
Sujit Pranav Reddy

Engineering Head

📍 Hyderabad, India ✉ contact@sujitpranavreddy.dev 📞 +91 7989393646

🌐 sujitpranavreddy.dev 🌐 in/sujitpranav 🌐 iamsujitpranav

Open to Engineering Leadership | AI/ML Platforms | Full-stack Development | SaaS Architecture | LLM Integrations

🔗 PROFESSIONAL SUMMARY

Engineering Head with 12+ years building scalable SaaS platforms, modernizing large-scale monoliths, integrating microservices, and engineering high-availability systems. Ruby on Rails developer with 4+ years hands-on delivering AI-powered applications using Python, FastAPI, TensorFlow, and PyTorch for intelligent search, recommendation systems, document processing, LLM integrations, and automation. Experienced building agentic workflows with Claude MCP, Codex MCP, OpenAI APIs, Anthropic Claude, and the Model Context Protocol (MCP). Track record leading architecture, sprint planning, code review, and mentoring across cross-functional teams, with DevOps delivery on AWS, Docker, Kubernetes, and CI/CD pipelines.

🏠 PROFESSIONAL EXPERIENCE

Engineering Head

Arccaa Analysis & IT Solutions | Oct 2025 - Present

Independent, part-time engagement. Leading engineering across three products - the AI recruitment platform, the AI services layer behind it, and the company's public web presence - architecture, delivery, and code review throughout. Works Claude MCP and Codex MCP into the development loop as repository-scale assistants, validating their output with scripts, tests, builds, and deployment checks.

- **Candidate Intelligence & Semantic Search Platform:** Architected an end-to-end recruitment intelligence system featuring vector-based candidate matching, automated scoring, and RAG-driven talent discovery. Implemented hybrid search pipelines combining full-text filtering with pgvector similarity to rank candidates against complex job specifications. Built end-to-end hiring workflows that transform unstructured applicant records into structured profiles, delivering real-time candidate insights, skill gap analysis, and automated recommendation pipelines across the ATS.
- **ARCCAA AI Engine - Distributed AI Microservice:** Engineered a high-throughput, containerized FastAPI service decoupling AI workloads from the core Rails application. Built a hybrid model routing gateway that dynamically dispatches tasks between local Ollama instances (preserving local 768-dim embeddings for PostgreSQL/pgvector consistency) and hosted OpenAI-compatible APIs for complex reasoning. Designed a centralized prompt registry with strict JSON schema enforcement, deterministic token/temperature configurations, code-fence sanitization, versioned multi-tier Redis caching, and heuristic fallback parsers for resilient document processing.
- **AAITS website & CMS:** Built the company website and CMS from scratch on Next.js 16, React 19, TypeScript, Payload CMS 3, PostgreSQL, Tailwind, and shadcn/ui behind Docker and Caddy - dynamic pages, reusable content blocks, blog posts, categories and tags, media, navigation, forms and submissions, redirects, SEO metadata, previews, and search. Replaced the legacy WordPress site without losing it: archived the original HTML, media, taxonomies, metadata, and redirects, then wrote idempotent importers that convert content to Lexical rich text, preserve internal links, match media, rebuild taxonomies and redirects, and refuse duplicates - public content, URLs, and SEO value carried over intact. Added admin TOTP/2FA, search indexing, cache invalidation, automated tests, and deployment workflows.

Technologies: Python, FastAPI, Next.js 16, Payload CMS, pgvector, Anthropic Claude, MCP, RAG, Ollama, Redis, PostgreSQL, Docker

IT Consultant

Magrofy | Apr 2026 - Present

Independent, part-time engagement. Two products built end to end, schema through UI.

- **AronFX VFX Tracker:** Built from scratch - an operations and billing platform for VFX production spanning clients, projects, bids, shots, work tracking, vendors, invoices, payments, and multi-currency project metrics, modelled around real VFX work such as roto, paint, camera and object tracking, and compositing. Designed a permission-key authorization model enforced in the backend and mirrored in the UI to drive page, column, and field visibility, alongside JWT refresh tokens, RBAC, TOTP, login-activity tracking, rate limiting, encryption, and audit logging; added invoice PDF generation and Excel import for legacy workbooks.
- **Magrofy Social Engine:** SiteCast is a multi-brand agentic content platform for the Magrofy website network. A run moves through Scout, Strategist, Writer, Illustrator, Guardian, Repurposer and Publisher stages with structured JSON outputs, risk tiering, and human approval gates at every exit - an operator picks a site, the model drafts the article and a hero image, a person approves it, and SiteCast publishes to that site's own API and promotes it from that site's own X account. Every website is a Publishing Channel - endpoint, auth type, encrypted credential and a JSON payload template - so adding a site is configuration, not code. Seven are wired up today. Nothing ships unattended: credentials are verified before every publish, scheduled runs are idempotent, and a global kill switch blocks all publishing.

Technologies: Python, FastAPI, SQLAlchemy, PostgreSQL, Celery, Redis, Next.js, Tailwind, JWT / RBAC, Anthropic Claude

Lead Architect

Aequor Information Technologies / Virtusa | Nov 2024 - Apr 2025

Led development on Storable Edge, an all-in-one self-storage operations platform. Drove sprint planning, effort estimation, and roadmap alignment; performed code reviews, pair programming, and mentoring. Contributed to microservice integrations, API design, and performance tuning; used Claude MCP for agentic development and improved deployments with Docker, GitHub Actions, and Jenkins.

