

Ajoy Sarker

Machine Learning Researcher — Robust AI, Reliable LLMs & Low-Resource NLP

ajoy.stu20181@juniv.edu | [Google Scholar](#) | [GitHub](#) | [Hugging Face](#)

[linkedin.com/in/ajoyrju](https://www.linkedin.com/in/ajoyrju) | [Portfolio](#) | Dhaka, Bangladesh

RESEARCH PROFILE

Researcher focused on robust and trustworthy machine learning, with interests in reliable LLMs and low-resource NLP. My work investigates model reliability under distribution shift, limited supervision, and noisy real-world conditions, spanning Bengali NLP, LLM adaptation and evaluation, and cross-subject EEG modeling.

RESEARCH INTERESTS

Robust & Trustworthy AI • **Reliable LLMs** • **Low-Resource NLP** • **Model Evaluation** • **Distribution Shift**

EDUCATION

M.Sc. in Computer Science & Engineering (AI/ML)

May 2024 – Sept 2025

Jahangirnagar University

GPA: 3.78 / 4.00

- Thesis: Bengali abusive-language detection using stacked ensembles with word- and character-level TF-IDF features.

B.Sc. in Computer Science & Engineering

Feb 2019 – Mar 2024

Jahangirnagar University

GPA: 3.68 / 4.00

RESEARCH EXPERIENCE

Bengali Abusive Language Detection [PDF]

2023 – 2025

Jahangirnagar University

- Studied **15-class abusive-language detection** in Bengali and expanded the dataset to approximately **58K samples**.
- Combined word- and character-level TF-IDF features in a stacked ensemble, reaching **91.2% accuracy** and **90.7% macro-F1—1.9 percentage points** above the previous result.

LLM Fine-Tuning for Low-Resource NLP [Hugging Face]

2024 – Present

Independent / HuggingFace

- Investigated **parameter-efficient adaptation** of language models for Bengali mathematical reasoning using **LoRA/PEFT**.
- Evaluated answer correctness on domain-specific benchmarks and developed a dual-encoder model for Bengali semantic similarity and retrieval; released the resulting models on Hugging Face.

EEG-Based Emotion Modeling [PDF]

2024 – Present

Jahangirnagar University

- Investigated **cross-subject generalization** of EEG-based emotion recognition under noisy recording conditions.
- Developed a preprocessing pipeline using filtering, ICA, and artifact removal, then compared CNN, RNN, and Transformer models across subjects and sessions on DEAP and MAHNOB-HCI.

Road Accident Severity Analysis and Prediction [PDF]

2023

Collaboration with the Accident Research Institute, BUET — Unpublished

- Studied accident occurrence and severity using police records from all 64 districts of Bangladesh, covering 2001–2020.
- Compared classification and regression methods; CatBoost produced the highest classification **F1 score of 0.6476**, while tuned Gradient Boosting produced the highest regression R^2 of **0.0818**.

PUBLICATIONS & MANUSCRIPTS

- **A. Sarker**, M. Z. Rahman. *Abusive Language Detection in Bengali Social Media*. WRFER 2026. **Accepted**. [PDF]
- **A. Sarker**, S. N. Riya, I. Islam. *Enhancing Emotion Recognition from EEG Signals*. **Under review**. [PDF]

RESEARCH PROJECTS & OPEN-SOURCE RESEARCH

- **Bangla Math LLMs** [Hugging Face] — Bengali mathematical-reasoning models fine-tuned with LoRA/PEFT and evaluated on structured benchmarks.
- **Bangla Sarcasm Bi-Encoder** [Hugging Face] — Dual-encoder model for Bengali semantic similarity and retrieval.
- **Intent-Based Hybrid Search** [GitHub] — Combined vector and keyword retrieval with LangGraph orchestration, structured outputs, and guardrails; won 1st place at the Brain Station 23 AI Hackathon 2025.
- **BD Document Parser** [Live] — Built schema-constrained extraction for Bangladeshi identity and financial documents using Gemini Vision.

INDUSTRY RESEARCH / ENGINEERING EXPERIENCE

Software Engineer (AI/ML)

2024 – Present

Brain Station 23, Dhaka

- Developed RAG pipelines over structured knowledge bases and evaluated retrieval and output reliability, reaching **92% retrieval accuracy** with latency below **200 ms**.
- Built LangGraph workflows with structured outputs, validation, and guardrails to reject malformed or unsafe responses; the project won 1st place at the company AI hackathon in 2025.
- Built Gemini Vision pipelines for schema-based KYC document extraction and validation, processing more than **100K records per day**; batching and inference changes reduced runtime by approximately **50%**.

HONORS & AWARDS

- NST Fellowship — National Science & Technology fellowship.
- University Scholarship (×4) — Awarded for academic performance.
- AI Hackathon 2025 — 1st place for a LangGraph-based hybrid retrieval system.
- CodeSamurai 2022 — Ranked 29th among more than 500 teams.

TECHNICAL SKILLS

- **Research Methods:** Dataset construction, preprocessing, feature engineering, model comparison, representation learning, benchmark evaluation, time-series and cross-subject evaluation
- **Machine Learning:** PyTorch, TensorFlow, scikit-learn, XGBoost, CatBoost
- **NLP & LLMs:** Transformers, Hugging Face, LoRA/PEFT, RAG, LangGraph, retrieval and output evaluation
- **Systems:** Python, FastAPI, PostgreSQL/pgvector, AWS, MLOps