Soumedhik Bharati

+91 8240947878 | soumedhikbharati@gmail.com | Linkedin | Github | West Bengal, India

Experience

Indian Institute of Technology Kharagpur — Kharagpur, India
Research Intern (Completion Certificate) — *May 2025 – July 2025*
- Architected a multi-stage caching IR reranking pipeline to accelerate large-scale, reproducible LLM experiments, improving throughput for complex query loads.
- Devised a novel parallelism strategy that slashed inference time by **33%** and **66%** over top-down and sliding window approaches, respectively, while improving or maintaining core IR metrics.
- Validated performance by executing over **2400** large-scale experiments on the Param Vidya HPC cluster using SLURM.

Xu Lab, Carnegie Mellon University — Pittsburgh, PA (Remote)
Researcher — *Sept 2025 – Present*
- Engineered a few-shot transfer learning pipeline by fine-tuning a 1.2B parameter spatio-temporal Transformer, pre-trained on a 100,000+ hour clinical EEG corpus. Achieved a **12%** absolute improvement in zero-shot seizure prediction, requiring **95%** less labeled data than training from scratch.
- Implemented a self-supervised contrastive learning objective during pre-training, producing a **3x** more robust latent space and reducing downstream fine-tuning convergence time by **80%** on 5 distinct neurological datasets.

Sister Nivedita University — Kolkata, India
Undergraduate Researcher — *Mar 2024 – Present*
- Engineered CADET, a BiLSTM architecture with multi-head attention, setting a new SOTA with a QWK of **0.98** on the ASAP essay dataset (under review).
- Pioneered **HCAT-Net**, a novel ResNet-Transformer architecture that delivered **99.8%** accuracy for EEG emotion classification (presented at CIACON 2025).
- Secured an **82% success rate** in optimizing employee training curricula by formulating a novel RL agent to solve the multi-objective task under a hard budget constraint.
- Cut trainable parameters by **55%** while achieving SOTA on the RESIDE benchmark by optimizing a multi-scale UNet with Mamba state-space models.

Exalt.ai — Remote
Product Engineer — *June 2025 – Present*
- Boosted contextual relevance by 20% for high-traffic news summarization via production RAG with hybrid BM25–Faiss retrieval.
- Improved model performance by **22%** across summarization and sentiment analysis tasks by fine-tuning LLMs with parameter-efficient fine-tuning (PEFT) techniques.
- Reduced model size by **40%** for edge deployment via combined quantization and distillation, retaining **90%** of original performance.

Raapid.ai — Remote
R&D Intern (Joining Letter) — *Apr. 2025 – June 2025*
- Improved HCC code extraction accuracy by **12%** using novel deep learning for unstructured clinical notes.
- Reduced ingestion and preprocessing time by **25%** on large-scale medical data via parallelism and optimized queries.

University of Lille (Collaboration) — Remote
Collaborative Researcher — *Feb 2025 – Present*
- Designed and implemented a deep learning surrogate model to map Hamiltonian control parameters to the final evolved quantum state, yielding an R^2 of **0.94** against the ground-truth solver (manuscript under review).

Education

Sister Nivedita University — Kolkata, India
B.Tech in Computer Science Engineering — CGPA: 8.68/10.0 — *Sept. 2022 – Sept. 2026 (Expected)*
- **Leadership:** Machine Learning Lead, SKEPSIS; Core Technical Team ML Lead, Google Developer Group (GDG) SNU.

Awards & Conferences

- **2nd Prize**, BRICS International Vocational Skills Offline Competition 2024, Shandong, China. (Certificate)
- **Top 3 Teams**, ICDMAI Offline Hackathon 2025 (from 1000+ teams). (Certificate)
- **1st Place**, SAP ICOE Hackathon 2024 (from 400+ teams). (Certificate)
- **1st Place**, Intel OneAPI Hackathon 2024 (100+ teams) (Certificate) and Pragati 2k24 ML Mania Competition (Certificate).
- **Selected for India Regional Bootcamp**, Google Solution Challenge 2024. (Certificate)
- **Best Presenter Award**, IEEE CIACON 2025 – Presented research on HCAT-Net for EEG Emotion Classification. (Certificate)
- **Attendee**, ICDMAI 2025 & BRICS International Big Data & AI Conference and AI Working Committee, Chongqing, China.

Skills

Languages: Python, R, C, C++, SQL
Frameworks & Libraries: PyTorch, TensorFlow, Hugging Face, Faiss, Elasticsearch, Scikit-learn, ONNX, Weights & Biases (W&B)
Research & Dev Tools: LaTeX, Git/GitHub, SLURM, Tableau, Power BI, Zotero
Specializations: NLP, CV, Time Series Forecasting, Reinforcement Learning, Information Retrieval, RAG, Vector Databases

Projects

Automatic Essay Grading System | *GitHub* — Hackathon Winner
- Attained SOTA performance (QWK of **0.674**) on the ASAP dataset by building a BiLSTM model with multi-head attention, outperforming previous models by **8.3%** in coherence assessment.

Assistive System for Blind People | *GitHub* — Hackathon Winner
- Reached **99.4%** accuracy at **25 FPS** on edge devices by integrating a multi-task vision system with YOLOv9, MIDAS, and ResNet50 for real-time object and currency detection.

Image-to-Music Synthesis System | *Live Demo*
- Synthesized music with high tonal and emotional alignment to visual input by fabricating a modular image-to-music pipeline using a Vision Transformer and diffusion models.

Additional Projects: Image Enhancement Autoencoder, Multi-Modal Face Tracking, Book Recommendation System