

# Amit Kumar

[amitkumar0902.github.io](https://github.com/amitkumar0902) | [ak6.amitkumar@gmail.com](mailto:ak6.amitkumar@gmail.com) | +91-8989691459 | [LinkedIn](#) | [GitHub](#)

## EXPERIENCE

### Lead AI Engineer

May 2023 — Present

*e42.ai — Bengaluru, Karnataka*

- Architecting production-scale generative AI platforms: multi-agent orchestration (LangGraph), distributed LLM inference engines with speculative decoding and INT4/INT8 quantization, and enterprise RAG systems.
- Designed Kubernetes-native AI platform toolkit with unified request routing, zero-trust service mesh, and automated GPU workload scaling for multi-tenant inference clusters on GKE.
- Leading and mentoring a team of engineers across the full GenAI stack — from model fine-tuning and evaluation to production deployment, monitoring, and scaling.

### NLP Engineer

Nov 2021 — May 2023

*Circlebase (Intelligent Health Outcomes) — Bengaluru, Karnataka*

- Developed NLP pipelines for patient entity extraction, social determinants of health, adverse drug events, and clinical trial QA pair generation using transformer-based models.
- Built and deployed a clinical trial chatbot with NER-based adverse event detection, Haystack QA retrieval, LLM-powered dashboard, and Dockerized deployment pipeline.
- Forecasted drug consumption using ensemble time-series methods (ARIMA, Holt-Winters, XGBoost), achieving MAPE below 20% across multiple drug categories.

### Machine Learning Engineer

Sept 2021 — Nov 2021

*Omdena — Internet & Jurisdiction Policy Network (Remote)*

- Built knowledge graph pipelines and triplet extraction algorithms for policy document analysis using SpaCy, Stanford CoreNLP, and NLTK.
- Developed the Datasphere framework for large-scale textual data analysis and visualization, enabling policymakers to navigate complex internet governance issues.

## SELECTED PROJECTS

- Abnormal Event Detection in Video** — Built a spatial-temporal anomaly detection model for surveillance video streams, achieving **95% precision** on the UCSD Anomaly Detection Dataset using CNN-based feature extraction and unsupervised learning.
- Frontier AI Research Presentations** — Delivered 6+ technical deep-dives on cutting-edge papers: StreamingLLM, AlphaGeometry, Infini-Attention, GaLore, WARP, and LARS — covering infinite-context inference, memory-efficient training, and RLHF alignment.

## EDUCATION

### M.Tech in Machine Learning & Computing

Aug 2020 — May 2022

*Indian Institute of Space Science and Technology (IIST) — Thiruvananthapuram, Kerala | CGPA: 8.29/10*

### B.E. in Information Technology

Aug 2016 — July 2020

*Jabalpur Engineering College — Jabalpur, Madhya Pradesh | CGPA: 7.44/10*

## TECHNICAL SKILLS

**Languages:** Python, C++, C, SQL, Bash

**ML / AI:** PyTorch, TensorFlow, Hugging Face Transformers, LangChain, LangGraph, vLLM, GGML, RAG, Multi-Agent Systems

**Infrastructure:** Docker, Kubernetes (GKE), GCP, Git, CI/CD, Service Meshes, LiteLLM

**Data & NLP:** PostgreSQL, SQLite, Redis, Elasticsearch, Vector Databases, SpaCy, Haystack, NLTK

## ACHIEVEMENTS

- ACM ICPC 2018 Regionalist — Qualified for Kolkata-Kanpur site contest at UIET, CSJM University, Kanpur
- CodeChef Rating 1883 (handle: amit\_9) — Competitive programming in algorithms and data structures
- Organized "So You Think You Can Code" — Competitive programming contest on HackerEarth
- Gold Medal in Chess — Avahan 2018, Jabalpur Engineering College