

PROFESSIONAL SUMMARY

Machine Learning Engineer specializing in NLP, LLMs, Deep Learning, and RAG systems. Expertise in PyTorch, TensorFlow, Hugging Face Transformers, and production ML pipelines. Proven experience in Reinforcement Learning, backend development with Spring Boot, REST APIs, and database optimization.

EDUCATION

- Savitribai Phule Pune University

Bachelor of Engineering in Information Technology; GPA: 9.32/10.0 (Class Rank: 4)

Pune, India

Sept. 2024 – May 2027

Government Polytechnic Pune

Diploma in Computer Engineering; GPA: 9.52/10.0

Pune, India

Aug. 2021 – May 2024

TECHNICAL SKILLS

- Programming:

Python, Java, C++, C, SQL, JavaScript, HTML, CSS

ML/AI:

Large Language Models (LLMs), Natural Language Processing (NLP), Deep Learning, Neural Networks, Reinforcement Learning (PPO, Q-Learning), Retrieval-Augmented Generation (RAG), Computer Vision, Transfer Learning, Fine-tuning, Sentiment Analysis, BERT, GPT, BART, LSTM, CNN, RNN, Named Entity Recognition (NER)

Frameworks:

PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face, LangChain, OpenAI API, spaCy, NLTK, XGBoost, Spring Boot, Node.js, Express.js, React

Databases/Tools:

MySQL, PostgreSQL, MongoDB, Neo4j, Redis, FAISS, Vector Databases, Git, Docker, Linux, CI/CD, Agile

Other:

REST API, Microservices, JWT, Spring Data JPA, Hibernate, NumPy, Pandas, JUnit, TDD, A/B Testing

PROFESSIONAL EXPERIENCE

- The Bodhi Tree

Artificial Intelligence Engineer Intern

Remote

Sept. 2025 – Present

Infaneat Digital Media

Backend Software Engineer Intern

Pune, India

May 2023 – Jul. 2023

Built end-to-end RAG and LLM pipelines with multimodal support (text, image, audio) using LangChain, FAISS vector databases, and semantic search, achieving 28% improvement in contextual understanding through optimized embeddings and retrieval algorithms

Developed transformer-based NLP models (BERT, GPT) with custom attention mechanisms for emotion-aware sentiment analysis and text classification, improving accuracy by 15% via transfer learning and fine-tuning using PyTorch and Hugging Face

Implemented production ML pipelines with data preprocessing, feature engineering, model evaluation (F1-score, precision, recall), hyperparameter tuning, and integrated computer vision and speech processing models for multimodal AI analysis

TECHNICAL PROJECTS

- Self-Learning Curriculum Engine for LLMs using Reinforcement Learning

Developed autonomous RL framework using Proximal Policy Optimization (PPO) and policy gradients enabling LLMs to self-evaluate and generate adaptive curriculum, achieving 22% accuracy improvement on benchmark NLP tasks — Python, PyTorch, RL, Hugging Face, PPO, Q-Learning, Neural Networks

Oct. 2025

DRONA – Predictive Maintenance System for Aircraft Using ML

Engineered predictive analytics using Kalman Filters and LSTM networks for time-series forecasting on 50+ aircraft component sensors, predicting failures 48-72 hours ahead with 87% accuracy and 1s latency — Python, TensorFlow, Keras, LSTM, RNN, Anomaly Detection, Pandas, NumPy

Jan. 2025 – Present

Briefly – AI Discord Bot for Abstractive Summarization

Built Discord bot using BART transformers for summarizing 1000+ message conversations with 92% coherence; optimized latency from 8s to 1.2s (85% reduction) via Redis caching, multi-threading, and batch processing — Python, Discord API, BART, Redis, NLG

Apr. 2025 – May 2025