KHONDKER SALMAN SAYEED

salkhon050@gmail.com | salkhon.github.io
Khondker Salman Sayeed | salkhon | salkhon | SalmanKhondker

Research Interest

• Natural Language Processing • Multimodal Learning • Information Retrieval • Low Resource Deep Learning

My research focuses on developing intelligent agents capable of understanding and interacting with the world through natural language. I specialize in natural language processing (NLP) and vision-language understanding, aiming to create trustworthy, interpretable, and continuously learning AI systems that can address complex, real-world challenges.

Education

• **Bangladesh University of Engineering and Technology (BUET)** *April 2019 - June 2024*
Bachelor of Science in Computer Science and Engineering Dhaka, Bangladesh
CGPA: 3.9/4.0, placed on the dean's list in all terms.
(session delay due to the COVID-19 pandemic)
Thesis Supervisor: Prof. Rifat Shahriyar

Publications

C = Conference, W = Workshop

C1 Haz Sameen Shahgir†, **Khondker Salman Sayeed**†, Abhik Bhattacharjee, Wasi Uddin Ahmad, Yue Dong, Rifat Shahriyar
Title: "IllusionVQA: A Challenging Optical Illusion Dataset for Vision Language Models".
Summary: The paper introduces IllusionVQA, a new dataset designed to test Vision Language Models (VLMs) on optical illusions and hard-to-interpret scenes. It evaluates VLMs on two tasks: comprehension and localization. The study also highlights the limitations of In-Context Learning and Chain-of-Thought reasoning in VLMs.
[webpage] [arXiv] [openreview]
COLM 2024**: The 1st Conference on Language Modeling* | ***SoCal NLP Symposium 2024

W1 Haz Sameen Shahgir†, Rownok Zahan Ratul†, Md Toki Tahmid†, **Khondker Salman Sayeed**, Atif Hasan Rahman
Title: "RNA-DCGen: Dual Constrained RNA Sequence Generation with LLM-Attack".
Summary: The paper introduces RNA-DCGen, a novel framework for generating RNA sequences with specific structural or functional properties. RNA-DCGen uses an RNA language model and can enforce conditions on generated sequences through gradient search and fixing conserved regions.
[bioRxiv]
***NeurIPS 2024 MLSB**: Workshop on Machine Learning in Structural Biology*

C2 Tamzeed Mahfuz†, Satak Kumar Dey†, Ruwad Naswan†, Hasnaen Adil, **Khondker Salman Sayeed**, Haz Sameen Shahgir
Title: "Too Late to Train, Too Early To Use? A Study on Necessity and Viability of Low-Resource Bengali LLMs".
Summary: This paper explores the need for Large Language Models (LLMs) specifically for Bengali, a low-to-moderate resource language. The authors comparatively evaluate existing LLMs and smaller Bengali encoder-decoder models and identify challenges such as inefficient tokenization and biases in machine-translated datasets.
[arXiv]
***COLING 2025**: The 31st International Conference on Computational Linguistics*

W2 H.A.Z. Sameen Shahgir†, **Khondker Salman Sayeed**†, Tanjeem Azwad Zaman†, Md. Asif Haider†, Sheikh Saifur Rahman Jony, M. Sohel Rahman
Title: "Ophthalmic Biomarker Detection Using Ensembled Vision Transformers – Winning Solution to IEEE SPS VIP Cup 2023".
Summary: This paper presents a method for detecting ophthalmic biomarkers OCT images using two vision transformer-based models, MaxViT and EVA-02. MaxViT is better at detecting local features, while EVA-02 excels at identifying global features. Their inference-time ensemble approach combines the best of both worlds.
[icip] [arXiv]
***ICIP 2023**: IEEE SPS Video and Image Processing Cup*

C3 **Khondker Salman Sayeed**, Haz Sameen Shahgir, Tamzeed Mahfuz, Satak Dey, M Saifur Rahman
Title: "Efficient Real-Time Video Colorization on Low-End CPUs via Pruning and Quantization".
Summary: This paper proposes a method to achieve real-time video colorization on low-end CPUs using pruning and quantization techniques. The authors use a U-Net architecture with an EfficientNet-B7 encoder and a lightweight decoder, then use a magnitude-based pruning method and activation-aware quantization.
[nsyss] [acm]
***NSysS 2024**: The 11th International Conference on Networking, Systems and Security*

