SURYA SG

AI Consultant | Co-Founder & CEO @ SpaceMarvel AI | ML Architect

Bangalore, India | +91 8144418881 | surya@spacemarvel.ai | surya20sg@gmail.com
suryasg.github.io | LinkedIn | 4 US Patents (AI/NLP)

PROFESSIONAL SUMMARY

AI leader and entrepreneur with 10+ years building production-grade NLP, Machine Learning, and Generative AI systems across healthcare, BFSI, conversational AI, and enterprise platforms. Founded SpaceMarvel AI to deliver end-to-end AI consulting from strategy and architecture to deployed, scalable solutions. Proven track record of architecting RAG pipelines, agentic AI workflows, and LLM-powered products for Fortune 500 clients. Holds 4 US patents in conversational AI and has trained 80+ professionals in AI/ML. Seeking co-founder, advisor, and senior consultant opportunities to help organizations accelerate their AI transformation.

CORE COMPETENCIES & AI SKILLS

Generative AI & LLMs: GPT-4o/4-Turbo, Claude 3.5/Opus, Gemini Pro, LLaMA 3, Mistral, Phi-3, Vicuna | Fine-tuning (LoRA, QLoRA, RLHF) | Prompt & Context Engineering | Chain-of-Thought, ReAct, Few-Shot/Zero-Shot Prompting

Agentic AI & RAG: LangChain, LangGraph, LlamaIndex, CrewAI | Agentic RAG Pipelines | Multi-Agent Orchestration | Vector Databases (Pinecone, Weaviate, ChromaDB) | Retrieval-Augmented Generation | Tool-Use & Function Calling

MLOps & Cloud AI: GCP Vertex AI, AWS SageMaker & Bedrock, Azure ML & Azure OpenAI | MLFlow, Kubeflow, Airflow | Docker, Kubernetes | CI/CD Pipelines | Model Monitoring & Retraining | GPU Programming (NVIDIA, AMD)

NLP & Conversational AI: Transformers, spaCy, NLTK | Speech Recognition (ASR, Whisper) | Chatbot Frameworks (RASA, Google CCAI, Dialogflow CX) | Entity Extraction, Intent Classification, Sentiment Analysis | Multi-lingual NLP

Data Engineering & Analytics: Python, PySpark, SQL, R | BigQuery, AWS Glue, Athena | ETL Pipelines | PyTorch, TensorFlow, Keras, Scikit-learn | Time Series, Predictive Modeling

Leadership & Consulting: AI Strategy & Roadmap Development | Technical Due Diligence | Team Building & Mentorship | Client-Facing Solution Architecture | Cross-Functional Collaboration | Product Delivery & R&D

PROFESSIONAL EXPERIENCE

SpaceMarvel AI | *Founder & CEO / AI Consultant*   Aug 2025 – Present
Bangalore, India

- Founded an AI consulting firm delivering production-grade AI systems for organizations across industries undergoing digital transformation.
- Architect end-to-end AI solutions including agentic RAG pipelines, LLM-powered automation, and multi-modal AI products that scale from prototype to production.
- Provide strategic AI consulting: technology assessment, AI readiness audits, build-vs-buy analysis, and implementation roadmaps for C-suite stakeholders.
- Build and deploy custom AI agents using LangChain, LangGraph, and CrewAI for enterprise workflow automation, reducing manual processes by up to 60%.
- Leverage GPU-optimized inference stacks (NVIDIA, AMD) and cloud-native architectures (GCP, AWS, Azure) for cost-efficient, scalable deployments.

Dell India R&D | *Sr. Systems Development Engineer*   Jun 2024 – Aug 2025
Bangalore, India

- Led GenAI solutions validation on Dell hardware, spanning bare-metal to application layer for enterprise AI workloads.
- Developed end-to-end LLM deployment pipelines with GPU programming across NVIDIA, AMD, and Intel stacks, optimizing inference performance by 35%.
- Contributed to Dell's authenticated model repository on HuggingFace, enabling enterprise-grade model distribution and validation.
- Benchmarked and validated large-scale model quantization (GPTQ, AWQ, GGUF) for edge and on-premise deployment scenarios.

Quantiphi, Inc. | *Associate Architect – Machine Learning Engineer*   Dec 2021 – May 2024
Bangalore, India

- Architected the Baioniq GenAI platform: built RAG pipelines, conducted LLM experimentation, and implemented LangChain-based conversational memory for enterprise knowledge workers.
- Led the Qollective.cx NLU module: developed knowledge base retrieval systems, slot-filling engines using Transformer models, and intent conflict detection via semantic similarity.
- Designed and deployed MLOps pipelines (Kubeflow, Vertex AI, MLFlow) for model training, monitoring, and automated retraining across multiple NLP products.
- Managed a team of ML Engineers, conducted code reviews, and defined product roadmaps for the NLP R&D division.
- Built healthcare AI (Risk Capture Platform): clinical NER, disease prediction from clinical notes, ICD code mapping with PySpark NLP, and knowledge graph research.

Vee Technologies | *Associate Data Scientist – NLP Lead*   Mar 2018 – Dec 2021
Salem, India

- Led a team of NLP engineers, translating video, audio, and text data into actionable insights across speech recognition, machine translation, and dialogue systems.
- Built production conversational AI chatbots (retrieval-based and generative) using PyTorch for healthcare and recruitment domains.
- Developed an AI-powered English speech assessment platform using ASR and phonetics-based (IPA) statistical models, deployed for large-scale candidate evaluation.
- Created NLP pipelines for essay grading (topic modeling, cohesion analysis, grammar detection), paraphrase generation, and healthcare claims analytics.
- Designed ETL data pipelines using PySpark, AWS Glue, S3, and Athena for large-scale data processing.

KyzerSoft | *IBM BPM Application Support Engineer*   Sep 2016 – Jun 2017
Mumbai, India

- Managed 28+ IBM BPM banking applications in BFSI domain, resolving issues via SQL and Java, handling runtime incidents and change management.

PATENTS (4 US PATENTS – CONVERSATIONAL AI)

- Method and system for generating intent responses through virtual agents (US20230350929A1)
- Method and system for training virtual agents through fallback analysis (US20230351257A1)
- Method and system for generating conversation flows (US20230351121A1)
- Method and system for training a virtual agent using optimal utterances (US20230342557A1)

EDUCATION

Indian Institute of Technology, Patna | *M.Tech – Big Data & Blockchain* Feb 2023 – Feb 2025

Indian Institute of Management, Trichy | *PG Certificate – Financial Management* Feb 2023 – Mar 2024

Vinayaka Missions University | *B.E. – Information Technology*   Sep 2012 – May 2016
Salem, India ***University Gold Medalist***

ACHIEVEMENTS & IMPACT

- Trained 30+ PhD-level students and 50+ post-graduate students (B-School) in AI, ML, and applied mathematics.
- Conducted 50+ technical interviews for ML Engineer roles and 5+ for Solution Architect positions.
- Received Best Performance Awards at both Quantiphi and Vee Technologies.
- University Gold Medalist – B.E. Information Technology.

CERTIFICATIONS

- Neural Networks and Deep Learning – Coursera (deeplearning.ai)