

Naga Venkata Sai Chennu

Fairfax, VA | 571-546-6207 | nchennu@gmu.edu | [linkedin.com/in/naga-venkata-sai-chennu](https://www.linkedin.com/in/naga-venkata-sai-chennu) | nagavenkatasai7.github.io/portfolio

Software Engineer • Scalable Systems, Test Automation, AI-Assisted Development • MS CS, May 2026

EXPERIENCE

Graduate Research Assistant — AI Systems & Test Automation Engineering

Aug 2025 – May 2026

George Mason University, Costello College of Business, Fairfax, VA

- Architected and built llm-forge, a configuration-driven 12-stage automation pipeline in object-oriented Python that lets researchers launch jobs from a single config file, cutting time-to-experiment by over 50% across a shared 4x H100 SLURM cluster.
- Designed an automated test framework with ~900 unit and functional checks (Pydantic v2 schema validation) at every pipeline stage, driving up coverage and catching configuration errors before runtime to eliminate costly job failures and reduce support requests.
- Implemented and delivered an extraction service (Python, Claude API) that converted 500+ multi-page legal PDFs into validated structured JSON, reducing manual processing time ~80%, and gated low-confidence output against human-labeled ground truth for quality control.
- Used Claude Code, GitHub Copilot, and Cursor to accelerate delivery, critically reviewing and refactoring AI-generated code before merge to keep the codebase maintainable.
- Instrumented pipeline telemetry and performance metrics, surfacing low-confidence outputs and run failures to drive operational visibility and quality.
- Partnered with faculty stakeholders to translate requirements into technical specs and led weekly progress reviews, delivering workflows with measurable accuracy improvements.

Undergraduate Researcher — Software Engineering

Jan 2022 – May 2024

KL University, Department of CSE, India

- Built an end-to-end machine-learning system in Python combining four classifiers with ensemble methods (bagging, boosting, stacking), reaching 93% accuracy and 92% F1, validated through t-test and ANOVA.
- Engineered an Ethereum land-registration prototype using object-oriented Solidity smart contracts and SHA-256 hashing for tamper-resistant transfers, validated across 200 transactions on a 12-node network.
- Evaluated competing models (ANN vs. CatBoost) and integrated the higher-performing CatBoost model (89% accuracy, 91% precision) into a web-based analytics dashboard.

PROJECTS

SmartRemit — Multi-Tenant Payments Platform | TypeScript, NestJS, Next.js, PostgreSQL

- Architected a multi-tenant SaaS backend connecting licensed money-service partners to end users through a conversational bot, built on object-oriented NestJS services, a Next.js front end, and PostgreSQL.
- Designed for scale and stability with a partner-abstraction layer, idempotent transaction handling, and clear service boundaries so new corridors can be added without destabilizing the platform.

JetBot — Edge AI Agent on Jetson | Python, SQLite/FTS5, systemd, Linux

- Designed and deployed a Telegram-controlled AI agent on Jetson Orin Nano edge hardware in Python, with SQLite/FTS5 full-text search and human-approval gates guarding destructive actions.
- Hardened runtime stability with systemd and cgroups v2 resource isolation, balancing LLM inference against constrained edge compute.

CUDA Matrix-Multiplication Kernel | C++, CUDA

- Implemented a tiled matrix-multiplication kernel in C++/CUDA and profiled memory and tiling strategies, benchmarking throughput against the cuBLAS baseline.
- Tuned shared-memory usage and thread-block sizing to improve arithmetic intensity, measuring the speed and occupancy tradeoff at each step.

TECHNICAL SKILLS

Languages: Python, Java, JavaScript/TypeScript, SQL, C++, HTML

Object-Oriented Design: OOP and design patterns, REST APIs, microservices (NestJS), service architecture

Testing & Quality: unit and functional testing, test automation frameworks, Pydantic v2 schema validation, code coverage, telemetry and metrics

AI Coding Assistants: Claude Code, GitHub Copilot, Cursor, Gemini; reviewing and guiding AI-generated code

Cloud & Platforms: AWS, Docker, PostgreSQL, FastAPI, Linux/SLURM

EDUCATION

M.S. Computer Science — George Mason University, Fairfax, VA

May 2026

B.Tech Computer Science — KL University, India | CGPA 8.7/10

2020 – 2024

PUBLICATIONS

Co-author of 5 peer-reviewed publications (2 first-author) and 1 patent, with 50+ citations across AI/ML, data analytics, and NLP.

CERTIFICATIONS

Anthropic — Claude Code in Action (Dec 2025) | NVIDIA Certified Associate, Generative AI and LLMs (Feb 2026) | AWS Certified AI Practitioner (Nov 2025)