YASH K. PATEL

Lead AI Engineer | AI Systems Architect

Surat, Gujarat, India | +91 8460320495 | Patelyash2504@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

AI leader with **8+ years** of experience taking AI from architecture to production: **Generative AI, agentic workflows, RAG and secure multi-tenant platforms** integrated with enterprise systems, CRMs, healthcare and third-party data APIs. Currently **Senior and Lead AI Engineer at Pragnakalp Techlabs**, owning AI strategy, architecture and delivery, and joining **Kiran Gems Pvt. Ltd.** as **Senior Head – Enterprise AI & Intelligence** from Dec 2026. Has built AI for the **diamond industry** (rough-to-polished prediction and annotation tooling). Former **CTO**, **Claude Certified Architect - Foundations (CCA-F)**, and **Gold Medalist in M.Sc. Artificial Intelligence (9.40 CGPA)**.

Known for turning real business workflows into **accurate, validated and security-conscious AI systems**, and for building the teams and in-house capability to own them - through hands-on team leadership, client-facing delivery, and 7+ years of training and mentoring professionals and university students in Surat.

LEADERSHIP PROFILE

What the role needs	Evidence from my track record
Diamond-industry domain exposure	Built a **diamond annotation and inspection tool** (v360-style) and an **ML model that predicts polished outcome (color and carat) from rough-stone parameters** such as COLOR, PCUT, PCARAT, PPOL, LCODE and SIZE, directly relevant to manufacturing yield and planning.
Own the AI & data architecture	Architected cloud-native, **multi-tenant AI platforms** (RAG, async enrichment pipelines, serverless and event-driven design) on Azure, AWS and GCP; defined the AI technology roadmap as CTO of Broader AI IT Solution LLP.
Connect AI to enterprise & operational data	Built **data-ingestion and integration pipelines** across provider directories, FHIR/HL7 healthcare systems, **CRMs (HubSpot, Zoho)** via REST/webhooks, and external economic and property data APIs. Converted three proprietary **Excel models into validated Python engines** producing automated 10-year forecasts.
Reliable analysis & decision support	Strong analytical base: time-series forecasting, ML/DL model comparison (RMSE, R^2), embeddings and semantic search. Designs for trust: **human-in-the-loop verification**, physician approval steps, and clarification flows that stop the agent from guessing or fabricating.
Validate accuracy & security; challenge partners	Led client-facing AI engagements end to end as AI Consultant. Implemented **RBAC, audit trails, encryption in transit/at rest, consent-aware access and PHI-safe logging** on healthcare-grade platforms, and led delivery of AI solutions from proof of concept to production.
Build internal AI capability	Built and led AI teams; mentor engineers; **Visiting Faculty** at Auro University and SCET; trainer and speaker at universities and corporates, with custom AI curricula for organizations.

PROFESSIONAL EXPERIENCE

Senior Head – Enterprise AI & Intelligence
Dec 2026 - Present

Kiran Gems Pvt. Ltd.

- Own **enterprise AI implementation, integration and operational deployment**, delivering agreed projects against **measurable business objectives**.
- Progressively build a **secure Company Intelligence System** connected to the company's ERP, operational data and business knowledge - capable of reliable analysis, root-cause investigation, anomaly and opportunity discovery, and **management decision support**.
- Coordinate with external technology partners and work with ERP, IT and operational teams to integrate AI into day-to-day operations.
- Establish **testing, validation, monitoring and knowledge-transfer processes**; maintain security and reliability standards.
- Help build the **internal AI implementation team**.

Senior and Lead AI Engineer
Dec 2025 - Nov 2026

Pragnakalp Techlabs Pvt. Ltd. · Full-time

- Lead **AI strategy and end-to-end delivery**, from concept to production, bridging research, engineering and business goals; mentor teams and guide AI adoption across projects.
- Design and build production-grade applications with **Claude Code, the Claude Agent SDK, Claude API and MCP**; develop multi-agent systems for orchestration, reasoning and tool execution using **Hermes Agent, LangChain and LangGraph**.
- Architect **multi-tenant SaaS and cloud-native AI solutions** with RAG, asynchronous enrichment workflows, serverless and event-driven processing.
- Build scalable **FastAPI** services integrated with AI pipelines and real-time inference; own the full lifecycle: data ingestion, embedding pipelines, retrieval optimization, response generation and automation workflows.

Senior Artificial Intelligence Consultant
Aug 2024 - Aug 2025

Inkey IT Solutions Pvt. Ltd. · Part-time

- Built, trained and led a high-performing AI team; managed multiple client engagements delivering **GenAI chatbots, document-intelligence systems and intelligent automation**.
- Translated business challenges into AI architectures with cross-functional stakeholders; optimized model performance, latency and cost for production deployments.

