# Vishvdeep

## AI Engineer

Ahmedabad, India

aiwithvd.com

## KEY ACCOMPLISHMENTS

- Introduction to LangGraph
- Data Science for Business

## SKILLS

LangGraph

Multi-Agent, Agentic RAG

LangChain

QA Chain, Summarization Chain, Custom Chain, Document Loaders, Output Parser

LangSmith

Chain Tracing, Evaluation, Prompt Mangagement

TensorFlow

Keras deep learning model development

PyTorch

Transfer Learning

HuggingFace

Transformers, PFT

LLMs

OpenAI, Anthropic, Ollama, LLaMa, Bedrock, Groq

MLOps

MLFlow, Weights and Biases

Cloud

AWS Services like EC2, S3, Lambda, Cognito, DynamoDB, Kinesis, VPC, IAM, Load Balancers, SNS, SES, CloudFormation, CloudFront, Route53, CodepPipeline, CodeBuild, CodeDeploy.

GCP Services like Kubernetes Engine, Cloud Storage, PUB/SUB, Cloud Scheduler.

## SUMMARY

5 years of experience in AI, Data Science, and MLOps, with a focus on development, validation, and deployment of Cloud-based data solutions.

- Proven track record in developing and optimising AI/ML microservices, reducing healthcare project microservice from 8 to 3 using Generative AI.
- Experience in transforming fintech objectives with tools such as LendAssist and Fintech-Lending, utilising Gen AI and Explainable AI (XAI) methodologies.
- Hands-on experience in managing MLOps, ensuring seamless model integration into production pipelines for robust, scalable solutions.
- Led healthcare product development from research stages to MVP, delivering impactful AI-driven solutions to meet industry needs.
- Skilled in leveraging LangChain, LangGraph, LangSmith, LlamaIndex, OpenAI, Anthropic, Llama, and other major LLM providers and open-source models.
- Proficient in production-grade tools such as MLFlow and DVC for model tracking and versioning.
- Developed and deployed Agentic AI and Retrieval-Augmented Generation (RAG) applications.

## WORK EXPERIENCE

2024 - present

### Senior Software Engineer

7Span

- Design, develop, and optimise GenAI and AI/ML applications to address complex business needs, utilising Python and advanced machine learning frameworks.
- Build and maintain robust ML and GenAI pipelines, ensuring efficient model deployment and alignment with technical requirements across various environments.
- Collaborate with cross-functional teams to integrate AI/ML solutions seamlessly into existing systems, driving impactful outcomes and enhancing operational efficiency.

2022 - 2024

### Data Science Engineer and Manager

fxis.ai

- Led a team of machine learning engineers, providing mentorship and technical guidance while planning, prioritising, and managing the execution of ML projects from development to deployment.
- Collaborated with cross-functional teams to ensure data availability and quality, overseeing preprocessing, feature engineering, and continuous model optimization in production environments.
- Communicated project insights and progress to stakeholders, documented methodologies, and stayed updated on industry trends to incorporate the latest ML techniques.

2019 - 2022

### Software Engineer

Accenture

- Analysed feature change plans, conducted impact assessments, and coordinated discussions with the solution plan author to address queries on upcoming features.
- Actively participated in Scrum ceremonies, including daily stand-ups, grooming, review, and retrospective meetings, to ensure alignment across the team.
- Managed task assignments within the Scrum Test team, ensuring all user stories were completed on time and met quality standards.

Vector Database

Pinecone, Qdrant, Weaviate, ChromaDB, FAISS

Analytical Tools

GrayLog, LangSmith

CI/CD for App Deployment

GitHub Action

Management Tools

JIRA, ClickUp, Rally, Notion, Confluence

## EDUCATION

2021 - 2023 **M.Tech in Data Science and Engineering**
Birla Institute of Technology & Science - Pilani

2016 - 2019 **B.E. in Computer Engineering**
L.D. College of Engineering

2013 - 2016 **Diploma in Computer Engineering**
Atmiya Institute of Technology & Science

## PORTFOLIO

**AI Trip Planner - Agentic AI**

**Description:**

- Customized Itineraries: Generates personalized travel itineraries based on user preferences and travel patterns, ensuring unique and tailored experiences.
- Data-Driven Suggestions: Utilizes real-time data from APIs and databases to suggest places, restaurants, activities, and local events.
- Output Structure: Persist output structure into json format.

