Md. Shahidul Salim

CSE, KUET, Khulna, Bangladesh
ss@cse.kuet.ac.bd Github Hugging Face Portfolio Linkedin Google Scholar

Job Experience

July 2022 – Present — **Faculty member,** Department of CSE, Khulna University of Engineering & Technology (KUET), Bangladesh

April 2022 – June 2022 — **Part-time Faculty member,** Department of CSE, Khulna University of Engineering & Technology (KUET), Bangladesh

April 2021 – March 2022 — **Faculty member,** Department of CSE, Uttara University, Bangladesh

Research Interest

- Natural language processing - Transformer models (Translation, Summarization, Text generation, Conversational Question answering), Large language models
- Machine learning, Deep learning, Multivariate and univariate time series
- Generative AI - GAN, Diffusion

Education

2016 – 2020 — **B.Sc. in Computer Science and Engineering,** Khulna University of Engineering & Technology (KUET),Bangladesh – **CGPA 3.86 out of 4.00** (4th position in my department)

Research Publications

Journal Articles

1. Md. Shahidul Salim, S. I. H. (2024). An applied statistics dataset for human vs ai-generated answer classification. *Data in Brief*.
2. Saad, A. M., Mahi, U. N., Salim, M. S., & Hossain, S. I. (2024). Bangla news article dataset. *Data in Brief*.
3. Ashiqussalehin, M., Jahan, K. N., Rahaman, M. A., & Salim, M. S. (2022). Human abnormal behavior detection using convolution neural network. *Specialusis Ugdymas*, *1*(43), 4076–4083.

Conference Proceedings

1. Bose, D., & Salim, M. S. (2024). Suggesting bengali words using masked language model. In *3rd international conference on computing advancements (icca)*.
2. Hossain, L., Hossain, I., Salim, M. S., Raju, S. M. T. U., & Saha, J. (2023). A novel technique for classification of motor imagery eeg signal based on deep learning approaches. In *Proceedings of the 2nd international conference on big data, iot and machine learning (bim 2023)*. (Accepted).
3. Nabil, A., Das, d., Salim, M. S., Arifeen, S., & Fattah, H. M. A. (2023). Bangla emergency post classification on social media using transformer based bert models. In *6th international conference on electrical information and communication technology (eict 2023)*. (Accepted).
4. Promi, R. T. H., Nazri, R. A., Salim, M. S., & Raju, S. M. T. U. (2023). A deep learning approach for non-invasive hypertension classification from ppg signal. In *2023 international conference on next-generation computing, iot and machine learning (ncim)* (pp. 1–5). doi:10.1109/NCIM59001.2023.10212940
5. Salim, M. S., Murad, H., Das, D., & Ahmed, F. (2023). Banglagpt: A generative pretrained transformer-based model for bangla language. In *2023 international conference on information and communication technology for sustainable development (icict4sd)* (pp. 56–59). doi:10.1109/ICICT4SD59951.2023.10303383
6. Salim, S., Islam, T., Zannat, R., Mia, N., Fuad, M., & Murad, H. (2023). Towards developing a transformer-based bangla typing error correction model: A deep learning-based approach. In *2023 international conference on information and communication technology for sustainable development (icict4sd)* (pp. 75–78). doi:10.1109/ICICT4SD59951.2023.10303361
7. Ahmed, T., Hossain, S., Salim, M. S., Anjum, A., & Azharul Hasan, K. M. (2021). Gold dataset for the evaluation of bangla stemmer. In *2021 5th international conference on electrical information and communication technology (eict)* (pp. 1–6). doi:10.1109/EICT54103.2021.9733662
8. Salim Shakib, M. S., Ahmed, T., & Azharul Hasan, K. M. (2019). Designing a bangla stemmer using rule based approach. In *2019 international conference on bangla speech and language processing (icbslp)* (pp. 1–4). doi:10.1109/ICBSLP47725.2019.201533

Under Review and Ongoing Research

Under Review

- **BConvQA: A Bangla Conversational Question-Answering Model Using Transformer-based Architecture(EMNLP 2024(Revision))**
  – Developing a Bangla Context-based Conversational Question Answering (CCQA) system faces challenges like limited data, inadequate translation methods, and a lack of pretrained models. To address these, we created a robust Bangla CCQA dataset using quality-controlled machine translation and LLM-based augmentation of English datasets, partitioned into training, validation, and test splits. We finetuned and evaluated various sequence-to-sequence models, appending conversation history to the input prompt to maintain context. The dataset and testing scripts are publicly available on GitHub, providing a foundation for further research in Bangla conversational AI.
- **LLM QA chatbot builder: A generative AI-based chatbot for Question answering(Submitted to softwareX)**
  – In this paper, we introduce the LLM QA builder, a web app simplifying the development of interactive chatbots using large language models (LLMs). The process involves data collection, LLM fine-tuning, and chat interface development. Our system supports fine-tuning of three LLMs: Zepyhr, Mistral, and Llama2, tailored for organization-specific information retrieval. It incorporates RAG techniques and retrieval ensembling for further enhancement. We showcase its capabilities through a university information chatbot. Comparative analysis using a benchmark dataset reveals Mistral 7B model with retriever ensemble achieving superior performance with a score of 4.8.
- **Agricultural Recommendation System based on Multivariate Weather Forecasting Model(Engineering Applications of Artificial Intelligence journal) (PRE-PRINT)**
  – This paper proposes a context-based crop recommendation system using a weather forecast model to improve farming practices in Bangladesh. The multivariate Stacked Bi-LSTM Network is used for accurate weather prediction, including rainfall, temperature, humidity, and sunshine. The system guides farmers in making informed decisions about planting, irrigation, harvesting, and more. It also alerts farmers about extreme weather conditions and provides knowledge-based crop recommendations for flood and drought-prone areas.
- **Comparing Prompt Based and Standard Fine Tuning for Bangla Text Classification(EMNLP 2024)**

Awards and Projects

- Dean's Award by Faculty of Electrical & Electronic Engineering

Projects

- **LLM based QA chatbot builder** A generative AI-based chatbot for question answering
- **Medical LLM Chatbot** - Chat with pdf using medical LLM langchain and streamlit
- **KUET Chat Bot** - Information about KUET
  – Students can chat with the bot and get information about KUET.
- **Efficient Backlog Routine Generator** - Python and Flask
- **Anonymity-Preserving Post Web Application** - Implementing Python and Flask for Secure and Confidential Content Sharing
- **Statistics exam**-Design and Implementation of a Python/Flask-Based Randomized Statistics Exam Generator
- **Counterfeit note detection** - Fake Bangladeshi Banknote Detection using Convolutional Neural Networks (CNN)

Miscellaneous Experiences

2023
- Ibex(Supercomputer) fine-tune Mistral for medical data, Hajj data and natural SQL question-answering system
- Research paper review (Worked as a reviewer for the conference AAAI-2024).

2023-2024
- Thesis coordinator

Technical Skills

- Programming Languages - Python, C, C++, Javascript, HTML, CSS, LaTeX , Java
- Frameworks - Pytorch, Tensorflow, Langchain, HuggingFace Transformer, Scikit-learn, Keras, Streamlit, Gradio, Flask