

Rishabh Jain

rishabh12j@gmail.com | Indore, India | rishabh12j.github.io | linkedin.com/in/rishabh12j | +91-9111440403

PROFESSIONAL SUMMARY

Applied AI Engineer with almost 5 years of experience bridging enterprise integration and production AI across GenAI, LLMs, RAG, NLP, and computer vision. Built cloud-deployed recommendation, document understanding, and safety-constrained robotics systems on AWS and GCP, engineering deterministic validation and guardrails around probabilistic models. Holds an MSc in Robotics and Embedded AI with First Class Honours from Maynooth University, Ireland. Based in India, open to relocation, and available to join immediately.

EDUCATION

Master of Science - Robotics and Embedded AI

Sep 2025 – Sep 2026 | Maynooth, Kildare, Ireland

Maynooth University

Relevant Coursework: Deep Learning, AI & Language Processing, Computer Vision, HRI

Bachelor of Technology - Information Technology

Aug 2016 – Oct 2020 | Indore, Madhya Pradesh, India

Shri Vaishnav Vidyapeeth Vishwavidyalaya - SVIIT

WORK EXPERIENCE

Maynooth University, Research Assistant

Apr 2026 – Jun 2026 | Maynooth, Kildare, Ireland

Screen-based robot controls can exclude older adults and people with disabilities, while unconstrained LLM voice interfaces risk turning a misunderstood command into unsafe physical motion.

- Integrated a safety-constrained voice-to-action pipeline with physical FarmBot Genesis XL systems at a research greenhouse, implementing firmware-confirmed, tick-and-verify execution so the system reports success only after the robot confirms an action has completed.
- Built the evaluation and logging setup for simulation and hardware testing, including full event-log coverage and a misclassification stress test used to validate that the safety layer blocks unsafe motion.

Infosys Limited, Senior Associate Consultant

Jul 2024 – Sep 2025 | Indore, Madhya Pradesh, India

A loyalty programme showed every member the same broad-tier offers, missing revenue from personalisation, but collaborative filtering alone cannot serve new members with no purchase history, and an LLM cannot sit in the critical path of a sub-second response at the pump.

- Contributed to client discovery sessions and defined the AI architecture, integration requirements, and deployment strategy for an LLM-powered recommendation engine, part of a loyalty platform redesign at a major oil and gas company.
- Architected and delivered a hyperpersonalisation system on AWS Bedrock for LLM contextual ranking of customer offers, solving new member cold start with a hybrid engine that combines collaborative filtering and LLM ranking to lift new user engagement.
- Engineered real-time data ingestion pipelines using Dell Boomi on AWS EC2, connecting legacy POS transaction data to AWS SageMaker inference endpoints and achieving sub-second recommendation latency via an Amazon ElastiCache caching layer.

Tata Consultancy Services

Systems Engineer - AI Engineer

Jan 2023 – Jul 2024 | Mumbai, Maharashtra, India

Template-based OCR broke whenever a supplier changed invoice layout, dumping large volumes into a manual review queue and driving an overdue invoice backlog that cost early payment discounts and strained supplier relationships.

- Developed a Python-based invoice document-understanding system using fine-tuned layout-aware transformer models, achieving 90%+ field-level extraction accuracy across varying invoice layouts and reducing manual processing time by 50%.
- Implemented a deterministic three-way-match gate across invoices, purchase orders, and goods receipts before payment routing, preventing incomplete, low-confidence, or inconsistent model outputs from entering the approval workflow.
- Deployed containerised inference services on GCP Cloud Run with Cloud Storage, Cloud Functions, Cloud Monitoring, and Cloud Logging, contributing to a 35% reduction in overdue invoices and improved SLA compliance.

Systems Engineer - DevOps Engineer

Feb 2021 – Dec 2022 | Mumbai, Maharashtra, India

Critical invoicing, payroll, warehousing, and inventory processes ran across disconnected legacy systems and data formats, with no reliable, monitored integration layer between them.

