



Pawan Kumar

APPLIED AI ARCHITECTURE &
TECHNICAL LEADERSHIP

📍 Cologne, Germany

📞 +49 176 8020 0719

✉️ contact@pawankumar.de

🌐 pawankumar.de

🌐 linkedin.com/in/pawan-kumar94

<> github.com/pawankumar94

10+

YEARS IN SOFTWARE

5+

YEARS ML & GENAI
IN PRODUCTION

20+

ENGINEERS GUIDED

~45%

LOWER LLM COST PER TASK

≡ PROFESSIONAL SUMMARY

Applied AI technical lead with **10+ years** in software and **5+ years** delivering production ML systems, spanning cloud platforms and governed AI agents. At Trimble Transportation, I guide architecture, technical direction, and shared engineering standards for **20+ engineers** across internal and customer-facing AI products. Previously led ML and MLOps across **5+ business units** at DISH. My work connects product goals with architecture decisions around reliability, governance, and cost, delivering **~45%** lower LLM cost per task and up to **60%** lower response latency at Trimble.

📁 EXPERIENCE

● Lead AI Engineer, Applied AI

Oct 2025 – Present

Trimble Transportation, Germany · Technical lead, Arc Agent

- Provide technical direction for Applied AI products across internal and customer-facing work, with architecture guidance and mentorship for **20+ engineers**, from junior to senior.
- Architect **Arc Agent**, a single governed AI agent with a catalog of skills for logistics back-office workflows, launched globally in August 2026. The design includes planner and sub-agent orchestration, sandboxed code execution, human-in-the-loop approvals, and enterprise SSO; it passed Trimble's security review.
- Reduced LLM cost per task by **~45%** and response latency by up to **60%** through multi-model routing and prompt caching across Anthropic, OpenAI and Gemini.
- Established the team's AI-assisted engineering practices across design, code review, debugging and verification; continue mentoring junior and mid-level engineers; and spoke at **Trimble AI Summit 2026**.
- Built cost and adoption analytics that gave leadership real-time visibility into AI spend and usage, informing pricing and go-to-market decisions.

○ Technical Co-Founder & Head of AI

Jul 2024 – Oct 2025

Arelis AI, Remote · Part-time, self-employed venture alongside DISH

- Co-founded an early-stage **GenAI startup** as Head of AI, owning product vision, architecture and delivery; recruited and mentored founding engineers.
- Built the platform around **AI governance**: ISO 42001, EU AI Act readiness, auditability, and responsible AI adoption.
- Set engineering, product and go-to-market direction under startup budget constraints; exhibited at **Web Summit Lisbon 2024**.

○ Senior Machine Learning Engineer

Mar 2022 – Sep 2025

DISH Digital Solutions GmbH, Germany · METRO AG subsidiary, 330,000+ restaurateurs in 16 countries

- Directed the technical strategy for **5+ end-to-end AI initiatives**, including agentic orchestration and multimodal LLM/ML pipelines; GenAI programmes cut manual processing and annotation workload by **40%**.
- Built a multilingual synthetic data engine generating **1M+ samples**, raising model accuracy in underrepresented and low-resource language domains.
- Standardised MLOps across business units (enterprise-grade CI/CD, observability, governance), shortening model iteration cycles by over **50%**.
- Architected a secure GCP ML platform using Terraform, Kubeflow, and Airflow, delivering **99.9% uptime** and **30% lower cloud costs**.

EARLIER EXPERIENCE

Apr 2021 – Feb 2022	ML Research Assistant <i>Leibniz University Hannover, L3S, Germany</i> Improved Kubernetes cluster resource utilization 40% with an RL-based scheduling algorithm for the EU-funded BRAINE project, researching advanced RL for real-time applications.
Nov 2015 – Aug 2018	Software Engineer <i>Infosys Ltd., India</i> Sustained 99.9% availability for enterprise Java applications (WebSphere, WebLogic), managing deployments and production rollouts.

TECH STACK

Agentic AI Platforms & Safety	LangChain, LangGraph, Deep Agents, LlamaIndex, CrewAI, Google ADK, MCP, Multi-Agent, Agent Skills & Plugins, Context Engineering, OpenWiki, RAG, GraphRAG, DSPy, Anthropic, OpenAI, Gemini, Prompt Caching, LangSmith, LangSmith LLM Gateway, Daytona Sandboxes, GCP Model Armor, Lakera Guardrails, Human-in-the-loop
Governance & Compliance	ISO 42001, EU AI Act, Responsible AI, Enterprise SSO
LLMOps & Evals	LLM Observability, Ragas, MLflow, Databricks, Datadog, CI/CD
Cloud & Infra	GCP Vertex AI, AWS Bedrock, Azure AI, Kubernetes, Docker, Terraform, BigQuery, Spark
Languages & Frameworks	Python, FastAPI, Flask, Streamlit, PyTorch, Hugging Face, SQL
Product & UX	Figma, Motion Design, AI-assisted UI Prototyping

EDUCATION

M.Sc. Digital Engineering Otto von Guericke University Magdeburg Focus: ML, Deep Learning, RL 2018 – 2022
Bachelor of Computer Applications GGSIPU, Delhi, India Focus: OOP, Web Dev, DBMS 2012 – 2015

LANGUAGES

English	C1 · Professional
German	B1 · Intermediate

REFERENCES

Dr. Javad Chamanara Project Lead · L3S, Leibniz University Hannover <i>"Exceptional technical depth and initiative."</i>

CERTIFICATIONS

Professional ML Engineer Google Cloud · 2024
Terraform Associate (003) HashiCorp · 2025
Databricks Platform Architect (AWS) Databricks · 2025

SPEAKING & VISIBILITY

Trimble AI Summit 2026 Speaker (internal)
Web Summit Lisbon 2024 Exhibitor · Arelis AI

Ramon Marrero Senior Head of Data/ ML Operations · DISH <i>"A force multiplier for enterprise-scale AI initiatives."</i>
