

ROHIT JAIN

✉ career.rohit@yahoo.com

🌐 [LinkedIn](#)

🐙 [GitHub](#)

📄 [Recommendations](#)



Senior Artificial Intelligence Engineer with a master's degree in AI from the National University of Singapore and **7 years of experience** building production-grade Generative AI, Computer Vision, NLP, and multimodal solutions across the *Finance, IT, Agritech, Construction, and Automotive* industries.

Skilled in architecting, developing, and deploying end-to-end ML systems, from research and model training to containerized microservices on cloud infrastructure.

AREAS OF EXPERTISE

Generative AI · Large Language Models (LLMs) · Retrieval-Augmented Generation (RAG) · AI Agents · Prompt Engineering · Fine-tuning · Embeddings · Machine Learning · Deep Learning · Computer Vision · Natural Language Processing · MLOps / LLMOps

LANGUAGES & DATABASES

Python · PostgreSQL · MongoDB · Vector Databases

ML LIBRARIES

PyTorch · TensorFlow · Transformers · HuggingFace · SentenceTransformers · OpenCV · spaCy · NLTK · scikit-learn · Pandas

LLM FRAMEWORKS

LangChain · LangGraph · LlamaIndex · OpenAI API · RAG Pipelines

TOOLS & CLOUD

Docker · Kubernetes · MLflow · ClearML · Flask · FastAPI · Git · GitHub · AWS · Azure

EXPERIENCE

Meinhardt Group

2025 – Present

Senior Data Scientist

Singapore

Multimodal · Agentic RAG · Graph RAG · RAGAS · LangGraph · MLOps · Azure

- Developed a **Multimodal Agentic RAG** system for Q&A over technical documents and engineering drawings, iterating from vector RAG to Graph RAG to resolve accuracy degradation on complex, relational queries, improving **multi-hop query accuracy by 35%+** over baseline vector retrieval.
- Designed a **Graph RAG pipeline** that extracts entities, relationships, and key claims from engineering documents into a knowledge graph, enabling multi-hop reasoning across interconnected concepts and **reducing hallucination by 40%+** by overcoming the structural context loss of chunk-based vector retrieval.
- Implemented **hierarchical, structure-aware chunking** to preserve tables and cross-references in engineering documents, replacing fixed-token splitting and cutting context-fragmentation retrieval errors by **30%+**.
- Benchmarked RAG pipeline quality using **RAGAS** (faithfulness, context precision, context recall, answer relevancy), achieving **95%+ accuracy, 99%+ consistency, and <5% hallucination rate** across production evaluation sets.
- Engineered a **scalable microservices architecture**, containerizing ingestion, partitioning, and query pipelines with Docker into 10+ independently deployable services.
- Deployed and orchestrated the application on a **self-managed Kubernetes** cluster, ensuring **99.9%+ uptime** and efficient auto-scaling under variable load (100+ concurrent requests).

Ailytics AI

2024 – 2025

Senior AI Data Scientist

Singapore

PyTorch · OpenCV · ML Pipeline · Vision LLM · Tracking · Pose Estimation · MLOps

- Built an **AI video-analytics** solution for worksite safety and productivity, implementing state-of-the-art object detection, multi-object tracking, and human/facial pose-estimation models with **90%+ precision and recall**.
- Designed and deployed a **Multimodal Vision LLM** (image-text-to-text) that flags safety violations in non-rule-based scenarios by analyzing images against user-configurable textual prompts.

- Accelerated inference through **model optimization and quantization** to ONNX, TensorRT, and NCNN formats.
- Created an **end-to-end ML pipeline** in ClearML covering dataset versioning, model training, evaluation, and benchmarking.

Vulcan AI Agritech

Artificial Intelligence Specialist

2022 – 2024

Singapore

↪ **SmartGrader** *Detection · Tracking · Segmentation · Classification · TensorFlow · Azure · GitHub Actions*

- Designed and implemented edge-deployed AI for automated **fruit counting and grading** using computer vision.
- Built object detection, tracking, segmentation, and classification models across agritech solutions, achieving **99%+ counting accuracy** and **95%+ grading accuracy** while significantly reducing manual labor.
- Optimized inference through model quantization, cutting model size by **40%** and improving inference speed by **60%**.
- Developed a **CI/CD pipeline** using Docker, Azure Container Registry, and GitHub Actions.

↪ **Chat with Fruit Doctor** *Large Language Models · Generative AI · PEFT · LangChain · HuggingFace*

- Implemented an **LLM-based chat platform** enabling farmers and stakeholders to query information on fruit health, growth, and nutrition.
- Fine-tuned a **7B-parameter LLM** on a domain-specific dataset, achieving high accuracy on agricultural Q&A.
- Curated a **10K-sample fine-tuning dataset** compiled from domain experts and government guidelines.
- Improved answer quality by incorporating **Retrieval-Augmented Generation (RAG)** over 100+ research papers.

