

NAGA VENKATA SAI CHENNU

AI Engineer

Fairfax, VA | (571) 546-6207 | nagavenkatasai.chennu@gmail.com

[linkedin.com/in/naga-venkata-sai-chennu](https://www.linkedin.com/in/naga-venkata-sai-chennu) | github.com/Nagavenkatasai7 | chennunagavenkatasai.com

EXPERIENCE & RESEARCH

SmartRemit | Remote

May 2026 – Present

Co-founder & Lead Engineer (sole engineer on a two-person founding team)

Pre-launch startup designed to let people send money abroad over WhatsApp through licensed money-transfer companies.

- Led development of the WhatsApp chatbot (Next.js, TypeScript, WhatsApp Cloud API) that completes a transfer in one chat, using LLM tool calling to quote exchange rates, run Persona KYC, collect recipient details, and send a payment link.
- Designed the payment backend so no payment is ever charged or paid out twice, using idempotency keys and a transactional outbox in PostgreSQL, with envelope-encrypted customer data and a full audit trail.
- Built a partner dashboard, REST API, and signed webhooks for reviewing transfers, checking compliance, and connecting partner systems; deployed on Vercel with Neon Postgres and Upstash Redis.

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

Aug 2025 – May 2026

AI Systems Engineer (Graduate Research Assistant)

- Partnered with Dr. Saurabh Mishra to automate his trademark research: built a Python scraper with caching and retry handling that collected ~200 USPTO opposition filings from TTABVue and TSDR into structured records.
- Classified trademark images with Claude Vision (multimodal LLM) at ~90% accuracy, cutting manual review ~60%.

KL University — Department of Computer Science & Engineering | India

Jan 2022 – May 2024

Undergraduate Researcher, Applied ML

- Implemented 4 classifiers (Decision Tree, Random Forest, SVM, Logistic Regression) tuned with grid search and cross-validation, then combined them into an ensemble for hair-fall prediction: 93% accuracy vs. 89% for the best single model.
- Published 2 first-author papers (hair-fall prediction, AI in cybersecurity); presented team research at IEEE ICICT 2023.

PROJECTS

VLM Chart-Understanding Study — CS 692, advised by Prof. Bo Han

- Evaluated GPT-5.4 on chart question answering with the ChartX benchmark (6,000 images, 18 chart types). Accuracy fell from 85.7% on 2D charts to 59.3% on 3D, significant under McNemar's test.
- Found why: 3D views hide or distort parts of the chart. Also found a benchmark flaw, where some questions asked for exact numbers the image never showed; adding data labels raised accuracy from 51.6% to 97.7%.

Finance-Tuned Llama 3.2 1B — Personal Project

- Fine-tuned it on 5,675 finance examples with LoRA (Hugging Face PEFT, TRL SFTTrainer) and published it on Hugging Face.
- Benchmarked each version on MMLU, GSM8K, and IFEval; an earlier run lost 7.4 MMLU points, so fixed the cause and the final model held at 45.9% vs. 46.1% base.

Mail Agent — Self-Hosted AI Mail Platform | TypeScript

- Created a web app that combines 5 email accounts (Gmail, iCloud, Fastmail, Zoho) into one inbox, with LLM triage and reply drafting (Gemma via Ollama, local or cloud).
- Encrypted credentials with AES-256-GCM and sandboxed email HTML in an iframe; follow-ups stop on reply or unsubscribe.

PUBLICATIONS

- [Enhancing Hairfall Prediction: A Comparative Analysis of Individual Algorithms and an Ensemble Method](#) — IJRITCC, 2023
- [Assessing the Effectiveness of Artificial Intelligence Techniques in Mitigating Cybersecurity Risks](#) — IJISAE, 2023

5 peer-reviewed papers (2 first-author, 50+ citations), including [IEEE ICICT 2023](#). Full list: orcid.org/0009-0000-8252-8682

EDUCATION

M.S., Computer Science — George Mason University | Fairfax, VA | GPA 3.57 / 4.00

May 2026

B.Tech, Computer Science & Engineering — KL University | India | GPA 8.7 / 10

May 2024

TECHNICAL SKILLS

Languages & Backend: Python, TypeScript, SQL, Bash, Next.js, FastAPI, PostgreSQL, Redis

LLM & Agents: Claude API, OpenAI API, Ollama, tool calling, RAG, embeddings, hybrid search, Chroma, prompt engineering

ML & Training: PyTorch, Hugging Face Transformers, TRL, PEFT, LoRA, QLoRA, DPO, GRPO, scikit-learn

Deployment & Evaluation: Docker, CI/CD, Vercel, AWS, vLLM, GGUF, quantization, LLM evaluation (MMLU, GSM8K, IFEval)

CERTIFICATIONS

[AWS Certified AI Practitioner](#) | [NVIDIA-Certified Associate: Generative AI LLMs](#) | [Anthropic: Claude Code in Action](#)