Relevant Coursework:** AP Computer Science A+, Artificial Intelligence, Web Development.

EXPERIENCE

Software Engineer June 2026 – Present
Unison McLean, VA

- Architected AWS infrastructure for an AI-powered federal contracting product called BidScale using AWS CDK, TypeScript, and GraphQL.
- Managed automatic provisioning of AWS CDK L2 constructs to dedicated VPCs and public and private subnets within availability zones to harden network security.
- Improved the developer experience in the company repository by adding devcontainers, Docker Compose local development configuration, TMUX configuration, and AWS CDK unit tests to test CloudFormation provisioning.

CS Research Lab Assistant Intern September – December 2025
W&M AI Lab Williamsburg, VA

- Designed Langchain RAG experiments to evaluate model accuracy of structured outputs on clinical medical trial datasets, then extracted classification metrics including Jaccard similarity, F1 score, and cosine similarity.
- Coauthored and published a paper in the *Journal of Software: Evolution and Process* publication and ran structured output inference experiments in Python on Ollama and HuggingFace models at scale to predict CVSS vulnerability scores of code snippets, attaining 83% accuracy.

Full-Stack Developer Intern June – August 2025
Howard Hughes Medical Institute (HHMI) Chevy Chase, MD

- Designed a Kubernetes-based distributed, generic services architecture system that monitored 112+ company internal websites for KPIs and Jenkins build pipelines statuses.
- Built a full-stack, real-time dashboard to visualize internal site KPIs and monitor Kubernetes cluster health using MongoDB, Express, TypeScript, Node, React, cron jobs and microservices communication via REST APIs.
- Architected a robust test-driven development framework by using Vitest for unit tests and Playwright for end-to-end tests, and then integrating CI/CD via Github Actions and automated EKS deployment.

Frontend Developer Intern June – August 2024
Two Six Technologies Arlington, VA

- Resolved 11 high-priority production bugs in an Agile-based team through systematic debugging, code reviews, and implementation of automated testing with Jest, improving system reliability by 25%.

PROJECTS

- VibeResume** | *Next.js, Gemini API, PostgreSQL, Clerk* | [Demo](#) 2025
- Won W&M hackathon (Best Use of Gemini API) with an AI-powered SaaS that generates portfolio websites from resumes using Next.js, Gemini API, Vercel hosting, Clerk authentication, and PostgreSQL with Neon.
 - Tracked real payments with Clerk billing and implemented robust, provider-agnostic LLM inference with Vercel AI SDK, long polling for realtime updates, and Inngest for background jobs.
- BetterCare Meals** | *React, TypeScript, Supabase, Stripe* | [Website](#) 2025
- Performed freelance work for a client by architecting a serverless E-commerce website selling healthcare-based meals that processes real payments; built with Supabase cloud functions, Google OAuth, Stripe, and hosted on Vercel.
 - Hardened security for the application by applying CORS and rate limiting for the serverless functions and applying row level security policies to the PostgreSQL database.
- MCP TypeScript Abstraction Library** | *TypeScript, Deno, Model Context Protocol (MCP) SDK* | [Repo](#) 2025
- Abstracted the MCP TypeScript SDK to provide object-oriented helpers over creating stdio and SSE based servers, making deployment and developer UX faster.
- Local OpenAI Whisper AI transcription GUI** | *Python, Streamlit, Docker, OpenAI Whisper* | [Demo](#) 2024
- Developed self-hosted audio transcription service using OpenAI Whisper model with Docker containerization, achieving 95% accuracy on 2+ hour audio files and reducing transcription time by 80%.

PUBLICATIONS

- On Predicting Vulnerability Severity Using In-Context Learning: An Industrial Case Study**
J. Software: Evolution and Process | [Article](#) 2026
- Co-authored a peer-reviewed paper on using large language models to predict CVSS vulnerability severity scores of code snippets, using in-context learning with local open-source LLMs.
 - Designed and ran the structured-output experiments, orchestrating batch concurrent Ollama and GPT-4o-mini API calls over the Big-Vul dataset using async I/O and multithreading in Python.
 - Enforced schema-constrained 0–10 score outputs using Pydantic with hard failure on non-conformance, then aggregated regression and stratified classification metrics, then ran the pandas-based analysis pipeline.
- Claude Skills Are THE Cheat Code of AI — Learn Them Now.** | *Stackademic* | [Link](#) 2026
- Wrote a deep-dive guide into how Claude/Agent skills work, tradeoffs between MCP and agent skills, custom skill creation, and how to integrate them with Claude Cowork, Claude Desktop, and Claude Code.
- Learn to Build a Neural Network From Scratch** | *Stackademic* | [Link](#) 2024
- Authored a 20,000-word technical article published in *Stackademic* on Medium explaining neural network mathematics from first principles, including linear algebra, gradients, optimizers, backpropagation algorithms and NumPy implementations, garnering 872+ claps and 60,000+ views.

TECHNICAL SKILLS

Languages: Python | JavaScript/TypeScript | Java | HTML | CSS | C#
AI Technologies: Claude Code | OpenAI Codex | Langchain | Antigravity | MCP | Ollama
AI/ML Frameworks: PyTorch | TensorFlow | NumPy | Pandas | Scikit-learn | HuggingFace
Web Technologies: Next.js | React | FastAPI | NodeJS | Express | Stripe | REST APIs | GraphQL
Databases & Cloud: PostgreSQL | MongoDB | DynamoDB | Redis | Firebase | Supabase | SQLite | Vercel | AWS
Enterprise: Jira | BitBucket | Jenkins | AWS multi-organizations | Agile
Developer Tools: ESLint | Vim | TMUX | Bash Scripting
DevOps: Docker | Kubernetes | Github Actions | Git | Linux | Pydantic | Playwright | CI/CD