

DINESH D

☎ +91 7019179783 ✉ dineshsolanki2799@gmail.com [in linkedin.com/in/dinesh-d](https://www.linkedin.com/in/dinesh-d) github.com/dineshsolankii

Summary

AI Specialist with 3+ years of experience building production-grade AI systems. **Published author in Springer (SmartCom 2025)** on generative AI for video summarization. Expertise in LLM orchestration, RAG pipelines, voice AI (VAPI, Twilio), and computer vision (Stable Diffusion). Architected solutions processing 50K+ daily requests using Flask, Kubernetes, HuggingFace. **Led Samsung R&D projects earning 3 Lakh prize and Excellence Award of the year.** Proficient in Python, TensorFlow, PyTorch, cloud infrastructure (GCP, Docker).

Experience

Trilliant Digital Pvt Ltd

April 2025 – Present

AI Specialist

Bengaluru, Karnataka

- Clients required 24/7 voice support at scale. Developed VAPI agents with ElevenLabs TTS, Twilio infrastructure, & RAG using FAISS indexing on PostgreSQL, supporting 900+ calls/day with 94% accuracy for manufacturing clients.
- Lead enrichment delays impacted sales velocity. Developed n8n workflows with OpenRouter Claude API, Airtable integration, & Firebase database, processing 5000+ leads/day and reducing enrichment time from 48 hours to minutes.
- IT clients required scalable AI inference infrastructure. Developed Kubernetes microservices with HuggingFace models, Google AI SDK, and Google Cloud Console with Looker Studio dashboards, supporting 50K+ requests per day.
- Marketing teams required integrated campaign analytics. Developed n8n pipelines with Claude analysis, PostgreSQL warehousing, and Looker Studio reporting on Firebase, aggregating 9 platforms and reducing decision time by 80%.

Samsung R&D Institute India - Bangalore

Oct 2023 – March 2025

Team Lead & POC

Bengaluru, Karnataka

- **Received 3 Lakh prize** for jailbreaking CIVITAI .safetensors checkpoints from SwarmUI dependencies. Reverse-engineered model structure with PyTorch, Diffusers library, custom CLI, allowing platform-agnostic inference.
- Stable Diffusion refiners in SwarmUI lacked flexible deployment. Isolated SDXL weights with Transformers pipelines, VAE loaders, LoRA adapters on standalone Python, preserving 98% image quality parity.
- **Excellence Award:** Developed self-sustaining fine-tuned models without Automatic1111. Designed modular inference engine with safetensors parsing, CLIP encoding, UNet scheduler optimization, boosting generation speed by 40%.
- Video content analysis required manual transcript analysis that took hours. Designed a pipeline using Whisper ASR, FFmpeg preprocessing, Sentence-BERT embeddings, ChromaDB indexing, reducing transcript generation time by 92%.
- Users required instant video insights without viewing entire content. Designed RAG-based chatbot using LangChain, HuggingFace transformers, FAISS vector search with GPT-4, enabling sub-3-second query responses with 89% accuracy.
- Long-form video summarization lacked critical highlight information. Developed NLP pipeline with spaCy NER, TextRank models, abstractive transformer models, producing context-aware summaries capturing 95% critical info.

Projects

AI-Powered Code Generation Platform | *Flask, SocketIO, Ollama, ChromaDB, Docker*

PreProd Corp.

- Built Flask-SocketIO platform with RAG-ChromaDB, generating Python codebases in 3 minutes with 95% accuracy.
- Engineered multi-agent orchestration using Ollama deepseek-coder-v2:16b with semantic similarity search.
- Designed automated requirement analyzer with NLP parsing & AST validation, reducing dependency conflicts by 82%.
- Implemented WebSocket asynchronous process management with ANSI terminal, reducing iteration cycles by 40%.

EEG Signal Analysis for Neuropsychiatric Diagnosis | *Python, TensorFlow, PyTorch, ChromaDB*

iBrains

- Analyzed 500+ EEG datasets using signal processing, PCA, and deep learning for ADHD, MDD, Alzheimer's diagnosis.
- Extracted temporal-spatial features with wavelet transforms and statistical methods for biomarker identification.
- Developed CNN-LSTM hybrid models achieving 72% accuracy for neuropsychiatric prediction using TensorFlow.
- Built diagnostic prototype with automated pattern recognition, reducing manual analysis time by 50% in clinical trials.

Technical Skills

Languages: Python, JavaScript, SQL, Java, C, HTML/CSS

AI/ML Frameworks: TensorFlow, PyTorch, HuggingFace Transformers, LangChain, Diffusers, spaCy, Sentence-BERT

Cloud & Infrastructure: Google Cloud Platform, Kubernetes, Docker, Firebase, Looker Studio

Databases & Vector Stores: PostgreSQL, ChromaDB, FAISS, Airtable

APIs & Communication: Flask, SocketIO, VAPI, Twilio, REST APIs, OpenRouter Claude API, Google AI SDK

Automation & Tools: n8n, Ollama, FFmpeg, Whisper ASR, ElevenLabs TTS, Git/GitHub, VS Code, Linux

Education

Cambridge Institute Of Technology

2021 – 2025

Bachelor of Engineering in Artificial Intelligence & Machine Learning

Bangalore, Karnataka