



ROSHAN ABRAHAM

AI ENGINEER

EDUCATION

Bachelors of Technology in Computer Science and Engineering,
2017 - 2021

St.Aloysius Institute of Technology
Grad with Honors

CONTACT



India | USA | UK



+91 8239778877



abraham.roshan.a@gmail.com

SKILLS

AI and Machine Learning



Cloud Infrastructure



Team and project Architecture



Backend Programming



LANGUAGES

English



Hindi



WORDS

As an AI engineer, I excel at AIML dev, cloud computing, and swiftly adapting to new technologies as they arise. Like Batman's utility belt, I possess the knowledge to craft advanced AI algorithms, machine learning models, and deep learning architectures. With innovative thinking, I conquer challenges and hurdles.

EXPERIENCE

AI Solutions Consultant

Deloitte USI

2024 - present

Responsibilities

- Handling Govt projects and transforming ai solutions for enterprise consumption
- Bulding Agentic solutions State Govt orgs adn bodies
- Winner of AI Hackathon for PR Agenti solutions

AI Team Lead

Appsynergies

2019 - 2024

Responsibilities

- Develop and Arch APIs systems for model inference
- Planning MLOps and all segments
- Delivered real-world optimized AI solutions leveraging custom industry expertise
- Planning sprint scope and ML artifact releases
- Mentoring junior AI engineers on the team

Project Lead AI Engineer

MyDiGiRecords

2022 - 2023

Responsibilities

- Planning sprint scope ML artifact releases
- EPIC and Cerner medical distributor integration
- Preprocessed 100+ GB of Doc AI model for training and fine tuning
- Develop and Arch APIs systems for ABHA FHIR integration
- Planning and Dev for end to end OCR NER into FHIR system

AI Consultant

Tasset

2022 - 2023

Responsibilities

- Large Scale Data scrapping and text dataset generations
- Blockchain ready chatbot provisioning for client needs

SKILLS AND TOOLS

VISUAL AI

Classification and Detection



- TensorflowServing and Pytorch, OpenCV, VisualAI, YOLOv4/YOLOv5, DeepLab, RetinaNet, FasterRCNN
- Segment Anything Model by Meta attached with Reasoning describer.
- ONNX: Open Neural Network Exchange Format
- NNEF: Neural Network Exchange Format
- ImageNet Augmentation and ControlNet

CLOUD

Experience and skilled with Teraform, Docker Engine, VM SSH, container CI/CD, artifact registry



AZURE AND IMB



GCP



Vertex AI, Language Bison Model Document AI, VisionAI, Cloud Build, DialogFlow, Cloud Run, Container registry, PubSub

VM serverless hosting, Quantum Computing, Cloud Foundry, Container Management,

AWS

Lambda, Fargate, SageMaker, AWS Transformer framework

ANALYTICAL AND VISUAL



Apache Airflow, Talend, Vector DB, GPT4 Graph, Looker, Google Data Studio

DATABASES AND ENGINEERING



NoSQL

- Firebase
 - MongoDB
-

90%

MySQL

- PostgreSQL
 - BigQuery
-

75%

Vector DB

- FAISS, chroma
 - Pinecone, Weaviate
-

75%

VISUAL AI

Segmentation and Generation Models



- Segment Anything Model, classification and detection, yoloV5
- TensorflowServing, Pytorch and CUDA base systems
- SD-XL2, ControlNet v1.1 for text2image and image2image
- Stable Diffusion Training and Eval Models using WandB

LLM KNOWLEDGE ENGINEERING



PROMPT ENGINEERING

LangChain, IndexLlama

- Agent and chains with multiD objective directives
- Specialized Chain of thought sequencing and tool attachment



KB and embedded indexing

- connecting chains and agents via vector store or embedding router
- pdf/csv/image data ingestion
- indexes, Embedded KB hosting locally and foreign



Autonomous System and tools

- Internet scraper and documentation tools
- File generation transformers for visual input and document generation
- Custom Py function tool integrations

LLM TRAINING TUNING FOUNDRY

Custom Training and fine tuning these chat , agent, and Instruct Models

LlAMA2 7B



CodeLlama



Beluga



Mistral



OpenAI API



Falcon 180B



WizardLM



30B



HUGGING FACE



GPU AND MACHINE TRAINING

CUDA

NeMO

NVIDIA A100/ T4,
remote and booted

vLLM

HFSpaces

RunPods LLM
deploy and host
text-gen-webui

RTX 3050ti

VIRTUAL MACHINE MASTERY



Docker/docker-compose/Helm



90%



Terraform

75%



GCP Cloud Build /
AWS SageMaker

75%

MLOPS



SKILLS

- Experience with ML frameworks such as PyTorch, TensorFlow, TFX
- Proficiency in experiment tracking and workflow versioning
- Skills in unit and integration testing, data, and model validation before adding to registry
- Post-deployment health monitoring
- Experience with end-to-end MLOps over platforms like SageMaker, GKE, MLFlow



TOOLS AND FRAMEWORKS

- Weights and Biases
- DataBricks
- MLflow
- AWS
- RunPod



databricks



- vLLM/openLLM
- neptune.ai
- GCP vertex AI

FUTURE STUDY AND SPECIALIZATION ASPIRATION

EXPLAINABLE AI

- I'm fascinated by XAI models that autonomously comprehend data without human supervision, analyzing and summarizing to draw novel insights.
- AI can bring entirely new, unbiased perspectives to problems by directly reviewing data, accelerating research and discovery.
- I want to pioneer black box AI techniques focused solely on reading, understanding, and deducing useful conclusions from knowledge, without exposing internal workings.

