Mohammad Ishaan Hasan Ansari

+91-9634038353 | mihansari.jh@gmail.com | ishaanansari.ai
ishaanansari | ishaan-ansari | iamihansari
New Delhi, India

Summary

Machine Learning Engineer with over 2 years of experience working at the intersection of computer vision and large language models.

Experience

- **Think Future Technologies Pvt. Ltd [🌐]** *February 2024 - Present*
  Software Engineer L1 (Machine Learning) Gurugram, India
  - **Gen-AI Health Coach** Developed a Generative AI–powered coach that automates routine doctor–patient interactions, deployed across 15 clinics.
  - **Food Logging** image-based calorie-detection feature achieving 94% estimation accuracy, optimized inference pipeline to cut latency by 65% (32 s → 11 s).
  - **Next meal suggestion** Built a personalized meal-suggestion system that increased users' daily caloric-goal compliance from 68% to 83%, driving a 12% uptick in weekly active users.
  - **Doctor clone** Created an interactive "Doctor Clone" agent capable of parsing 100+ medical reports per day with 90% insight accuracy, reducing physicians' report-review workload by 40%.
  - **Digital trends** Designed visual dashboards used by over **200** clinicians and **1,500** patients to track biomarker trends—enabling a **20%** faster identification of abnormal readings and supporting proactive interventions.
  - **HIPPA compliance** Implemented end-to-end PII anonymization and secured third-party integrations, resulting in zero compliance incidents across two external audits and safeguarding data for **5,000+** users.
  - **Document-Parsing Search Optimization:** Redesigning search algorithms to index and retrieve document sections 3× faster, cutting average parse time from 6 s to 2 s %.
  - **API-Call Reduction Logic:** Implementing relevance-based chunking to limit OpenAI API calls to critical text segments—reducing total monthly calls by 60% (from 10,000 to 4,000), while maintaining 95% answer accuracy.
  - **Layout Detection Model:** Developed a layout detection model to classify and segment newspaper content into articles, images, titles, and subtitles. Fine-tuned the model to achieve an accuracy exceeding **97%**.
  - **Optical Character Recognition (OCR):** Implemented OCR to extract textual data from perso-arabic images and documents, achieving overall accuracy of **83%**.
  - **Language Detection:** Built and fine-tuned a language detection model using FastText, optimized for multiple Perso-Arabic languages.
  - **Language Translation:** Conducted R&D on translation models like **NLLB** and **M2M100**. Developed training, fine-tuning, and evaluation pipelines for multilingual translation tasks.
  - **Proctoring System:** To ensure fairness during remote assessments developed a proctoring system, improved accuracy by **22%** and reduced latency by **40%**.
  - **Recruitment system:** Trained **Machine Learning Ranking Models** for ranking developers for jobs using **Gradient Boosting Decision Tree (GBDT)** and **Logistic Regression** models using **pandas, numpy, scipy, scikit-learn** and **lightGBM** frameworks
  - Led the **GenAI** project initiatives based on OpenAI **GPT-3. GPT-4** and open-source models hosted using **Text Generation Inference (TGI)** framework
  - **Fine tuned Llama 3** using **PEFT LORA** technique, **Python**, **Huggingface** frameworks, to reduce hallucinations in a chat application
  - Deployed **Mistral based GenAI** solution on **Microsoft Azure T4 GPU** in **Docker container**
  - **Retrieval Augment Generation:** Developed a RAG pipeline using **LangChain framework, FAISS, ChromaDB, Pinecone and FastAPI**
  - Developed an end-to-end **GPT-3.5 Fine tuning** pipeline to enhance a coding-practice platform's question generation.
  - Trained **Vision Transformer** and **Convolutional Neural Network (CNN) model** using **PyTorch** framework on **(GPU)**, for real-time surveillance
  - Developed a heuristic approach for masking first four digits of government-id proof, achieved around **92%** of accuracy score
- **Think Future Technologies Pvt. Ltd [🌐]** *August 2023 - January 2024*
  Software Trainee (Machine Learning) Gurugram, India
  - Designed **CNN** architecture which reduces HEVC video compression artifacts by **6%** and leveraged GoogleNet's Inception blocks to reduce the number of learnable parameters by **50%**
  - Experimented with **GAN** architecture using 3 different loss functions (perceptual loss, smooth loss, and MSE loss) and performed hyperparameter tuning.