**Technology:** LangGraph, Agentic RAG, LangSmith, LangGraph Studio, OpenAI's GPT, Llama 3.2-Vision

**Role**: Data Science Engineer

**Development**:

- **Development:**
  - Agent Development: Develop and refine the trip-planning agent to deliver optimized itinerary results, ensuring seamless user interaction.
  - API Integrations: Implement and manage API connections for real-time data retrieval on destinations, hotels, and activities.
  - Agent Testing and Optimization: Conduct testing and performance tuning of agents using LangGraph Studio and LangSmith, ensuring robust and reliable output.

**Healthcare Document Extractive Summarization - GenAI**

**Description**:

- This is a groundbreaking AI-powered platform specialising in medical record extraction and summarization.
- Utilising patent-pending technology and trained on millions of data points, This product swiftly and accurately extracts critical information from medical records, offering comprehensive patient case summaries in minutes.
- By automating this process, This saves resources, time, and costs associated with manual record summarization. This achieved significant optimization by reducing its microservices from 8 to just 3 (Document splitter, Text classifier, LLM service), enhancing efficiency and performance in processing medical documents.

**Technology**: Micro-service, AWS EC2, SQS, EKS, Terraform, AWS Bedrock LangChain, AzureOpenAI, Pinecone, Open-Search, RAGatouille, ECR, Graylog, Grafana, LangSmith, DeepEval

**Role**: Data Science Engineer

**Development**:

- Designed and implemented a streamlined backend system, reducing eight microservices to three core services—Document Splitter, Text Classifier, and LLM Service—significantly boosting efficiency and lowering operational costs.
- Developed the Document Splitter and Text Classifier microservices with Python, enabling quick document parsing and accurate classification of medical records.
- Built the GenAI Application on AWS Bedrock for the LLM Service, providing automated, high-quality patient case summaries in minutes.
- Leveraged the DeepEval framework to assess GenAI accuracy with GEval and other LLM metrics, ensuring reliable performance.
- Implemented CI/CD pipelines for streamlined deployments, and enforced HIPAA-compliant data security.

**LendAssist**

**Description**:

- LendAssist is an innovative project on the B2B lending marketplace platform that enhances the loan application process using artificial intelligence and chatbot technology.
- It provides applicants with real-time updates, document retrieval, and personalised support, making the experience smoother and more transparent.
- With LendAssist, applicants can easily track their applications, access information, and receive assistance, all in one user-friendly interface.

**Technology**: Open-Source LLM, OpenAI, VectorDB, LangChain, AWS EC2, Docker, FastAPI, Streamlit, Pinecone, ChromaDB

**Role**: AI Engineer and Consultant

**Development**:

- Developed a GenAI-powered assistant to analyse bank statements, helping underwriters quickly extract valuable information.
- Leveraged LangChain and other AI agents to assist underwriters in the loan underwriting process, acting as a chatbot for interacting with bank statements and other financial documents.

**B2B Lending Marketplace - GenAI**

**Description:**

- B2B Lending Marketplace is a groundbreaking platform that simplifies lending processes for businesses and lenders. It acts as a lending marketplace, connecting merchants needing quick loans with interested lenders.
- B2B Lending Marketplace also offers a seamless SaaS platform for lending businesses, eliminating the need for new development efforts.
- Implemented Gen AI for matching lenders and merchants, XAI for underwriting loan offers, and a risk scanner to assist lenders. Collaborated across teams to streamline loan processes and provide accurate data insights for informed decision-making and risk assessment.

**Technology:** Unsupervised Learning, AWS, Microservice, Docker, OpenAI, Open-Source LLM, LangChain, HuggingFace, FastAPI

**Role**: AI Engineer and Consultant

**Development**:

- Built a B2B Lending Marketplace platform to connect businesses seeking loans with lenders, simplifying the lending process and reducing time-to-funding.
- Developed a SaaS platform for lending businesses, providing a ready-to-use solution that eliminates the need for additional development efforts.
- Integrated GenAI to match lenders with merchants efficiently and utilised Explainable AI (XAI) for transparent loan underwriting.
- Implemented a risk scanner tool to help lenders assess risk accurately, enhancing decision-making and risk management.
- Collaborated with cross-functional teams to optimise loan processing and deliver valuable data insights for better risk assessment and decision support.

