Hiyansh Chandel

chandelhiyansh@gmail.com
+91 96802 11132 • Jaipur, India
GitHub • LinkedIn

EDUCATION

National Institute of Technology (MNIT), Jaipur **2023 – 2027**
Bachelor of Technology in Artificial Intelligence and Data Engineering
Relevant Coursework: Deep Learning, Machine Learning, Data Structures & Algorithms, Computer Vision, Natural Language Processing, Database Systems, Software Engineering
St. Xavier's School, Jaipur **2020 – 2022**
Higher Secondary Education (10th – 12th)

TECHNICAL SKILLS

Programming Languages: Python, C++, C, SQL, Bash
ML & AI Frameworks: PyTorch, TensorFlow, Scikit-learn, NumPy, Pandas, Hugging Face Transformers
Deep-Learning Focus Areas: Large Language Models (LLMs), Computer Vision, NLP, Transformer Architectures
GenAI & LLM Frameworks: LangChain, LangGraph, LlamaIndex, OpenAI API, RAG Pipeline Design
Vector Databases: Pinecone, Qdrant, Weaviate, FAISS
Cloud & MLOps: Azure ML, Azure STT, Docker, Git, API Development, Model Deployment
Additional Tools: OpenCV, Matplotlib, LangFlow, n8n, CrewAI, AutoGen, SmolAgents, Twilio API

ACADEMIC PROJECTS

LLaMA-2 Architecture Re-implementation **2025**
- Built core **Transformer** components—**attention mechanisms**, **KV caching**, and **grouped-query attention**—to internalize modern LLM design principles.
- Implemented **rotary positional embeddings (RoPE)**, **RMSNorm**, and **SwiGLU-based** feed-forward layers.
- *Tech Stack:* Python, PyTorch, Git

PaLI-Gemma Vision–Language Model **2025**
- Recreated Google's PaLI-Gemma multimodal architecture by fusing a **ViT-based visual encoder (SigLIP-style)** with a **Gemma-style causal language model** for image captioning.
- Added **masked self-attention** and a projection layer for seamless vision-to-token embedding mapping.
- *Tech Stack:* Python, PyTorch

VoiceCare: Agentic Healthcare AI System **2025**
- Designed an end-to-end conversational system for elderly assistance; processed 500+ voice queries with 92% intent-classification accuracy using **LLaMA-70B on Groq**.
- Deployed a production **RAG pipeline** on **Pinecone** to reduce response latency by 40%.
- Integrated multi-agent workflow with **Azure STT** for transcription and **Twilio** for emergency escalation, enabling family notification in under 15 s.
- *Tech Stack:* Azure STT, Groq, Pinecone, Twilio API, TTS, OpenAI API, Python, FastAPI

Micrograd: Autograd Engine Recreation **2024**
- Re-implemented Andrej Karpathy's Micrograd engine: **computational graph construction**, **automatic differentiation**, and gradient-based optimization from first principles.
- Achieved 95% accuracy on binary-classification tasks by building a lightweight training framework.
- *Tech Stack:* Python, NumPy

LEADERSHIP EXPERIENCE

Founder – MNIT-SPARK **Aug 2024 – Present**
- Established a research society focused on cutting-edge machine-learning applications; hosted a tech exhibition with 300+ participants.
- Ran an inaugural workshop for 60+ members featuring hands-on demonstrations.

Executive Lead – Data Science Club, MNIT **Oct 2024 – Present**

- Directed the merger of SPARK with the Data Science Club, boosting active membership from 45 to 150+ in three months.
- Coordinated an inter-college hackathon featuring 40+ teams.

ADDITIONAL INFORMATION

Technical Interests: Large Language Models, Multimodal Systems, Computer Vision, Reinforcement Learning, MLOps
Languages: English (Fluent), Hindi (Native)
Certifications: Deep Learning Specialization (in progress)