# Soumik Deb Niloy

Computer Science Researcher

d19niloy@gmail.com | linkedin.com/in/soumik-deb-niloy-22ab431a5 | github.com/S18-Niloy | +880 17963 11815
Dhaka, Bangladesh

## Experience

**Researcher (Project Based)** *Jan 2025 - Ongoing*
*BRAC University, Dhaka, Bangladesh*

- Conduct interdisciplinary research applying AI/ML to biological data analysis and NLP tasks.
- Develop models for protein/peptide classification, gene expression analysis, and cancer diagnostics using machine learning and graph neural networks.
- Work on multilingual NLP projects, including text summarization, sentiment analysis, and code generation.
- Contribute to conference/journal manuscripts and collaborative research projects.

**Research Asistant** *Jan 2025 - Ongoing*
*Center for Computational and Data Sciences, Independent University, Dhaka, Bangladesh*

- Conduct in-depth literature reviews and data collection to support research projects in computational biology, NLP, and medical image analysis.
- Develop and implement algorithms and machine learning models using Python, PyTorch, and TensorFlow to analyze complex biological and clinical datasets.
- Design experiments, prepare datasets, and perform statistical analyses to validate research hypotheses and improve model accuracy.
- Collaborate with multidisciplinary teams to interpret results, prepare technical reports, presentations, and research publications.

**Natural Language Processing Engineer** *Feb 2024 - Jan 2025*
*Startup; Synteo.AI, Dhaka*

- Developed intelligent chatbots using Retrieval-Augmented Generation (RAG) and MLDotNet to enable context-aware conversational AI.
- Conducted research on AI hallucination to identify, analyze, and mitigate erroneous outputs in large language models.
- Designed and implemented fintech-based NLP applications to automate and enhance financial data processing and user interactions.

## Technical Skills

**Data Science:** NumPy, Pandas, Matplotlib, Seaborn
**AI Frameworks:** Tensorflow, Pytorch, Sci-kit learn, Huggingface transformers, Spacy, NLTK, OpenCV, PyTorch Geometric (PyG), Keras
**Mathematics and Statistics:** Linear Algebra, Calculus, Matrix, Probablity, Neumerical Methods
**Programming:** Python, Java, C, C#
**Data Scraping:** BeautifulSoup, Selenium
**Miscellaneous:** Git, Latec

## Projects

**AI-Driven Financial Fraud Detection (FinTech GNN Project)**

- Stack: PyTorch Geometric, DGL, XGBoost, Neo4j, NetworkX
- Developed a Graph Neural Network (GNN) to detect fraudulent financial transactions by modeling complex relationships among users, devices, IPs, and merchants.
- Implemented graph-based anomaly detection and integrated explainable AI techniques for model interpretability.
- Leveraged Neo4j for graph storage and visualization, and NetworkX for network topology insights.
- Enhanced fraud monitoring via real-time alert streaming using Kafka and an interactive LangChain-powered dashboard.
- Demonstrated strong skills in graph data modeling, fraud analytics, and **interpretable deep learning**.

**EfficientLiteNetB0: Lightweight CNN-Transformer for Video Processing**

- Developed a hybrid CNN+Transformer model for video frame analysis and image classification with low GFLOPs (0.7) and 5MB model size.
- Implemented dataset preprocessing, video frame extraction, and Grad-CAM visualization for model interpretability.

- Achieved 99.67% training accuracy, 94.24% validation accuracy, and 45ms inference latency, outperforming T2T-ViT, CSWin, and SqueezeNet.
- Designed and optimized a complete video testing pipeline for edge devices and mobile deployment.
- *Technologies: Python, PyTorch, OpenCV, Grad-CAM, CNN, Transformer, Video Processing.*

**Second Brain — Personal Knowledge Curator**

- Developed a personal knowledge management system that ingests PDFs and URLs, generates embeddings using local SentenceTransformers, and stores them in Chroma for efficient retrieval.
- Built a Streamlit UI with a FastAPI backend to enable retrieval-augmented generation (RAG) queries on personal data.
- Implemented a fully local setup with no cloud dependencies, supporting offline operation and private data handling.
- Demonstrated proficiency in embedding-based search, vector databases, and full-stack Python deployment.

**Products Chatbot API**

- Developed a RESTful Chatbot API using FastAPI to provide human-like responses about product details.
- Integrated Groq LLM for natural language understanding and RAG-style reasoning to fetch product data from DummyJSON.
- Designed modular project structure with separate layers: API routes, services, models, utils, and configuration.
- Implemented endpoints:
  - `GET /api/products` – Fetch all product data.
  - `POST /api/chat` – Process customer messages and respond intelligently.
- Managed API secrets securely using .env files and `python-dotenv`.
- Tested API endpoints via Swagger UI and ensured robust error handling.

**Python Code Generation From Bangla Instructions**

- Iterative Agentic AI system to generate python code from bengali instructions without translator.
- A lightweight system using Ollama server
- Generator, Checker and Debugger are also created to generate the best response
- Possible corner cases for a coding problem is also set to generate best response
- Notable LLMs: Gemini-1.5-flash-updated, qwen1.5, gpt-oss-20B, StarCoder, Gemma
- Best test performence: *Code generator:* Gemini-1.5-flash-update, *Checker and debugger:* gpt-oss

**OCR - HandWritten Prescription**

- Applied PaddleOCR for multilingual handwritten text recognition.
- Designed post-processing algorithms for spell correction and drug name validation.
- Built an intuitive web/mobile-compatible interface for easy prescription uploads.
- Improved processing speed and accuracy through image pre-processing (noise removal, contrast enhancement).
- Enabled digitized storage and retrieval of prescriptions for pharmacies and healthcare providers.

