PARTH KOMALWAD

Senior AI Engineer | Pune, India

pkomalwad@gmail.com | linkedin.com/in/parth-komalwad-b482a71b0

SUMMARY

Senior AI Engineer with 3+ years building and shipping production Generative AI systems for enterprise clients across manufacturing, automotive, sustainability, and real estate. Specialises in multi-agent orchestration, Retrieval-Augmented Generation (RAG), NL2SQL, and LLM fine-tuning, with production deployments on Azure AI Foundry and Microsoft Fabric. Designed a 12-agent orchestration system that cut workflow creation from 4-8 hours to 2 minutes and reduced error rates from 30-40% to under 10%. Also builds governed data platforms for AI consumption, including a medallion warehouse over 1.5M rows of manufacturing data. Anthropic-certified in Model Context Protocol (MCP).

TECHNICAL SKILLS

Generative AI and LLM: Agentic AI, AI Agents, Multi-Agent Orchestration, RAG (Retrieval-Augmented Generation), GraphRAG, Prompt Engineering, LLM Fine-tuning (Llama 3), Model Context Protocol (MCP), NL2SQL, Semantic Search, Embeddings, LangChain, Wren AI, Ollama, Anthropic and OpenAI APIs

Data and Vector: Vector Databases, pgvector, Vector Indexing, Similarity Search, PostgreSQL, DuckDB, Neo4j (Certified Professional), MySQL, SQLite, Parquet, Knowledge Graphs, Medallion Architecture, Semantic Layers, Data Modelling, ETL Pipelines, Data Reconciliation

Azure and Cloud: Azure AI Foundry, Azure AI Document Intelligence, Azure Functions, Azure Blob Storage, Azure Key Vault, Microsoft Fabric, OneLake, Microsoft Graph API, MSAL, Row-Level Security

Backend and MLOps: Python, FastAPI, Node.js, REST APIs, GraphQL, Hasura, Docker, CI/CD, nginx, AWS EC2, Air-Gapped Deployment, Structured Logging, Observability, Linux, tmux

Full-Stack Integration: React 19, TypeScript, WebAssembly, ECharts, Zustand, Tailwind CSS, ElectronJS

Other: PaddleOCR, Keycloak SSO, SAP Integration, Git

PROFESSIONAL EXPERIENCE

Equations Work IT Services Pvt. Ltd. | Pune, India Mar 2023 - Present

Senior Software Engineer *Oct 2025 - Present*

Software Engineer *Mar 2023 - Oct 2025*

Client Engagement: Excellon | Multi-Agent Orchestration and AI Analytics Platform (2025)

- Led architecture and development of two production AI systems for one of India's largest Dealer Management System networks, owning design decisions end to end.
- Built a 12-agent multi-agent orchestration system generating production-ready workflow action JSONs (500-2000 lines) and UI form configurations from natural language, cutting workflow creation from 4-8 hours to 2 minutes and reducing error rates from 30-40% to under 10%.
- Designed 6 Workflow Domain agents (Schema Analyzer, Task Selection, Task Chain Architect, Path Expression Generator, Payload Builder, Workflow Assembler) and 5 UI Domain agents spanning 118+ task types, with agent-to-agent handoff and structured output validation.
- Delivered an AI analytics platform on Microsoft Fabric and Azure AI Foundry using Medallion architecture (Bronze, Silver, Gold on OneLake), integrating 23+ source tables across MySQL, SAP, and REST APIs.
- Implemented NL2SQL through Fabric Data Agent with Row-Level Security across a 6-tier dealer hierarchy, serving 65,000+ job card records per dealer per month.

Client Engagement: Jabil | AI Manufacturing Defect Analytics Platform (2025)

- Designed and delivered an end-to-end medallion data platform (bronze, silver, gold) turning an 865 MB, 1.53M-row MES test and defect export into a governed PostgreSQL warehouse, orchestrated in Docker with one-command bring-up and a ~3 minute full rebuild.
- Diagnosed an undisclosed table join fan-out in the source export that inflated every test event ~3.5x and silently corrupted all downstream KPIs; rebuilt the grain with windowed de-duplication to recover 433K true test events and 139K defects, validated by exact silver-to-gold reconciliation.
- Achieved 10.5x row compression (1.53M to 145K) with zero dimensional loss, cutting agent query cost and latency while preserving full drill-down.
- Engineered a governed semantic layer for LLM and NL2SQL access with a gold-only read-only database role, encoding domain footguns (cross-grain aggregation, non-defect bookkeeping rows, volume-versus-rate station ranking) as hard agent constraints to prevent plausible-but-wrong analytics answers.
- Handled late-arriving facts where 53% of defect analyses fell outside the test window, using idempotent per-day restatement of affected test dates instead of naive incremental loads.