Research & Projects

- **MIRAGE: [Multimodal RAG framework for clinically grounded medical reasoning]** *April 2025*
  Tools: [Python, PyTorch, LangChain, BioViL, MedCLIP, FAISS, Med-PaLM, Python] [GitHub]
  - Fused medical-image embeddings with retrieved clinical evidence, raising factual accuracy by 22% and cutting hallucinations 28%.
  - calibrated Rethinking and Rearrangement module that dynamically adjusts the quantity of retrieved contexts to manage factual risk.
  - Orchestrated AI agents through LangChain chains, achieving transparent, step-wise clinical reasoning in under 300 ms per query.
  - Implemented continuous fine-tuning of MedCLIP/BioViL encoders on EMR and vital-sign streams, improving critical-case detection by 18%
  - Evaluation on the MedVQA dataset shows reduced factual errors when using online search augmentation and adaptive chain-of-thought (CoT) verification.
- **Nudge: [Personalized product recommendation engine]** *October 2024*
  Tools: [Python, Pandas, Langchain, RAG, Streamlit] [GitHub]
  - Built a query engine using Chroma as Vector DB and RAG using LangChain for semantic search on BigBasket's product data.
  - Utilized BAAI's general embedding (bge) for creating vector embeddings and Llama-3.1-70B as LLM for accurate, context-driven responses. Deployed the solution as a Streamlit API for seamless real-time querying.
- **GeoMorph: [Satellite-to-Map Translation using Pix2Pix GAN]** *July 2023*
  Tools: [Python, PyTorch, NumPy, matplotlib, Streamlit] [GitHub]
  - Implemented a Pix2Pix GAN for mapping satellite/aerial images to it equivalent Map-View image (as on Google maps)
  - Built an optimized data pipeline processing 257 MB of imagery, reducing training time by 25% over 50 epochs.
  - Deployed the model on HuggingFace platform for realtime Satellite-to-Maps image translation task
- **Captionix: [Generating Captions for images using CNN & LSTM and attention.]** *March 2023*
  Tools: [Python, PyTorch, Numpy, matplotlib, FastAPI, Github Actions, Docker] [GitHub]
  - Implemented CNN encoder and LSTM decoder to extract textual cues from images, leveraging Flickr8k dataset.
  - Compared performances of InceptionV3 and EfficientNetV2, achieving the best BLEU-1 score of 0.57.
  - Developed a web application utilizing Flask and deployed the model on HuggingFace to generate real-time captions.
- **TGIC: [Text-Guided Image Clustering.]** *August 2022*
  Tools: [Deep ImageNet Models, Transfer Learning, Sentence Embeddings, Captioning, Visual Question Answering] [GitHub]
  - Benchmarked classical features (SIFT, Canny, Color Histograms, LBP, HOG) against deep embeddings (ResNet-50, EfficientNet-B0) on Food-101, achieving a baseline ARI of 0.62.
  - Generated semantic captions via BLIP and VQA features via ViLT, then fused SBERT sentence embeddings with image vectors to enrich clustering inputs.
  - Fine-tuned BLIP on domain captions and applied DBSCAN to combined embeddings, boosting ARI to 0.8041 and improving cluster purity by 18%.

Education

- **Jamia Hamdard University** *August 2019 - July 2023*
  Bachelor of Technology in Computer Science - 1st division with distinction New Delhi, India

Skills

- **Programming Languages:** Python, C, C++, JavaScript, SQL
- **Web Technologies:** FastAPI, Django, React, Next.js, Node.js
- **Database Systems:** PostgreSQL, MongoDB, MySQL
- **Data Science & Machine Learning:** PyTorch, TensorFlow, scikit-learn, Exploratory Data Analysis, Gradient Boosting, Decision Trees, Clustering, Regression, Statistical Analysis
- **Cloud Technologies:** AWS, Google Cloud Platform, Microsoft Azure
- **DevOps & Version Control:** Docker, Git, Jenkins, GitHub Actions
- **Specialized Area:** Natural Language Processing, LLMs, Computer Vision
- **Other Tools & Technologies:** Redis, RabbitMQ, Apache Kafka, MLflow, Apache Airflow
- **Research Skills:** Literature Review, Experimental Design, Data Visualization, Algorithm Development, Model Evaluation