**Predict Disease From Retinopathy**

**Description:**

- The project aimed to develop a deep learning solution for automated grading of diabetic retinopathy using transfer learning and modified classification layers.
- The goal was to improve the accuracy and speed of DR diagnosis and treatment using tensorflow, keras, and image thinning.
- The developed solution was containerized using Docker and served using FastAPI, helping to prevent vision loss from diseases such as diabetes and hypertension.

**Technology:** FastApi, Docker, TensorFlow, Keras, Pandas, MLFlow, OpenCV, AWS, Redis, AWS S3

**Role:** Data Science Engineer

**Development:**

- Developed a deep learning solution for automated diabetic retinopathy grading, leveraging transfer learning and modified classification layers to enhance diagnosis accuracy and speed.
- Utilised TensorFlow, Keras, and image thinning techniques to improve model performance for detecting diabetic retinopathy.
- Containerized the solution with Docker and deployed it via FastAPI, facilitating easy access and integration.
- Aimed to support early diagnosis and treatment, helping prevent vision loss due to diabetes and hypertension.

**Cover Letter Generation**

**Description:**

- The Cover Letter Generation project aimed to create an automated system that generates a personalised cover letter based on a job description and linked personal information.
- The Cover Letter Generation project used FastAPI, a chatbot API, and Docker to create a web application that scrapes job descriptions and generates customised cover letters for job applications. Docker ensured smooth deployment and operation. The end result was a concise cover letter, tailored to the specific job application.

**Technology**: FastAPI, Docker, OpenAI, LangChain, AWS EC2

**Role**: Data Science Engineer

**Development**:

- Created an automated system to generate personalised cover letters based on job descriptions and personal information, enhancing the job application process.
- Built a web application using FastAPI, OpenAI, and LangChain to scrape job descriptions and generate customised cover letters tailored to specific roles.
- Utilised Docker for streamlined deployment and reliable operation on AWS EC2, ensuring scalability and smooth performance.
- Enabled applicants to produce concise, job-specific cover letters with minimal effort.

**MenuMaster**

**Description:**

- MenuMaster revolutionises recipe and menu management with advanced ML capabilities. It automates recipe scraping for quick costing and optimises menu pricing, nutrition analysis, and special diet labelling for compliance and profitability.
- Its ML-driven Named Entity Recognition (NER) extracts pricing data from supplier lists, ensuring accurate ingredient cost updates.

**Technology**: Transformers, MLFLow, FastAPI, Docker, GitLab, Jenkins, SOC2-compliance, Grafana

**Role**: Software Engineer

**Development**:

- Designed and implemented an ML-powered platform to automate recipe scraping and menu management, optimising pricing, nutrition analysis, and dietary labelling for food service compliance and profitability.
- Developed Named Entity Recognition (NER) models using transformers to accurately extract ingredient pricing from supplier lists, ensuring precise cost updates.
- Built a scalable application with FastAPI and containerized it using Docker, enabling smooth deployment on SOC2-compliant infrastructure for enhanced data security.
- Utilised MLflow for tracking model performance and integrated monitoring with Grafana, while leveraging GitLab and Jenkins to streamline CI/CD processes.

**PilotQA System**

**Description:**

- The project's main focus is harnessing the power of the Language Model (LLM) to create an advanced question-answering system.
- By tapping into the extensive knowledge and natural language processing capabilities of the LLM, the project endeavours to develop a robust system capable of understanding and providing accurate responses to a diverse array of user-generated questions.

- The system will analyse the content of uploaded documents to provide relevant and comprehensive answers, enhancing user interaction and facilitating efficient information retrieval.

**Technology:** Open-Source LLM, OpenAI, Vector DB, FastAPI, Docker, LangChain

**Role**: Data Science Engineer

**Development**:

- Developed an advanced question-answering system leveraging an open-source Language Model (LLM) and OpenAI, enabling accurate responses to diverse user queries.
- Built capabilities for document analysis, allowing the system to extract relevant information from uploaded files to provide comprehensive answers, improving user interaction and information retrieval.
- Utilised Vector DB for efficient storage and retrieval of embeddings, enhancing response accuracy and speed.
- Implemented the system with FastAPI and Docker for seamless deployment and scalability, and integrated LangChain to manage interactions with the LLM.