CTO | Senior Data Scientist
May 2022 - Jul 2024

Broader AI IT Solution LLP · Full-time

- Defined and executed the **AI technology roadmap and innovation strategy**; led end-to-end ML and AI solution development from data engineering to deployment.
- Led research in Generative AI, Agentic AI and predictive modelling; integrated AI into products and business processes while ensuring solutions were scalable, ethical and aligned with business goals.
- Mentored data-science teams and established engineering and experimentation best practices.

Machine Learning Engineer
Nov 2021 - Apr 2022

Rejoicehub Solutions · Full-time

- Developed ML models for predictive analytics and automation; built data pipelines and feature engineering for structured and unstructured data; deep learning, NLP and computer-vision models; testing, validation and performance tuning for production.

Full Stack Developer
2018 - 2021

ItSoulInfotech

- Built full-stack web applications and Python (Flask/FastAPI) REST APIs, and delivered early ML and NLP solutions (prediction, classification, chatbots, information extraction) into production.

ACADEMIC, TRAINING & ENTREPRENEURIAL ROLES

AI Trainer & Instructor
Oct 2024 - Present

Part-time

- Designed and delivered advanced programs in **Agentic AI, Generative AI and LLMs, Machine Learning and AI product development**, guiding learners from idea to deployment.

Visiting Assistant Professor
Jul 2022 - Present

Auro University, Surat · Part-time | Sarvajanik College of Engineering & Technology (SCET) · Part-time

- Teach ML, deep learning, transformers, LLMs, Generative AI and Agentic AI; design curricula that connect theory to real-world systems; mentor student projects and research with emphasis on ethical AI.

Corporate Trainer | AI Expert Speaker | Mentor
2018 - Present

Freelance

- Delivered workshops, seminars, expert talks and custom corporate training in AI, Data Science, ML, Generative AI and Agentic AI; judge and panelist at university and industry forums.

FLAGSHIP AI PROJECTS

1. Diamond Rough-to-Polished Prediction & Annotation Platform

Stack: Machine learning, computer vision / annotation tooling, Python

- Built a **diamond annotation tool in the style of v360.in** for marking and reviewing stone characteristics.
- Developed an ML model that takes rough-stone parameters (COLOR, PCUT, PCARAT, PPOL, LCODE and SIZE and other key attributes) and **predicts the final polished color grade and carat weight**, supporting better planning and valuation decisions.

2. AI-Powered Healthcare Provider Discovery & Referral Platform

Stack: LLMs, RAG, Django REST, React/Next.js, FHIR/HL7, OCR, Docker, CI/CD

- Conversational provider search (LLM + RAG) by specialty, symptoms, location, insurance and availability; intelligent **provider-ranking logic** and ingestion pipelines consolidating directories, insurance and appointment metadata.
- Secure referral workflow with document attachments, status tracking and notifications; **FHIR/HL7-compatible integrations** with third-party healthcare and scheduling systems; OCR-based document processing.
- Healthcare-grade security: role-based portals, audit trails, encryption in transit and at rest, consent-aware access, PHI-safe logging.

3. AI-Powered Event Lead Capture & Engagement Platform

Stack: Django REST, PostgreSQL + pgvector, Celery, Redis, Azure AI Vision OCR, Azure OpenAI, Next.js, GitHub Actions

- Converts business cards and event interactions into enriched, qualified leads: multilingual OCR, **human-in-the-loop verification**, and asynchronous LLM enrichment and qualification pipelines.
- Personalized outreach via WhatsApp Business API, email and Telegram; **CRM sync with HubSpot and Zoho** (REST + webhooks); secure **multi-tenant data isolation**, offline-capable capture, real-time dashboards.

4. AI-Powered Hotel Valuation SaaS Platform

Stack: Python, Pydantic, Django REST, PostgreSQL, Next.js, OpenAI, Stripe, Jinja2, Nginx

- Converted three proprietary Excel valuation models (ADR, Expense, Equity) into **validated Python engines**, preserving model logic and automating 10-year financial forecasts.
- Enriched valuations through TripAdvisor, Google Places, BEA, FRED and OpenAI APIs (property intelligence, macro indicators, AI investment narratives); generated investor-ready PDF reports; implemented Stripe subscriptions and credit-based add-ons.