- Deployed and managed Dell Boomi runtime atoms on AWS EC2, configuring VPC networking, IAM roles, and CloudWatch alarms for production monitoring across invoicing, payroll, warehousing, and inventory pipelines.
- Implemented Dell Boomi and JavaScript middleware integrations across 10+ connector types (Salesforce, REST, SOAP, SFTP, HTTP, FTP, Disk) and formats (JSON, XML, Flat File) for a leading life sciences brand.

- Leveraged AWS S3 as a staging and archival layer for Disk/FTP connector workflows; provided real-time production support maintaining high-availability SLAs.

Earlier: Data Science Intern, TezMinds Software (Jan 2020 – Apr 2020) - DeepFake video classification with a custom CNN in TensorFlow, trained on 100,000+ videos from the Kaggle/Facebook Deepfake Detection Challenge dataset.

PROJECTS

GrowMate-BT (Safety-Constrained Voice-to-Action Framework),

MSc Dissertation Research Project · Robotics · LLM Safety

- Designed and built a ROS 2 voice-to-action framework for FarmBot garden robots, where a locally hosted LLM is restricted to intent classification while deterministic behaviour trees own all planning, tool selection, and safety-critical execution through a guard chain.
- Achieved 94.2% success in simulation and 92.78% on physical FarmBot Genesis XL hardware with zero unsafe states, including zero unsafe motions under a stress test of 120 deliberately wrong intents.

FlowGuard (Agentic Integration Incident Copilot), *Agentic AI · Enterprise Integration*

- Developing an agentic incident response copilot that investigates API and enterprise data pipeline failures by correlating application logs, runbooks, API specifications, deployment records, dependency maps, and historical incidents.
- Designing a stateful investigation workflow for incident triage, evidence retrieval, dependency tracing, root cause synthesis, remediation planning, and human approval, with scoped tool access and auditable, evidence-linked recommendations.

DepthLens (Depth Aware Scene Descriptor), *Computer Vision · Assistive Technology*

- Built a real-time assistive-vision pipeline combining a 1–3B vision language model with Depth Anything V2 and YOLOv8n to generate scene descriptions enriched with object locations and relative depth.
- Fused detection and monocular depth outputs into structured spatial context before VLM inference, enabling the system to distinguish foreground, mid-ground, and distant objects while running on a consumer GPU.

Multilingual Discord Content Moderator, *Reinforcement Learning · AI Safety*

- Formulated multilingual content moderation as a Constrained Markov Decision Process across 13 languages, combining sentence embeddings, calibrated XLM-R toxicity scores, and user infraction history in a 403-dimensional observation space.
- Trained a graduated five-action policy (Allow, Warn, Delete, Timeout, Ban) with hard action masking to prevent invalid or disproportionate enforcement decisions.

TECHNICAL SKILLS

- **AI & Machine Learning:** LLMs, RAG, AI Agents, NLP, Computer Vision, Transformers, Recommendation Systems, Model Evaluation, Guardrails
- **Cloud & Production AI:** AWS Bedrock, SageMaker, EC2, S3, ElastiCache, GCP Cloud Run, Cloud Functions, Docker, Vector Databases, Model Deployment
- **Frameworks & Libraries:** Python, PyTorch, TensorFlow, Hugging Face, LangChain, CrewAI, OpenCV, Scikit-learn, FastAPI
- **Integration & APIs:** REST APIs, Dell Boomi, Enterprise Integration, Data Pipelines, JSON, XML, SOAP
- **Robotics:** ROS 2, Behaviour Trees, Raspberry Pi, Sim-to-Real Testing

CERTIFICATIONS

- Generative-AI Cohort 1 by 100xEngineers
- Generative AI with Large Language Models by DeepLearning.AI
- Deep Learning Specialization by DeepLearning.AI

LANGUAGES

English — Fluent

Hindi — Native