

## JATIN MEHRA

LinkedIn: [linkedin.com/in/jatin-mehra119](https://www.linkedin.com/in/jatin-mehra119)  
GitHub: [github.com/Jatin-Mehra119](https://github.com/Jatin-Mehra119)  
Portfolio: [jatin-mehra119.github.io](https://jatin-mehra119.github.io)

Email: [jatinmehra@outlook.in](mailto:jatinmehra@outlook.in) | Phone: +91 9910364780 | Address: Delhi, India

---

### SUMMARY

Generative AI Engineer with hands-on experience in building **LLM-powered production systems**, architecting **RAG pipelines**, and deploying **modular AI workflows** across domains like education, healthcare, and content intelligence. Proficient in **LangChain**, **FAISS**, **FastAPI**, **Streamlit**, and **containerized deployments**. Delivered scalable GenAI applications with embedded context retrieval, evaluation frameworks, and cloud-native architecture. Ranked top 1–5% in global ML challenges and published high-traffic open-source models. Passionate about designing **human-like AI systems** grounded in performance, scalability, and user empathy.

---

### SKILLS SUMMARY

**Languages & Frameworks:** Python, SQL, FastAPI, Streamlit, LangChain, LlamaIndex, Hugging Face, PyTorch, TensorFlow  
**GenAI & LLMs:** RAG, Vector DBs (FAISS, Pinecone), Prompt Engineering, Multi-agent Systems, Sentence Transformers, Groq API, OpenAI  
**DevOps & Deployment:** Docker, Linux, Git, REST APIs, Modular Pipelines, Asynchronous Processing  
**Cloud & Platforms:** GCP, Hugging Face Hub, (familiar with Databricks-style orchestration)  
**Tools:** MLflow, Whisper, PyMuPDF, TavilySearch, Power BI  
**Soft Skills:** Cross-functional Collaboration, Product Thinking, Rapid Prototyping

---

### RELEVANT PROJECTS

#### *PDF Insight PRO – Agentic RAG PDF Chatbot | Aug 2024-June 2025 | [GitHub](#) | [Live Demo](#)*

**LangChain** | **FAISS** | **ChatGroq** | **Streamlit** | **Docker** | **FastAPI**

- Designed an **agentic RAG-based system** integrating FAISS, Sentence Transformers, and **multi-agent LLM reasoning** using LangChain.
  - Enabled **semantic search with sub-second latency** using FAISS vector database and embedded document chunking pipeline achieving **85.3% evaluation accuracy** and **0.852 mean semantic similarity score** across document analysis tasks.
  - Used **LangChain agents** and tools for enhanced LLM orchestration and context awareness, with web augmentation via Tavily API.
  - Deployed modular components using Docker; prepared for real-world multi-client (PDF + Android) usage.
- 

#### *AI-Powered Podcast-to-Blog Generator | April 2025-May 2025 | [GitHub](#) | [Live Demo](#)*

**Groq API** | **LangChain** | **FastAPI** | **Modular Pipelines** | **Whisper** | **Streamlit** | **Llama3/4**

- Built a cloud-native pipeline using **Whisper Large v3 Turbo** + LLMs to transcribe and convert multilingual podcast audio into **SEO-ready blogs, FAQs, and newsletters**.
  - Integrated **external knowledge grounding** using TavilySearch and asynchronous processing for scalable, human-like content generation.
  - Delivered in-browser app with real-time blog generation—highlighting **cloud-scale LLM orchestration** using modular APIs deployed on Hugging Face Spaces using a **CI/CD pipeline powered by GitHub Actions**.
- 

#### *Plagiarism Detector using Fine-tuned LLM | Nov 2024-Jan 2025 | [GitHub](#) | [Model card](#) | [Live Demo](#)*

**SmolLM** | **Hugging Face Transformers** | **Streamlit** | **Docker** | **PyTorch**

- Fine-tuned a lightweight LLM (SmolLM 135M Parameter) on the MIT Plagiarism Dataset and deployed as a **containerized Streamlit app** with 96% accuracy on test set.
  - Achieved 1200+ monthly model downloads, demonstrating strong GenAI deployment and real-world adoption.
- 

#### *Nexar Dashcam Collision Predictor – Kaggle Top 1% | May 2025 | [GitHub](#) | [Model card](#) | [Live Demo](#)*

**VideoMAE-2** | **PyTorch** | **Hugging Face Transformers** | **TimeSformer**

- Fine-tuned transformer-based models (VideoMAE-2) for video classification with **0.80mAP Scores**.
  - Extracted feature using TimeSformer.
  - Built a modular pipeline with reusable training components, **tracking via MLflow** and leaderboard ranking (11th place).
- 

### Certificates

- NVIDIA – Building RAG Agents with LLMs**  
Hands-on certification focused on Retrieval-Augmented Generation (RAG) pipelines, LLM orchestration, and agent design.
  - Hugging Face – Agents Course Certification**  
Covered tool calling, agent workflows, multi-step reasoning, and integration using Transformers and LangChain.
  - Harvard Business School Online – Presentation, Business Planning, and Digital Intelligence**  
Developed executive-level skills in product storytelling, stakeholder alignment, and digital transformation strategies — complementing technical AI projects with strong product thinking.
- 

### Education

**Masters of Science in Data Science** | 2025-2027

Chandigarh University, Punjab

Relevant course work- Deep learning, Machine Learning, SQL Programming, NLP, Data Engineering, Big Data, Calculus & Linear Algebra

**Bachelors of Computer Applications** | 2022-2025

Chandigarh University, Punjab

Grade-7.89

Relevant course work- Data Structures and algorithms, Python & Java Programming, Cloud Computing, System Designs, Probability & Statistics