Naga Venkata Sai Chennu

Prompt Engineer | LLM Apps & RAG | Python, LangChain| MS CS @ GMU ('26)

nchennu@gmu.edu
www.linkedin.com/in/naga-venkata-sai-chennu | nagavenkatasai.github.io/portfolio/ |
cal.com/nagavenkatasai-c-h24tx0/15min
Fairfax, Virginia, US

I am an aspiring AI engineer who enjoys connecting the dots. I connect ideas from different fields, people from different teams, and applications from different industries. I have strong technical skills. I have an academic background in computer science, machine learning, and prompt engineering.

I like solving business problems with AI solutions and algorithms. I explain complex ideas to non-technical people. I can work across different industries to deliver high-performing AI solutions.

As a Conversational AI Engineer, I designed systems that handled over 1000 queries daily. I achieved 95% user satisfaction. I developed a system that processed 200+ incidents weekly with 78% automation. I also built an AI market research platform. It generated over 3,700 words in professional reports in under 5 seconds. This project used advanced multi-agent prompt work.

One of my best projects was developing a real-time American Sign Language detection system. This project showed my skills in multi-modal prompt engineering. It also showed my commitment to making technology accessible for everyone.

In my graduate studies, I have taken on various leadership roles. I am currently pursuing my Master's degree in Computer Science at George Mason University. I stay updated on industry trends through online courses and community meetups.

I'm graduating in May 2025. I'm interested in full-time AI engineering or machine learning roles. Please feel free to get in touch with me via email at nchennu@gmu.edu.

Education

George Mason University
Jan 2024 - Present

Master's degree Computer Science

KL University

Bachelor of Technology - BTech Computer and Information Sciences and Support Services

Work Experience

Graduate Research Assistant
Aug 2025 - Present

George Mason University | Virginia, United States

- Research focus: Emotional Learning Mechanisms of Large Language Models (LLMs).
- Tech: Python, PyTorch, TensorFlow/Keras, Transformers/Hugging Face, LangChain, LangGraph, FAISS, OpenAI API, RAG, Vector DBs, FastAPI, Streamlit, Docker, Git, SQL, AWS.

Computer Vision & Multi-Modal Prompt Engineer
Apr 2025 - Present

Self-employed

- Developed transformative accessibility technologies using Python programming and data science methodologies.
- Created an ASL detection system that integrates multi-modal prompt engineering with TensorFlow, OpenCV, and CNN architectures.

- Applied advanced image processing techniques and machine learning frameworks to achieve 92% accuracy in gesture recognition through deep learning and neural networks, bridging communication gaps for the hearing impaired.
- Engineered specialized prompts for visual-to-text translation using Python, TensorFlow, and deep learning architectures.
- Combined CNN outputs with advanced NLP processing and neural networks to enhance AI gesture interpretation capabilities.
- Designed adaptive machine learning models that enabled seamless real-time gesture classification through image processing and natural language generation.
- Built end-to-end data science pipelines using Python, TensorFlow, and OpenCV with robust API development practices.
- Implemented computer vision models using CNN and deep learning that process video input streams through sophisticated image processing algorithms.
- Applied machine learning and data science principles to optimize neural networks performance, resulting in faster inference times.
- Deployed solutions using AWS cloud infrastructure and managed code repositories with Git version control systems.
- Utilized Docker containerization for scalable deployment of Python-based machine learning models and API development.
- Leveraged TensorFlow, OpenCV, and neural networks to analyze model performance through comprehensive data science approaches and NLP techniques.
- Facilitated inclusive AI applications by developing accessibility-focused solutions using cutting-edge Python programming, CNN architectures, image processing, neural networks, and API development methodologies powered by TensorFlow, OpenCV, machine learning, deep learning, data science principles, Git version control, AWS deployment, and Docker containerization.

Conversational AI Engineer & RAG Systems Specialist
Self-employed

Jan 2025 - Mar 2025

- Designed and deployed production-ready conversational AI systems with sophisticated prompt routing and context management.
- Built context-aware research chatbot serving 1000+ queries daily with 95% user satisfaction through intelligent tool selection and source-grounded response generation.
- Engineered multi-modal prompt routing system automatically selecting between web search, RAG, and mathematical tools
- Optimized retrieval-augmented generation prompts for domain-specific knowledge bases with FAISS integration
- Implemented session-aware prompt engineering maintaining conversation context across interactions
- Developed comprehensive prompt evaluation frameworks measuring faithfulness, relevance, and grounding

Intelligent Automation & Prompt Engineering Specialist
Self-employed

Nov 2024 - Jan 2025

- Built sophisticated incident triage system using LangGraph with intelligent prompt-driven decision making and confidence scoring.
- Designed stateful workflow processing 200+ incidents weekly with 78% automation rate through advanced prompt engineering and tool integration.
- Created conditional prompt strategies for evidence gathering, analysis, and escalation decisions
- Implemented confidence-based prompt routing with retry logic and human-in-the-loop integration
- Designed prompt templates for technical documentation extraction and root cause analysis
- Developed observability prompts for system monitoring and performance optimization

LLM Operations & Prompt Optimization Lead   Aug 2024 - Nov 2024
Self-employed

- Spearheaded development of AI-powered market research platform generating 3,700+ word professional reports through advanced multi-agent prompt orchestration.
- Achieved sub-5-second report generation using CrewAI framework with GPT-4, serving enterprise clients with 98% accuracy in financial modeling and competitive analysis.
- Designed specialized agent prompts for market sizing, competitive intelligence, and financial projections
- Implemented dynamic prompt adaptation based on industry vertical and market complexity
- Optimized prompt efficiency reducing API costs by 45% while maintaining output quality
- Created prompt templates for TAM/SAM/SOM calculations and strategic recommendation generation

Senior AI Research Engineer & Prompt Architect   Jan 2023 - Apr 2024
Freelance

- implemented advanced prompt strategies for multi-agent systems, achieving 40% improvement in response accuracy through systematic prompt refinement and evaluation frameworks.
- Architected prompt chains for 5 published research papers, optimizing LLM outputs for domain-specific tasks
- Developed novel prompt engineering methodologies for psychological stress detection using social media data
- Created specialized prompts for technical documentation generation, reducing manual effort by 60%
- Implemented few-shot and chain-of-thought prompting strategies for complex AI model evaluations

Core Skills

n8n, Gemini 2.5 Pro, OpenRouter (GrokFast), Prompt Engineering, Conversational AI, Multi-Modal Systems, Python, Machine Learning, Natural Language Processing, TensorFlow, Intelligent Automation, Deep Learning, Computer Vision, Accessibility Technologies, Multi-Agent Orchestration, PyTorch, TensorFlow/Keras, Transformers/Hugging Face, LangChain, LangGraph, FAISS, OpenAI API, RAG, Vector DBs, FastAPI, Streamlit, Docker, Git, SQL, AWS, OpenCV, CNN, LangGraph, CrewAI, GPT-4

Awards

Research Excellence Awards

Best Accessibility Innovation Award

Publications

Comparative Analysis of Psychological Stress Detection: A Study of Artificial Neural Networks and Cat Boost Algorithm

Enhancing Hairfall Prediction: A Comparative Analysis of Individual Algorithms and An Ensemble Method

Assessing the Effectiveness of Artificial Intelligence Techniques in Mitigating Cyber security Risks

A Novel Strategy for Streamlining Land Registration using Ethereum Blockchain

Speech Quality Assessment and Control in Indian Languages