PYTHON DEV

Experience in vast python environments and distros.

Specialized in

- Data science and Engineering



- API and SaaS systems



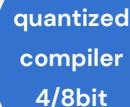
- VM, python compilers and container images



- AI-ML frameworks



- Machine language and compiler instruction



Projects, Responsibility and Specializations

AI Engineering / ML Dev and Training / DeepLearning Integrations / MLOps / Data Engineering / Prompt Chain Engineering / Data Science / Data Analysis

Project Experience



Custom LLM foundry

Custom Models Solo Project



Medical admin AI

AI Medical Infrastructure



Stable Diffusion pipeline

SD-XL2.1.2, ControlNetv1.1



ChatBot System

Complex LLM and model system wrapped in rasa



Data Ops

Dev and design Data pipeline, agent handler, data set cross creations and validation



Dynamic Mlops Engine

End to End Model dev and deploy, Cloud Infrastructure

- Preparing suite of Custom Training data collection and instruction Tuning for base, instruct and chat formats
- Rostering LLMS architecture parameters of open-sourced model
- Design and Management of LLM registry management and LLMOps processes over docker and WandB
- Integrating Experimental Integration into Local LLMs (llama2, codellama)
- Integrations with Doc to token KnowledgeDB (embedded-routing) for personalized coding and research chain of actions
- Designing and Dev Document text OCR engine
- Designing and Dev Medical NER parser for complex Medical Documents
- Built Ingestor Doc translation for PDF/Json/csv/raw corpus
- Translating raw data corpus to FHIR doc to FHIR & SMART documents and resources
- LLM powered Knowledgebase and Bedside manners protocols
- FineTuning for instructional text and image diffusion on stable diffusion 1.1 for a new artistic style addition (WDW style : nour cyberpunk)
- Prompt Chain simple request conversion into complex set of Prompt instruction
- Dev QR code ControlNet powered Stable Diffusion
- Using LangChain, Rasa and Sanic framework as basis of bot integrations
- Arch and training Seq2Seq and LSTM connection to model stack
- LLM integrations into model stack, with Eval param
- Plan and dev Different Models deployed into system
 - Agent Specialized Medical PromptRouters
 - Agent FHIR Query manager
 - Fine tuned Classify model
 - Instruction prompter Context Gen for formating standard formats
- Cross platform data conversion and traffic protocols
- Internal VPC and secure external data mutation
- Server, db, Service Chain handler for running and management
- Meta and composed data generation
- Terraform and Docker Helm based cloud infrastructure
- Model Deployment over several options of serverless and edge machine for end user delivery
- Leveraging Large machine GPUs and TPUs segmentation for training dev ops

Responsibility and skills

- **Specialized Knowledge extraction for projects**
LLM managed data extraction paired with extractors, evaluation metrics and judgement criteria for PDF/CSV/TXT/IMG Knowledge tree conversion (HWSF/FlastL2)
- **Working with different combo of Model Stack and new integration**
TFX, CUDA, HF transformers, DataBricks, WandB, Edge device encoder and decoder, Gradio, StreamLit, Reveries
- **Data Collection via Crawling, scrapping, modulation and generation**
Selenium, Agentsmanager, Indexer timestamped and routed
- **Converting information into a personal Knowledge base and project assistant**
Local and Cloud image model training and hosting for KB, Indexes and Embedded Memories

Platforms

HuggingFace

Replit / Wolfram

AutoGen

LangSmith

Certifications and Courses

Studies conducted with Institutes, online educators , Open Knowledge and Open Source communities

Specializations Courses taken

Natural Language Processing Specialization

- Classification and Vector Spaces
- Probabilistic Models
- NLP with Sequence Models
- Attention Models

Generative Adversarial Networks (GANs) Specialization

- Build Basic Generative Adversarial Networks
- Build Better Generative Adversarial Networks
- Apply GAN deployments

AI for Medicine Specialization

- AI for Medical Diagnosis
- AI for Medical Prognosis
- AI For Medical Treatment

TensorFlow: Advanced Techniques Specialization

- Custom Models, Layers, and Loss Functions with TensorFlow
- Custom and Distributed Training with TensorFlow
- Advanced Computer Vision with TensorFlow
- Generative Deep Learning with TensorFlow



Review some of my AI interactive environments.

Please contact for demoing more complex systems needed Hardware bootup

Roshan Abraham
(dweebster/projects/list)

Certifications

Artificial Intelligence and Deep Learning

(01/2020 – Present)

An online course on building and deployment of neural network systems

Python Developer – Microsoft Technology Associate

(04/2020)

Examination conducted by Microsoft for validation of skills and industry presets required

End-to-End Machine Learning with TensorFlow on GCP

(08/2020)

coursera.org/verify/7BM64VLTXCGH

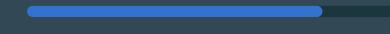
Reliable Google Cloud Infrastructure: Design and Process

(07/2020)

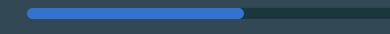
coursera.org/verify/JXTJTFR4WC2N

Strengths

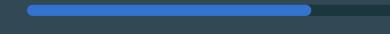
Data handling and engineering



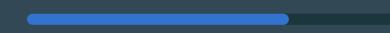
System designs and solution arch



MLOps



LangChain frameworks Systems



Medical diagnosis Administration systems

