

# KANHAIYA KUMAR

Data Scientist & GenAI Engineer — ML | NLP | Azure | Automation

Pune, Maharashtra, India | Phone: +91 8405929349 | Website: kanhaiya.site

## PROFESSIONAL SUMMARY

Data Scientist & GenAI Engineer with 4+ years of experience designing AI-driven automation systems for document intelligence, quoting workflows, and engineering teams. Skilled in ML, NLP, Generative AI, Azure, and Python. Experienced in building production-ready pipelines for structured extraction, summarisation, and validation. Focused on scaling GenAI systems, refining prompts for real user needs, and strengthening data engineering foundations.

## CORE SKILLS

• Machine Learning	• NLP & LLMs	• Generative AI
• Python, Pandas, NumPy	• Azure, Azure Blob, DevOps	• CI/CD Pipelines
• Document AI	• Prompt Engineering	• QA / UAT Testing

## PROFESSIONAL EXPERIENCE

**Xoriant — Software Engineer / Data Scientist** (Sep 2022 – Present)

- Developing GenAI-powered document intelligence workflows for RFQs and engineering specifications.
- Building Python + Azure OpenAI pipelines for extraction, summarisation, classification, and entity recognition.
- Designing refined prompt-engineering systems aligned with Application Engineer expectations.
- Integrating Azure Blob, file converters, validation layers, and automation modules.
- Scaling AI workflows used daily by engineering and quoting teams.
- Conducting QA/UAT and improving pipeline robustness and accuracy.

**Xoriant — Associate Software Engineer** (Aug 2021 – Sep 2022)

- Optimised build performance using BuildGrid, RECC, and Bazel.
- Enhanced automation in build and release pipelines.
- Contributed to Azure deployments and CI/CD enhancements.

## PROJECT HIGHLIGHTS

- **Seal Quote Automation — Data Extraction Pipeline** — Built an end-to-end RFQ processing workflow that structures engineering & commercial fields, improving quoting efficiency.
- **Specification Log Extraction System** — Automated extraction + rule-based validation producing high-quality structured data.
- **Spec-Extract GenAI System** — LLM-powered summarisation & keyword-driven extraction for voltages, pressures, materials, and technical attributes.

## EDUCATION

**B.Tech in Computer Science Engineering**

Orissa Engineering College (2017–2021) — CGPA: 8.65

## CERTIFICATIONS

- Microsoft Azure Fundamentals (AZ-900)
- Microsoft Azure Administrator (AZ-104)
- Microsoft Azure AI Fundamentals (AI-900)
- DevOps / CI-CD Certification
- Linux / RHEL Certification
- Python Programming Certification