Abhishek Singh

Email: abksingh2004@gmail.com — Phone: +91 9650950014
LinkedIn: linkedin.com/in/abhishek-singh202220260204
GitHub: github.com/Abhi2april
Portfolio: abhi2april.github.io/portfolio/

Summary

A Junior in Computer Science Engineering at Bennett University with 2+ years of applied AI/ML experience through academic projects and internships. Proficient in Pytorch framework. Developed 5+ projects in Medical Imaging, Finance and Educational AI. My coursework projects and Papers -abhi2april.github.io/portfolio/

Skills

- Programming: Python, C++, SQL
- Machine Learning: TensorFlow, PyTorch, Keras, Scikit-learn
- Generative AI: Diffusion Models, Model Fine-tuning, Retrieval-Augmented Generation (RAG)
- NLP: LLMs, Transformer models, Text embeddings, Sentiment analysis
- Data Structure and Algorithms
- Dev Tools: Git, Streamlit, Google Colab, Jupyter, Visual Studio
- Other: Team Management, Active Listening, Multitasking

Work Experience

NLP Team Lead — Artificial Intelligence Society, Bennett University August 2024–Present

- Led team of 5 developers in building **Minstral-AI-based Finetuned profanity filter API → 2.3s avg response time** (WhatsApp/YouTube integration pipeline)
- Designed **Disease Recognizer** using `all-MiniLM-L6-v2` embeddings + logistic regression → **87% diagnosis accuracy** on symptom input
- Architected ML training pipeline: Curated 50k+ symptom-disease pairs → K-means clustering → 15% faster inference vs traditional classifiers

AI Research Intern — LLM Specialist, IIIT Dharwad April–August 2024

- Fine-tuned **LLaMA3** via **LoRA adapters → 92% relevance** in curriculum-aligned QGen (NeuralIPS-2021/NCERT dataset)
- Developed a RAG pipeline: Sentence Transformers/all-MiniLM-L6-v2 + ChromaDB embedding vectorization + NCERT Book Text + LangChain agent.
- Reduced GPU memory usage by **40%** through 8-bit quantization (NVIDIA P100 clusters)

Projects

3D U-Net Tumor Segmentation w/Spatial Attention GitHub
(TensorFlow, Medical Imaging)

- Developed 3D U-Net model with spatial attention layers in Conv3D blocks → **85% mIoU** on BraTS dataset
- Trained using custom Dice-Cross entropy loss → **18% fewer false positives** vs baseline U-Net
- Impact: Advanced tumour boundary delineation for medical imaging workflows

Disease Diagnosis Engine GitHub
(Scikit-learn, Hugging Face, NLP, StreamLit)

- Encoded 1200+ patient symptoms via `all-MiniLM-L6-v2` → 512-dim embeddings
- Trained logistic regression + K-means clustering → **87% accuracy** in symptom-to-disease classification

Curriculum-Aligned Question Generator GitHub
(PyTorch, LLaMA3, Minstral-7B, RAG, Fine-Tuning)

- **Fine-Tuning Approach**:
  - Optimized 8-bllion Parameter LLaMA-3 model with Low Rank Adaptation for efficient inference → **92% relevance** in Math/Science questions
  - Reduced GPU P100 training costs by **40%** via 8-bit quantization
- **RAG Pipeline**:
  - Structured using ChromaDB for vector embedding storage, utilising PyMuPDF for NCERT textbooks to achieve a retrieval speed comparable to Facebook AI Similarity Search.
  - LangChain-driven QA system → **around 95% accurate syllabus alignment** (CBSE/ICSE standards)

Education

Bennett University September 2022 - April 2026
B.Tech in Computer Software Engineering 3.4/4 GPA