**Mitigation of Hallucination and Interpretations of Self Attention of Mistral 7B**
Documentation

- Reproduce and Customize the COVE pipeline without external resources.
- The aim was to improve the generation of LLM without using RAG
- Dataset "RACE" was used
- Finetuned Mistral 7B using 4-bit quantization
- Analyzed attention layers in eight labels (Worst of worst, worst, worse, bad, good, better, best, best of best)
- Evaluated and measured Hallucinated AI generations
- Interpreted the causes behind LLM hallucinated generations both before and after correction

**Cervix Cancer Region Analysis by Ensembling Customized Graph Neural Networks**

- Designed a hybrid GNN combining GCN and GAT for medical image-based graph classification.
- Integrated heterogeneous data sources by constructing graphs from image features and clinical metadata.
- Implemented attention-based edge analysis to identify key relationships affecting classification.
- Applied beam search to extract best and worst attention edges for model interpretability.

- Developed visualization tools using NetworkX and PyG to interpret graph structures and attention maps.
- Achieved improved classification accuracy on cervix cancer images using SwedeFinal-based labels.
- Achieved highest performance on the selected dataset

**KEMP-PIP: A feature-fusion-based approach for pro-inflammatory peptide prediction**
Github Link Web Surface
- Features: K-Mer, ESM embedding, ModlAMP, PhysioChemical
- Zero-variance filtering is used for unnecessary feature elimination and co-efficient pruning is applied to peek best threshold
- Logistic Regression is used as classifier
- Deployed using Gradio

**Cell clustering based on the pattern of the genome sequence**

- Each cell is treated as Node and edge connection is created by
- Self-supervised Graph clustering was used to cluster nodes
- ARI value: 0.4

**AI Based ERP Solution**

- Developed AI-driven modules for demand forecasting and inventory optimization.
- Integrated predictive maintenance models to reduce downtime and costs.
- Built NLP-based chatbots for enhanced user support within ERP systems.
- Automated data analytics and reporting to enable real-time business insights.
- Designed scalable and modular architecture for seamless ERP integration.

**Automated Text Summarization - Web App**
Github Link
- CNN daily news dataset from kaggle was used and one thousand data were used for device space.
- Fine-tuned Bert-base model.
- Deployed using Streamlit

**Vegetation Plot Segmentation from Aerial Images**
Github Link
- Satelite dataset was collected from kaggle
- Designed a hybrid vision model using UNet, UNet++, ResUNet
- Test accuracy : 77% (Checked using Respberry Pi on real data)

## Education

**B.Sc., Computer Science** *Status: Graduate*
BRAC University, Dhaka, Bangladesh

**Relevant Coursework:** Data Structures and Algorithms, Software Engineering, Artificial Intelligence,Natural Language Processing, Computer Architecture,Linear Algebra,Automata And Computablity,Numerical methods,Operating Systems, Distributed Computing , Simulation and modelling , Quantum computing , Computer network , Compiler Design , Machine vision

**Academic Extracurricular Activity:**
Senior Member of "AI and Autonomous Team" at BRACU Mars Rover Team, Mongol Tori.
Here, I was engaged to
1. Build models for rock detection, mechanical tool detection and rocket landing tool detection using YOLOv5 and TensorFlow.
2. Tested models on Jetson Nano and Raspberry Pi.

## Publications and Research

1. Taki; S.M. Abrar Mustakim, Kar; Showmick, A.M., Niloy; Soumik Deb, Rakib; Mazharul Islam, & Biswas; Abdullah Al Nahid (2023). **Mitigation of Hallucination and Interpretations of Self Attention of Mistral 7B AI to Analyze and Visualize Context Understanding Ability of Large Language**

**Models.** Handle: 10361/22762.
Available at `https://dspace.bracu.ac.bd:8443/xmlui/handle/10361/22762`

2. Soumik Deb Niloy, Md. Fahmid Ul Alam Juboraj, Swakkhar Shatabda (2025). **KEMP-PIP: A Feature-fusion Based Approach for Pro-inflammatory Peptide Prediction. (Under Review)**
Web Surface Journal Link

3. *Md. Fahmid-Ul-Alam Juboraj, Soumik Deb Niloy, Mahbub-E-Sobhani Himel, Farig Yousuf Sadeque.*
**BRACU_CL at BLP-2025: CodeMist: A Transformer-Based Framework for Bangla Instruction-to-Code Generation.**
**Under Review at AACL 2025 (Bangla Language Processing Workshop)**.
Workshop Details

## Honors & Achievements

– **5th Place** – *BLP Workshop 2025: Code Generation in Bengali*, organized by **IJCNLP–AACL**; among **152 participants**.
Field: *LLMs, Prompt Engineering, Computational Linguistics*

## References

**Dr. Swakkhar Shatabda**
Professor, Dept. of Computer Science And Engineering, BRAC University, Dhaka
Email: swakkhar.shatabda@bracu.ac.bd

**Dr. Farig Yousuf Sadeque**
Associate Professor, Dept. of Computer Science And Engineering, BRAC University, Dhaka
Email: farig.sadeque@bracu.ac.bd