

Yash Kalwar

Agentic AI Developer

+91 9057357954 | New Delhi | yashkalwar.dev@gmail.com | [Linkedin](#) | [Github](#)

Profile

AI/ML Engineer with experience across enterprises (Amdocs), applied ML (Brainywood), and AI consulting (Cogniwide UK). Skilled in building production-ready ML pipelines, RAG/multi-agent architectures, and scalable backend systems. Delivered AI solutions that optimized system performance by ~30%, cut LLM hallucinations, and won hackathons for agentic architectures.

Work Experience

1) Machine Learning Engineer – AI Solutions at Cogniwide UK (July 2025 – present)

- Partnered with stakeholders across 3 enterprise clients to define AI problem statements and translate into scalable ML pipelines, supporting use cases handling 1M+ records daily.
- Evaluated and benchmarked 5+ LLMs and vector DBs, reducing infrastructure costs by ~20% through optimal architecture selection.
- Architected and deployed production-ready AI models and services, covering data preprocessing, training, evaluation, and client integration.
- Piloted and scaled proof-of-concepts, scaling successful designs into maintainable systems adopted in enterprise environments.

2) ML engineer at Brainywood (Apr 2024 – Sept 2024)

- Built and deployed regression, classification, and clustering models on student performance datasets, identifying key drivers of learning outcomes and recall.
- Applied behavioral analytics and A/B testing to optimize recommendation engines, boosting student retention by 12% and lowering task completion time by 15%
- Formulated scalable pipelines and dashboards serving 5k+ students, cutting manual reporting by 20+ hours per week for product and pedagogy teams.

3) Software Developer at Amdocs (Project: AT&T- USA) (Jul 2022 – Apr 2024)

- Migrated legacy COBOL codebase to C within 3 months, improving system maintainability and execution speed while Minimized technical debt.
- Engineered rating and discounting modules for 10M+ AT&T subscribers, Mitigated billing errors by 15% and improving throughput.
- Authored test strategies that reduced post-deployment defects by ~25%, ensuring stable releases for a system processing millions of transactions daily.
- Tech stack: COBOL, C, Python, Unix, SQL, Shell scripting, Git, Postman.

Projects

1) Mem1 – Next-Gen Memory Layer for Multi-Agent Systems

- Engineered a layered memory architecture (short-term, semantic, episodic, credential vaults) with deterministic + LLM-based routers for low-latency, high-accuracy recall.
- Integrated hybrid RAG pipelines (vector, symbolic, metadata filters) using Sentence-Transformers + FAISS/ChromaDB, lessen hallucinations by ~30% and enabling personalized, context-aware agent orchestration.

Skills – Python, FastAPI, LangChain, LangGraph, Sentence-Transformers, FAISS/ChromaDB, GitHub Actions, Advanced RAG Architectures, Multi-Agent Orchestration, Vector Databases, Context-Aware AI

2) Avatar Podcast – AI Native Content Creator

- Developed an AI-first media engine where a virtual avatar autonomously generated scripts and narrated podcasts on demand.
- Implemented multimodal LLM pipelines for instant script generation and neural speech synthesis, delivering lifelike narration.
- Orchestrated an end-to-end autonomous content production stack showcasing synthetic media and generative workflows and Generated 50+ autonomous podcast episodes.

Skills – Python, Generative AI, Speech Synthesis (TTS), Natural Language Generation, Multimodal LLMs, Content Automation, End-to-End AI Product Development

Awards

1) ORBYTT - A Multiagent Life-Management tool

Winner – Imperial AI Hackathon 2025: Devised ORBYTT, a multi-agent life management tool with RAG and context-aware scheduling/decision-support agents. [Link](#)

2) Disease Diagnosis from Symptoms

Winner – ScoeThon21: Developed ML model for disease diagnosis from symptoms using decision trees & random forests on healthcare datasets.

3) Best Use of LangFlow

LangFlow + ElevenLabs + Mistral Hackathon: Created “Life Orchestrator,” a workflow of intelligent agents integrating Twilio + ElevenLabs via LangFlow. [Link](#)

Education

1) MS in Artificial Intelligence (sept 2024- sept 2025)

King’s College London, United Kingdom | Tentative Distinction

2) B.Tech in computer Science (Jul 2018 - Aug 2022)

Sinhgad College of Engineering, SPPU, Pune | 8.71 CGPA