# Piyush Singh

9205973514 | piyushsingh50105@gmail.com | https://www.linkedin.com/in/piyush-singh-328302289/ |
https://github.com/PiyushVIT346

## Summary

Aspiring AI engineer and CSE (AIML) student at VIT Bhopal, driven by a passion to solve real-world problems through intelligent systems. Throughout my engineering journey, I've built ML and NLP projects, participated in national hackathons, and led collaborative teams, all while exploring the depths of artificial intelligence, web technologies, and innovation. My goal is to contribute to impactful AI solutions that improve lives and push the boundaries of what technology can achieve. Constantly learning, building, and growing to become a well-rounded tech leader of tomorrow.

## Technical Skills

**Programming Languages**: Python, C++, SQL, Java, HTML, JS, ReactJS

**AI and ML skills**: ML, DL, Neural Network, GenAI, LLM, Advance RAG, Transformer(HuggingFace), NLP, Computer Vision

**AI ML Framework and Tools**: LangChain, LangGraph, LangSmith Monitoring, CrewAI, Ollama , MLflow,DVC,Docker, Vercel, CI/CD,Scikit-learn, OpenCV, NLTK, Git, Docker

**Backend Developement and APIs**: API development using FastAPI and Flask

**Data Science**: Python, Statistical Analysis & Modeling, ML, pandas, Numpy Data Visualisation using PowerBI,

## Projects

**Water Potability Detection: MLOPS and CI/CD**
*Project Link*

- Data ingested (in csv format) and performed preprocessing and clean it to extract useful information for further model training.
- Performed hyperparameter tuning, model training, and versioning using MLflow, with proper model registry, tracking, and monitoring integrated into the pipeline.
- Developed and tested RESTful APIs using FastAPI for serving machine learning models and ensuring robust backend integration.
- Implemented Continuous Integration (CI) workflows using GitHub Actions for automating code testing and deployment.
- Conducted experiment tracking and version control through DagsHub for collaborative ML development.
- Built and compared multiple ML model versions, evaluating performance through accuracy, F1-score, and other key metrics.

**Advance RAG**
*Project Link*

- Implementing Hybrid RAG Search (using keyword search $+$ Chroma DB retrieval $+$ vector search) using LangChain and HuggingFace transformer and also having memory.
- Implementing Reranking (rank_bm25, cross encoder and cohere).
- Implement Lost in middle phenomena and overcoming using merger and compression retriever.
- Implement RAG Fusion to combining results from multiple query variations.
- Flash Reranking as it ensures the most useful ones guide the generation.
- Implementing Context Compression Retriever, Self Retriever, Parent-child Retrieval, Sentence Window Retriever and HyDE Retriever.

**Kidney Tumor Detection and MLOps**
*GitHub Link*

- Developed a VGG16-based convolutional neural network (CNN) achieving $94\%$ accuracy on CT scan images for early-stage kidney tumor classification. Used data augmentation and fine-tuning techniques to enhance model generalization.
- Designed and implemented a complete MLOps pipeline covering the full machine learning lifecycle: data and model versioning using DVC, experiment tracking using MLflow and DagsHub, and automated testing and deployment with GitHub Actions.
- Ensured reproducibility, scalability, and collaboration through modular pipeline design, model registry setup, and continuous integration workflows.

• Containerized the trained model using Docker and deployed it through a Flask-based REST API for real-time tumor prediction in production-like environments.

• Conducted performance evaluation using metrics such as accuracy, precision, recall, and confusion matrix, and iteratively improved model using hyperparameter tuning.

**Trained Model for CT-Scan, X-Ray and Brain tumor**
*GitHub Drive*

• Trained a deep learning model to classify medical documents and scans, accurately identifying whether an image was an X-ray, CT scan, or Brain MRI using image classification techniques.

• Enhanced model performance by applying Convolutional Neural Networks (CNN) and transfer learning with ResNet-50, improving accuracy on complex medical image datasets.

• Developed a chatbot for disease detection that interacts with users through symptom-based input, identifies potential illnesses using NLP, and provides basic first-aid suggestions and care tips.

• Built a personalized food timetable generator that considers the user's health history (e.g., past diseases), age, dietary preferences (veg/non-veg), and chronic conditions to generate nutritionally appropriate meal plans.

## Research Papers

• **EMG Signal Disease Detection** – Published in JETIR. View Paper

• **Skin Disease Detection** – Research paper pending publication.

• **EEG Signal Analysis for Disease Detection** – Research paper pending publication.

## Experience

**Intern and Trainee (Machine Learning and Data Science)** Sep 2024 – Feb 2025
*LearnFlu* *Remote*

• Completed a 4-month internship and training program focused on Python programming, machine learning, statistics, and data analysis.

• Gained hands-on experience in data preprocessing, feature engineering, model training, evaluation, and deployment using libraries such as Pandas, NumPy, Matplotlib, and Scikit-learn.

• Built multiple ML solutions for business use cases, including:

  – Iranian Churn Prediction using classification algorithms (Logistic Regression, Random Forest, XGBoost).

  – Online Retail Customer Segmentation using RFM analysis and clustering (KMeans).

  – Real Estate Price Prediction with regression models.

  – Wholesale Customer Segmentation using unsupervised learning (PCA, DBSCAN).

• Strengthened understanding of customer behavior analytics, business decision-making, and predictive modeling for real-world applications.

**CDC Summer Internship** June 2025 – July 2025
*Remote*

• Completed two industry-recognized certifications under the CDC Summer Internship program:

**1. IBM GenAI Certification:** Learned to design and build chatbots using natural language processing (NLP) concepts. Gained practical exposure to IBM Watsonx.ai platform for creating and fine-tuning large language models (LLMs) on IBM Cloud.

**2. Microsoft DP-900 Certification:** Acquired foundational knowledge of data concepts, cloud architecture, and Microsoft Azure tools. Deployed models using Azure ML Studio and worked with SQL, NoSQL databases, Synapse, and Power BI for data analytics and visualization.

**Technical Core Member** Jan 2025 – Present
*Freelancing Club* *VIT Bhopal*

• Collaborated with other core members to organize technical events and workshops focused on freelancing, software development, and AI.

• Gained leadership and project coordination experience by guiding junior members and facilitating project-based learning initiatives.

• Worked on internal tools and helped members explore freelance platforms and technical upskilling.

## Education

**Vellore Institute of Technology (VIT), Bhopal** — Bhopal, India
*B.Tech in Computer Science and Engineering (Artificial Intelligence and Machine Learning)* — *2023 – 2027*
• Current CGPA: 9.05

**Vanasthali Public School** — Ghaziabad, India
*CBSE – Class 12th* — *2022*
• Scored 89.6% in CBSE Board Examinations

**Vanasthali Public School** — Ghaziabad, India
*CBSE – Class 10th* — *2020*
• Scored 96.4% in CBSE Board Examinations

## Certifications

• IBM GenAI Course

• Microsoft DP-900: Azure Data Fundamentals

• LangChain, LangGraph and Ollama-RAG – Udemy

• Intermediate Machine Learning – Kaggle

• Computer Vision – Great Learning

• Data Structures and Algorithms (DSA) – Simplilearn

• Advanced C++ – Simplilearn