↪ **Super Resolution of Satellite Images** *GAN · Generative AI · Remote Sensing · Python · PyTorch*

- Designed a **patented Super-Resolution GAN** that generates high-resolution (HR) crop imagery from low-resolution satellite images, achieving **5× upscaling**.
- Reduced annual HR-imagery acquisition **cost by 12×** and improved crop health and nutrient prediction **performance by 50%** using the generated HR images.
- Improved output quality by designing a **custom loss function** with perceptual (feature) loss, optimizing for visual similarity over pixel-space metrics.

Panasonic R&D Center Singapore

AI Engineer (part-time, during master's)

Jan – Jun 2022

Singapore

↪ **Autonomous Vehicle Vision** *Neural Networks · OpenCV · GAN · PyTorch · Video Analytics*

- Built a deep learning model to detect street objects occluded by raindrops, enabling safety-critical toggling of autonomous vehicle (AV) self-driving mode.
- Applied **GANs for image de-raining** alongside OpenCV image-processing techniques to minimize raindrop occlusion across varying sizes, shapes, and intensities.
- Implemented a **Siamese neural network** for one-shot verification of predicted objects at inference.

↪ **Topic Transition System** *Speech Processing · NLP · HuggingFace · AWS*

- Built a topic-transition and sentiment-detection system for analyzing conversations in a meeting-room environment.
- Developed a state-of-the-art **Wav2Vec2-based Automatic Speech Recognizer (ASR)** dedicated to the Singaporean English accent.

- Achieved a **5.9% word error rate (WER)** on the NSC dataset, outperforming the leaderboard Wav2Vec2 baseline.
- Applied NLP techniques to ASR-transcribed text to design the topic-transition detection model.

Staple AI

Data Scientist Intern

Apr – Jul 2021

Singapore

OpenCV · Feature Detection · NLP · Flask · Docker

- Built image-processing pipelines for document image correction using OpenCV and traditional computer vision techniques.
- Designed deep learning models to identify key features within documents using CNNs such as *Faster-RCNN* and *Mask-RCNN*.
- Enhanced data-extraction algorithms for financial documents; applied NLP techniques to compare extracted data against ground truth and visualized results with Plotly.
- Developed REST APIs and deployed scalable, dockerized Flask applications on AWS.
- Managed independent projects with direct client engagement across multinational banks.

Hewlett Packard Enterprise R&D

Data Scientist

2018 – 2021

Bangalore, India

Linear Regression · Decision Trees · Boosting · MLOps

- Designed, modelled, and maintained an **AI-powered prediction engine** that proactively detects and auto-resolves infrastructure issues spanning *availability, performance, security, and capacity*.
- Built and automated an **end-to-end ML pipeline** on HPE *Ezmeral* for data ingestion, cleaning, feature engineering, model training, and deployment.
- Proposed new data-collection points and developed interfaces to ingest sensor data.

EDUCATION

National University of Singapore

Master in Computer Science, *Artificial Intelligence* specialization | CAP 4.45 / 5

2021 – 2022

Singapore

Manipal University

Bachelor of Technology in *Electrical & Computer* Engineering | CGPA 9.36 / 10

2014 – 2018

Karnataka, India

PROJECTS

Video Visual Relationship: Video Captioning

CNN · RNN · InceptionV4 · ResNet128

- Designed a *seq2seq* architecture to identify visual object-relationship pairs in video, achieving **75% top-5 accuracy**.
- Extracted and sequenced video-frame features using *InceptionV4* and *ResNet128* CNN models.
- Modelled a bidirectional, multi-stacked RNN decoder to generate captions; improved overall accuracy via model ensembling.

Question Answering on Stanford SQuAD v2.0

BERT · RoBERTa · Transformers · NLP

- Built a transformer-based system to answer questions from Wikipedia reading comprehensions.
- Explored architectures including BERT, ALBERT, ELECTRA, RoBERTa, and XLNet as base models for the QA system.
- Achieved a **top-35 global leaderboard rank** with an **F1 score of 87.18** using a custom fine-tuned RoBERTa-large.

Live Face Mask Detection

MobileNet v2 · Transfer Learning · OpenCV · Computer Vision

- Built a **98%-accurate** real-time face-mask detector that handles masks of any pattern and colour.
- Applied transfer learning with a *MobileNet v2* model on extracted faces to classify mask vs. no-mask.

CERTIFICATIONS

- **Deep Learning Specialization** – deeplearning.ai 
- **Machine Learning in Production** – deeplearning.ai 
- **Deep Learning with PyTorch** – FreeCodeCamp.org 
- **Python for Data Science, AI Development** – IBM 

ACHIEVEMENTS

- Received the **HPE Employee Recognition Award** three times for innovation and impact delivered through work.
- Awarded the **Director's Medal for Academic Excellence**, finishing in the top 2% of a class of 190 students.
- Finished in the **top 1%** of all participants in the 10th International Mathematical Olympiad (IMO).
- Won **2nd prize** in an eye-controlled robotics event at the Indian Institute of Technology (IIT) Madras.