5. Healthcare AI Medical Dictation & Clinical Documentation Platform
Stack: Azure AI Speech, Azure OpenAI, Azure AI Language, Django, React, PostgreSQL, HL7 FHIR

- Near-real-time medical transcription with generated SOAP notes, clinical summaries and treatment plans; **medical entity extraction** and FHIR resources (Patient, Observation, Condition, Encounter, Medication Request, Care Plan) with ICD-10 / SNOMED CT suggestions.
- Physician review and approval, RBAC, audit logging, encryption and HIPAA-ready cloud storage.

6. WhatsApp-Based Real-Estate Property Search Agent
Stack: Hermes Agent skill runtime, Python, Filesystem MCP, Azure MCP, Azure Speech STT

- LLM-driven intent and parameter extraction from natural-language messages; **controlled clarification flow** (max. two follow-ups) that prevents the agent from guessing cities or fabricating prices.
- Live multi-source discovery with automatic fallback, Google Maps link generation and concise WhatsApp results; modular, Git-tracked skill architecture with voice-note transcription.

Other notable projects: AI Resume Analysis System (Llama 2, OCR, RAG, Django); German-to-English Audio Translation & Information Extraction (fine-tuned Whisper, Hugging Face, RAG); Digital Marketing RAG Chatbot (150+ eBooks via OCR, vector DB, quantized LLMs); Time-series Stock Forecasting (six models compared on RMSE/R^2).

TECHNICAL SKILLS

GenAI & Agentic AI	Claude (Code, Agent SDK, API, MCP), OpenAI, Gemini, Llama; RAG design & evaluation; prompt engineering (CoT, ReAct, few-shot); multi-agent orchestration (Hermes Agent, LangChain, LangGraph, AWS Bedrock); memory, context management, tool calling; voice agents; n8n automation
Machine Learning & Data	ML, deep learning, NLP, computer vision (YOLO, Mask R-CNN, OCR/Document AI), time series, recommender systems; TensorFlow, Keras, Pytorch, scikit-learn, OpenCV; ETL/ELT, embeddings, hybrid/semantic/vector search (Pinecone, Weaviate, FAISS, Milvus, pgvector); MLflow, model monitoring, latency & cost optimization
Backend & APIs	Python, FastAPI, Django, Flask; REST, WebSockets, async Python; microservices; secure API design (RBAC, auth); Celery/Redis; PostgreSQL, MySQL, SQL Server; pytest, unittest
Cloud & DevOps	Azure & GCP (AI services, App Services, IAM); AWS (Lambda, API Gateway, Step Functions, Event Bridge, S3, DynamoDB, ECS/Fargate, CloudWatch, Secrets Manager); Docker, CI/CD (GitHub Actions, Code Pipeline)
Frontend & Visualization	React, Next.js, Tailwind CSS, HTML/CSS/JS; Tableau, Plotly, Matplotlib, Seaborn
Foundations	Probability, statistics, linear algebra, calculus; R, SQL, C, C++

CERTIFICATIONS & EDUCATION

- **Claude Certified Architect - Foundations (CCA-F)** - Anthropic, 2026. Validates design, orchestration and scaling of production Claude systems: agentic architecture, tool & MCP integration, Claude Code workflows, prompt engineering and context management [Verify]
- **Machine Learning Specialization** - Stanford University, Coursera [Verify]
- **Deep Learning Specialization** - DeepLearning.AI, Coursera [Verify]
- **TensorFlow Developer Specialization** - DeepLearning.AI, Coursera [Verify]
- **Machine Learning with TensorFlow on Google Cloud Platform Specialization** - Google Cloud, Coursera [Verify]
- **Data Visualization with Tableau Specialization** - UC Davis, Coursera [Verify]
- **Practical Data Science with MATLAB Specialization** - MathWorks, Coursera [Verify]
- **Computer Vision & Image Processing** - NPTEL [Verify]

M.Sc. in Artificial Intelligence - Auro University, Surat | 9.40 CGPA, Gold Medalist (Distinction)

B.Sc. in Information Technology - Veer Narmad South Gujarat University, Surat | 7.72 CGPA (Distinction)

SPEAKING, THOUGHT LEADERSHIP & RESEARCH

- **Agentic AI: Beyond Chatbots to Autonomous Agents** - seminar, Auro University, Surat
- **Modern AI: Use-Cases, Challenges, Future Pathways & Careers** - seminar, SSAIT Diploma College, Surat
- **Judge & Expert Talk on the Future of AI** - CB Patel College, Surat
- **Careers in Generative AI: Skills, Tools and Opportunities** - seminar, SASSIT Engineering College
- **The Future of Generative AI in Education and Research** - faculty and AI-expert panel discussion
- **Research:** Automation in vegetable cultivation [Paper]; Crop recommendation using LightGBM [Paper]