Tulasi Venkata Sri Varshini Padamata

AI / Machine Learning Engineer · San Francisco Bay Area, CA

585-623-7981 · srivarshini09.edu@gmail.com · linkedin.com/in/ptvsv · github.com/SriVarshini09 · portfoliosri-chi.vercel.app

SUMMARY

AI/ML engineer (M.S. in AI, RIT) building LLM agents, retrieval, evaluation and production ML infrastructure. Shipped document AI and semantic search on AWS that cut LLM inference cost 35% and analyst lookup time 60%. Published researcher; building SHAPI.

EDUCATION

M.S. in Artificial Intelligence — Rochester Institute of Technology, Rochester, NY · Aug 2024 – May 2026

Coursework: Machine Learning Applications, Distributed Systems, Engineering Scalable Systems

Bachelor's in Computer Science and Engineering — R.M.K. Engineering College, India · Aug 2019 – May 2023

EXPERIENCE

Graduate Researcher — Rochester Institute of Technology, Rochester, NY · Jan 2026 – Apr 2026

- Built a zero-shot multi-agent pipeline that detects RAG hallucinations by verifying each claim against source passages.
- Reached 0.642 F1 and 0.764 recall on 600 RAGTruth records (QA, summarization, data-to-text) with no training data or GPU.
- Designed task-aware decision rules that gave the largest gain (+0.13 F1) at about $0.0003 per record.

Machine Learning Engineer — Novacis Digital LLC, India · Jun 2023 – Feb 2024

- Built a SageMaker document AI system with LayoutLM and BERT, improving F1 by 18% on 10K+ legal documents.
- Deployed an OCR + LLM pipeline on Amazon EKS that cut document review time to under 10 minutes per document.
- Built FAISS-based semantic search with 300 ms retrieval, reducing analyst lookup time by 60%.
- Benchmarked GPT-4 against Claude for contract analysis, reducing LLM inference cost by 35%.

PROJECTS

SHAPI — Self-maintaining APIs · TypeScript, Node.js, MCP, Claude API, PostgreSQL · 2026 – present · shapi.dev

- Maps every third-party API call site into an api.lock and flags the ones a known breaking change hits.
- Proves each fix with a contract test before opening a pull request; ships as a CLI, MCP server and agent skill.

Nori — AI job-search agent · TypeScript, Next.js, MCP, Composio, PostgreSQL + pgvector, Playwright · Oct 2026 · jobs.norinote.xyz

- Career-graph agent: every tailored resume bullet traces to a verified fact; fills ATS forms with review-before-submit.
- Built at the Cline × Composio × Parallel hackathon, San Francisco; driven from Slack, Telegram, WhatsApp, email or MCP.

Kerna — Trust layer for AI agents · Rust, MCP, SQLite · kerna.run

- Local-first runtime enforcing fail-closed tool permissions, human approvals and hard budgets; logs a receipt of every run.

Multi-Agent Document Intelligence · Python, LangChain, LangGraph, Qdrant, FastAPI, Docker

- 82% extraction accuracy on 150 documents at under 5 s each; 85% top-3 search relevance; 84% less manual data entry.

PUBLICATIONS

- P. B. Kanaparthi, **T. V. S. V. Padamata**. "Lightweight Channel Attention for Efficient CNNs." arXiv:2601.01002, Jan 2026. arxiv.org/abs/2601.01002
- **T. V. S. V. Padamata**. "EDAT-Lite: Predictive Difficulty-Aware Training for Neural Networks."

SKILLS

- **Languages:** Python, TypeScript, Rust, SQL
- **LLMs & Agents:** Claude API, OpenAI API, MCP, multi-agent systems, RAG, LangChain, LangGraph, LLM evaluation
- **Machine Learning:** PyTorch, Hugging Face Transformers, LayoutLM/BERT, CLIP, OCR, scikit-learn, Pandas, NumPy
- **MLOps & Cloud:** AWS (SageMaker, EKS), Docker, Kubernetes, MLflow, Ray Serve, Prometheus, Sentry
- **Data & Search:** PostgreSQL, pgvector, Qdrant, FAISS, Redis, Kafka, Feast, SQLite
- **Backend & Web:** FastAPI, Node.js, React, Next.js, Playwright, pytest, Vitest

CERTIFICATIONS

Edge AI Specialization — NVIDIA · Machine Learning Specialization — Stanford University