Technologies: Rails 7, Python, Ruby 3.2, GraphQL, Sidekiq, AWS, Docker, Grafana

Senior Architect

Emami Frankross | Apr 2023 - May 2024

Led backend modernization of a large-scale e-commerce monolith powering thousands of daily transactions. Built a DevOps stack (AWS + Docker + GitHub Actions), raised release velocity 2x, and cut incident response >40% with Grafana + New Relic. Designed a PyTorch + FastAPI recommendation engine on user/product embeddings, integrated with the Rails platform, and improved search relevance with Algolia.

Technologies: Rails 6.1, Python, PyTorch, FastAPI, Kubernetes, Algolia, New Relic

Senior Software Engineer

Inmar Intelligence - Collective Bias | Feb 2022 - Sep 2022

Designed predictive-analytics models with Python and TensorFlow to estimate influencer engagement, reach, and campaign ROI. Built budget-optimization algorithms on historical + costing data, contributed to architecture and sprint planning, and streamlined microservice deployment with existing Kubernetes pods and Docker images.

Technologies: Rails 5.2, TensorFlow, Python, Kubernetes, GraphQL

Associate Consultant

Global Logic | May 2021 - Feb 2022

Designed architecture for new enterprise HR features used by high-volume global clients. Led sprint planning across cross-functional teams and improved CI/CD via Jenkins + AWS, reducing deployment failures. Supported multi-client delivery for Workday, SuccessFactors, Marriott, and J&J.

Technologies: Ruby 3.0, MySQL, Redis, Sidekiq, AWS, Kubernetes, Jenkins

Senior RoR Developer / Associate Tech Lead

SunRay Construction Solutions | Oct 2019 - Apr 2021

Designed architecture for critical new features, optimized queries, and improved search with Elasticsearch & Kibana. Raised overall performance >10% and built a complete DevOps pipeline (Docker + GitHub Actions + AWS) enabling zero-downtime deployments.

Technologies: Rails 5.0, React, Elasticsearch, Kibana, AWS, MySQL

Project Lead / Application Architect

Denken Solutions - Auto Refund Claims | May 2018 - Sep 2019

Architected and developed the platform end-to-end - design, DB schema, APIs, auth, workflows. Built CI/CD with Jenkins on AWS, introduced Docker-based environments, and implemented CloudWatch-driven alerting to improve uptime.

Technologies: Rails 5.2, PostgreSQL, Redis, Elasticsearch, AWS, Docker, Jenkins

Software Engineer

LotVue India (LotVue LLC, USA) | Jun 2017 - Apr 2018

Restructured the SaaS multi-tenant architecture, cutting maintenance costs 80% by consolidating multiple codebases into a unified Rails Engines platform. Built CI/CD with Jenkins + Docker and integrated New Relic APM for production monitoring.

Technologies: Rails 5.0, MySQL, Rails Engines, AngularJS, AWS, New Relic

Software Engineer

Emami Frankross (earlier tenure) | Apr 2016 - May 2017

Built and maintained the backend for a high-traffic e-commerce application; optimized the product catalog and API performance for Android/iOS mobile apps.

Technologies: Rails 4.0, Ruby 2.0, AWS, Redis, Sidekiq, PostgreSQL

Ruby on Rails Developer

Carveniche Technologies | Jun 2015 - Mar 2016

Installed and configured BigBlueButton for live audio/video, screen share, and virtual classrooms; upgraded legacy Rails apps from 3.2 to 5.2 and integrated conferencing APIs.

Technologies: Rails 3.2->5.2, BigBlueButton, MySQL, Capistrano, AWS

Software Engineer

Kreatio Software Pvt. Ltd | Mar 2014 - Jun 2015

Worked on Ruby on Rails applications - feature development, debugging, and enhancing CMS-based products.

Technologies: Ruby on Rails, CMS

🔗 TECHNICAL SKILLS

Full-stack: Ruby on Rails, Python, Golang, FastAPI, SQLAlchemy, Alembic, Celery, GraphQL, REST, Microservices, Hotwire, React, Next.js, Tailwind, RSpec

AI & ML: Anthropic Claude, MCP, TensorFlow, Pytorch, RAG, LangChain, OpenAI API, Structured outputs, Multi-agent pipelines, Embeddings, Semantic search

Cloud & DevOps: AWS, Kubernetes, Docker, GitHub Actions, Jenkins, Nginx, Capistrano

Data & Storage: PostgreSQL, pgvector, MySQL, Redis, Elasticsearch, Algolia

Security & Auth: RBAC, JWT, Refresh tokens, TOTP 2FA, Rate limiting, Encryption, Audit logging

Observability: New Relic, Grafana, Kibana, CloudWatch

Agentic tooling: Claude MCP, Codex MCP, Cursor, Windsurf, GitHub Copilot

☆ LEADERSHIP FOCUS

Architecture direction, delivery planning, code review, mentoring, platform modernization, and building practical AI capabilities that can be operated in production.

🎓 EDUCATION

B.Tech, Jawaharlal Nehru Technological University, Hyderabad (2008-2012).

🌐 LANGUAGES

English, Telugu, Hindi, Kannada, Tamil

♥ PERSONAL DETAILS

Date of Birth: 6th February 1991

Hobbies: Scuba Diving, Long Drives, Travel, Cricket, Badminton

Marital Status: Married