# Manish Kumar

Senior AI Engineer / Tech Lead

✉ kmanishofficial@gmail.com    in linkedin.com/in/master-fury    ⌂ github.com/master-fury

## Summary

Senior AI / ML Engineer with **6+ years of experience** delivering **enterprise GenAI and data platforms**. Expertise in LLM training and fine-tuning, scalable inference and secure cloud deployment. Led cross-functional teams of 10 engineers and drove architectural decisions. Delivered **production AI systems across global markets** including Japan, Canada and US.

## Technical Skills

**Backend**
Python, FastAPI, Django, SQL

**Cloud**
Azure, AWS

**Databases / Platform**
Snowflake, PostgreSQL, Azure Data Lake, Azure Delta Lake, Chroma Vector DB, Azure Databricks

**Libraries**
Pandas, NumPy, Matplotlib, OpenCV, Spacy, scikit-learn, XGBoost, LightGBM, Faker, MSAL, PyAtlan

**Generative AI**
LangChain, LangGraph, MCP, RAG, Agentic AI, AWS Bedrock, Microsoft Azure AI Studio, Agent2Agent (A2A) Protocol

**Deep Learning Frameworks / Tools**
PyTorch, Triton Inference Server, TensorRT, Vllm, Text Generation Inference, Deepspeed, PEFT, Flash Attention, Transformers, Ollama

**DevOps / CICD**
Jenkins, Terraform, Docker, Git, AquaScan, SonarQube, JFrog, HashiCorp Vault, Bitbucket, Azure DevOps

## Professional Experience

**CoForge Ltd.** 🔗
*Lead AI Engineer*

07/2024 – Present
Noida, India

- Led 2 GenAI data products - **Synthetic Test Data Generator (STDG)** and **Automated Data Classifier** - delivered to *Philip Morris International* (Global FMCG enterprise) and onboarded into *Coforge Data Cosmos*, an enterprise data platform comprising 12 production data products.
- Built **LLM-powered synthetic data generation using Snowflake metadata**, replacing production data in dev/QA and reducing test-data preparation effort by 70–80%, while enabling compliant rollout across initial 10 global markets, with platform adoption scaling further.
- Delivered an automated **column-level data classification system**, tagging 100% of Snowflake columns using LLM-driven metadata intelligence (Snowflake + Atlan), **with PII detection, explainable reasoning, natural-language controls** and multi-format exports.
- Defined **solution architecture and design documentation (HLD/LLD)**, functional specifications, and non-functional requirements including scalability, security, performance, and compliance for enterprise deployment.
- Owned **end-to-end product delivery** - architecture, AWS infrastructure (ECS, ALB, Bedrock, S3, CloudFront, GuardDuty), Terraform IaC, CI/CD pipelines, security reviews, and led Frontend and DevOps engineers, enabling scalable onboarding of new markets.

**ValueLabs LLP** 🔗

*Consultant AI & Machine Learning*

11/2023 – 04/2024
Hyderabad, India

- Led development of **AiDE (AI-Driven Engineering)**, an enterprise LLM platform powering **code copilots and NL→SQL analytics**, optimized for cost and latency across multiple client deployments.
- Led **LLM pre-training and instruction tuning (3B to 40B)** using FSDP/DeepSpeed and PEFT (LoRA/QLoRA), training on large codebases and domain datasets; built AiDE Copilot achieving internal HumanEval performance exceeding GPT-3.5.

- Delivered **production NL→SQL systems** for Postgres/MySQL, improving query accuracy by 30% and deploying to 3 enterprise clients generating **$500K+ yearly savings**.
- Productionized high-throughput inference using **vLLM (TensorRT)** and **TGI**, supporting 250 concurrent users with enterprise safety guardrails and monitoring.

*Associate Consultant AI & Machine Learning*

10/2021 – 10/2023
Hyderabad, India

- **Trained and fine-tuned LLMs** across open-source models (Falcon, CodeLlama, LLaMA2, Pythia, MPT, Flan-T5), using NVIDIA A100 GPU clusters.
- Delivered over **20 POCs** and **8 MVPs** across GenAI, NLP, CV and predictive analytics, transitioning multiple solutions from experimentation to production.
- Recognized as **Super Star of the Quarter** for consistent delivery of high-quality solutions under aggressive timelines and budget.
- **Microsoft Certified:** AI-900, DP-900, AZ-900, SC-900

*Technical Consultant AI & Machine Learning*

07/2020 – 09/2021
Hyderabad, India

- Built **intelligent document extraction systems** using fuzzy logic, regex and OCR pipelines, achieving **97% accuracy** across Indian and international identity documents.
- Developed **invoice and legal document extractors** using ELECTRA and BERT-based QA models, reaching **92% accuracy** in production workflows.
- Designed and deployed **20+ RESTful APIs** using Django and FastAPI, containerized with Docker and hosted on **AWS EC2**.

**Defence Terrain Research Laboratory (DTRL, DRDO)**
*ML Research Trainee*

12/2019 – 05/2020
New Delhi, India

- Developed a **novel unsupervised band-selection** approach for hyperspectral imagery leveraging **multi-objective optimization** and **Boltzmann entropy**.

## Education

**Bachelor of Technology - Computer Science and Engineering**
*SRM Institute of Science and Technology* 🔗

07/2016 – 05/2020
Chennai, India

## Projects

**Claims Automation Solution**
*Python, AWS SageMaker, Snowflake*

10/2021 – 03/2022

- Designed and deployed a **Smart Virtual Adjuster** for a leading **P&C insurance enterprise**, automating claims assessment and accelerating end-to-end processing.
- Processed **400k historical claims (2 years)** from Snowflake; trained an **XGBoost classifier** via AWS SageMaker to identify adjuster-modifiable claims.
- Achieved **87% accuracy and recall**, reducing manual workload by **40%** and accelerating claim processing through an end-to-end **AWS SageMaker** ML pipeline.

**Fraud Monitoring Solution**
*Python, Azure, Databricks, Power BI*

05/2021 – 08/2021

- Built an anomaly detection platform for procurement tenders in collaboration with the **Inter-American Development Bank (IDB)** and **Microsoft**, analyzing 10 years of contracting data to detect irregular spending patterns across South America.
- Engineered ML pipelines using **iForest + statistical inference** on **Azure Databricks, ADLS Gen2,** Azure DevOps to detect unusual payments, transaction variability and abnormal bill patterns; integrated real-time dashboards in Power BI.
- Flagged **$10B worth of tenders** for investigation, significantly strengthening fraud detection and compliance oversight.