- Built a 9-view analytics dashboard (React 19, TypeScript, Tailwind, ECharts, Zustand) with natural-language-to-SQL, 6M Fishbone root-cause analysis, unit-level traceability, and Excel export, served as a zero-backend static app querying Parquet client-side via WebAssembly.
- Containerised the full stack for air-gapped deployment with offline image tarballs, nginx, and EC2 deploy scripts.

Client Engagement: Bray | Enterprise RAG Platform (2024 - 2025)

- Designed and shipped a production Retrieval-Augmented Generation platform enabling semantic search across enterprise technical documentation for a global industrial valve manufacturer.
- Built vector indexing and similarity search enabling natural language querying over large unstructured datasets, removing the SQL knowledge barrier for business users.
- Automated document ingestion with Azure Blob Storage and Azure Functions, including custom structured table extraction for high-precision parsing of complex technical content.
- Developed FastAPI services with retry logic, structured logging, and full observability from ingestion through query response.

Client Engagement: SmartCap | AI Investor Intelligence and Automated Reporting (2024)

- Built AI agents that interpret complex Excel financial models and surface structured investor insights without manual analyst intervention.
- Developed FastAPI ingestion APIs triggered by Azure Blob Storage workflows, with schema-driven ETL pipelines and ranking logic for intelligent property scoring.
- Shipped a GenAI pipeline generating layout-aware, investor-ready PowerPoint reports directly from structured data, eliminating manual report creation.
- Secured the backend with MSAL authentication and Azure Key Vault; containerised with Docker and deployed through automated CI/CD pipelines.

Client Engagement: Optimizer 3.0 and Neural Fabric | LLM Fine-tuning and GraphRAG (2023 - 2024)

- Fine-tuned Llama 3 on domain-specific sustainability and cement project data using Neural Fabric infrastructure for model training and API deployment.
- Built a GraphRAG framework over Neo4j graph data, enabling structured retrieval across complex relational datasets for AI-generated decision support.
- Deployed to production for cement production optimisation, replacing manual expert analysis with AI-assisted recommendations.

Client Engagement: Jabil | JDoc Multilingual Document Converter (2023 - 2024)

- Technical Architect for a multilingual conversion system translating Visual Aids and Work Instructions from English into Indian regional languages for a global electronics manufacturer.
- Integrated Azure AI Document Intelligence with the JDoc API to preserve technical terminology, formatting, and document structure at scale; deployed to 7,000+ concurrent users across Jabil Pune manufacturing operations on the Jabil Azure tenant.

PROJECTS

Sable - Python, Linux, LLM APIs, tmux, SQLite, bubblewrap In Progress

- Open-source Linux login shell that replaces /bin/bash: routes each line to bash or to an LLM orchestrator (Anthropic, OpenAI, Ollama), previews every generated command, and keeps scp/rsync/git untouched via an SSH bypass.
- Multi-turn agent runtime: an orchestrator that plans, runs, or delegates to sub-agents in their own tmux windows, each isolated in a bubblewrap kernel sandbox (workspace read-write, host read-only), with pause/resume/attach and crash reconciliation.
- Self-learning skill system: clusters command history from the audit log, has the LLM crystallise repeated procedures into reusable skill files, and ranks them by a confidence score updated from outcomes.
- Safety layer: destructive-pattern blocklist with typed confirmation, entropy-based secret redaction, per-session and daily token budgets, append-only audit log; PTY execution for fully interactive programs (vim, htop, nested ssh).
- Engineering: 85-feature roadmap with test gates per phase, mock-LLM mode for zero-cost end-to-end tests, router accuracy corpus (500+ labelled lines, 99% bash recall), GitHub Actions CI, Docker playground and devcontainer.

ACHIEVEMENTS

Second Place, Metron Security Code Jam 2025 Jan 2025

- Placed 2nd of 30+ developers across Pune in a 6-hour timed challenge, building an Elastic ML and real-time security dashboard from scratch.

CERTIFICATIONS

- Model Context Protocol: Advanced Topics, **Anthropic** (Aug 2025)
- Introduction to Model Context Protocol, **Anthropic** (Aug 2025)
- Certified **Neo4j** Professional

EDUCATION

Genba Sopanrao Moze College of Engineering Pune, India
Bachelor of Engineering, Computer Engineering *2019 - 2023*