

Karandeep Singh

+91-9588339067 | karandeep07012004@gmail.com | linkedin.com/in/karandeep-singh | github.com/karan-0701

PROFESSIONAL SUMMARY

Machine Learning Engineer experienced in **NLP, Generative AI, Monte Carlo Simulation, and AdTech optimization**. Skilled at building end-to-end ML pipelines, fine-tuning transformer models (BERT, LLaMA), and improving system performance through simulation and algorithm design. Proven record of delivering measurable improvements in accuracy, efficiency, and scalability.

EDUCATION

LNM Institute of Information Technology (LNMIIT)

B.Tech Computer Science Engineering

Jaipur, Rajasthan

Aug. 2022 – May 2026

CBSE Board

12th Grade – Percentage: 79.60/100

Ambala, Haryana

2022

CBSE Board

10th Grade – Percentage: 95.40/100

Ambala, Haryana

2020

EXPERIENCE

Teaching Assistant – Introduction to Scripting Language

LNM Institute of Information Technology

Jan. 2024 – Apr. 2024

Jaipur, Rajasthan

- Mentored **100+ students** in Python and ML, driving a **20% improvement** in final project quality and submission rates.
- Designed and delivered **10+ ML-focused assignments** (Linear Regression, SVMs), adopted in the course curriculum.
- Directed weekly peer-led sessions which promoted teamwork and helped students resolve technical issues; increased weekly attendance by 20% as students found collaborative solutions together.

PROJECTS

End-to-End Text Summarization | Python, Hugging Face, Natural Language Processing

Sept. 2025 – Present

- Built an **end-to-end abstractive text summarization system** using **Hugging Face Transformers (BART, T5)** with a modular, config-driven pipeline for training, evaluation, and inference.
- **Fine-tuned Transformer models** on benchmark and custom datasets (e.g., SAMSum), achieving competitive performance measured by **ROUGE** metric.
- Implemented scalable architecture with **YAML-based configs, and GPU/CPU support**, enabling reproducible experiments and easy extension to new datasets.

RAG-Based Customer Shopping Chatbot | LangChain, LLaMA, ChromaDB

Jul. 2025 – Present

- Built a Retrieval-Augmented Generation chatbot for order tracking and FAQs, reducing manual support workload.
- Enhanced response grounding by integrating ChromaDB with optimized chunking and recursive document splitting.
- Designed an interactive Streamlit UI enabling contextual multi-turn conversations for better user experience.

ML Pipeline for Network Security | Python, MLflow, GitHub Actions, AWS

Aug. 2025 – Sept. 2025

- Built an end-to-end ML pipeline with data ingestion, validation, transformation, training, and evaluation.
- Automated experiment tracking and model versioning using MLflow.
- Integrated GitHub Actions for CI/CD and deployed workflows on AWS.
- Designed modular, production-ready components with error handling and logging.

TECHNICAL SKILLS

Languages: Python, SQL

Machine Learning & AI: Deep Learning, NLP, Generative AI

Frameworks & Libraries: PyTorch, , Hugging Face, Scikit-learn, NumPy, Pandas, MLflow

Developer Tools: Git, GitHub, Docker, Kubernetes

Specializations: MLOps, LLMOps, Deep Learning