Work Experience

- **IQVIA [🔗]** *June 2024 - Present*
  Machine Learning Engineer Remote
  - Working on the development and deployment of IQVIA's in-house Large Language Models for generating expert domain-specific insights. Developing a Mixture-of-Agents based AI Assistant pipeline using LangGraph, CrewAI. Serving business users with data-informed intelligent and interactive responses.
- **BUET CSE NLP Lab [🔗]** *March 2023 - June 2024*
  Research Assistant Dhaka, Bangladesh
  - Prepared a benchmark illusion dataset that tests the visual understanding capabilities of Vision Language Models to identify visual illusions in images and correctly describe them. Published this first-author work at COLM 2024.
  - Worked on the multimodal extension of the XLSum dataset from BUET CSE NLP group to include images and BBC article-summary pairs. Explored the potential improvement for abstractive summarization by including article images in addition to text using Vision Language Models.

  Supervisors: Prof. Rifat Shahriyar, Prof. Tahmid Hasan, CSE BUET & Wasi Ahmad, Senior Research Scientist, NVIDIA.

Honors and Awards

- **IEEE Upsilon Pi Epsilon Honor Society Award 2023 (Awardee)** *February 2024*
  IEEE Computer Society [🔗]
  - One of four international winners of this prestigious award in recognition of my academic results and success in national and international machine learning competitions. Received a $1000 scholarship.
- **IEEE Video and Image Processing Cup 2023 (Champion)** *October 2023*
  IEEE Signal Processing Society, ICIP 2023 [🔗] [arXiv]
  - Champion in this annual international computer vision competition. 2023's edition was based on biomarker prediction from retina OCT scans. Received travel grant to present our solution at ICIP 2023, Kuala Lumpur, and a $5000 scholarship prize.
  - **Summary:** In this work, we employed two vision transformer-based models: MaxViT and EVA-02. Our ensembled solution achieved a patient-wise F1 score of 0.81 in the first phase and 0.85 in the second and final phase, setting the state-of-the-art ophthalmic biomarker detection using OCT images.
- **DLSprint 2022 (Champion)** *September 2022*
  Optimizely, Bengali.AI, and CSE BUET [🔗] [arXiv] [GitHub]
  - Champion in this Kaggle community competition based on Automatic Speech Recognition on the Bengali Common Voices Dataset. Received a $1300 scholarship.
  - **Summary:** In this work, we fine-tuned wav2vec 2.0 for Bengali speech recognition using the Bengali Common Voice Speech Dataset. Our model achieved impressive results, with a word error rate (WER) of 0.25 on the validation set and a Levenshtein Distance of 2.61 after further training. This work bridges the gap for a widely spoken language.
- **EEE Day Datathon 2023 (First Runner-up)** *March 2023*
  EEE BUET, Apurba Technologies Ltd. [🔗] [arXiv] [GitHub]
  - Runner-up in this Kaggle community competition based on Automatic Grammatical Error Tagging on Bengali Text. Received a $500 scholarship.
  - **Summary:** In this work, we propose a method for detecting grammatical errors in Bangla using a Text-to-Text Transfer Transformer (T5) Language Model. We fine-tune the small variant of BanglaT5 on a corpus where errors are surrounded by the dedicated tokens "$". Despite the T5 model's primary design for translation, we achieve low Levenshtein Distance in tagging grammatical errors in Bangla.
- **AI For Bangla 2023 (Honorable Mention)** *March 2023*
  EBLICT, Bangladesh Computer Council (BCC) [🔗] [arXiv] [GitHub]
  - Prize winner in this national competition calling for advancements in AI for Bengali for creating the first Bengali Sign Language video dataset and preparing a baseline on that dataset. Received a $500 scholarship.
  - **Summary:** In this research, we introduce a new word-level Bangla Sign Language dataset, BdSL40, comprising 611 videos across 40 words. We prepare two baselines for classification: one using a 3D Convolutional Neural Network (CNN) model and another employing a novel Graph Neural Network (GNN) approach.
- **Robi Datathon 2024 (First Runner-up)** *May 2024*
  Robi Axiata Ltd, Huawei [🔗]
  - Runner-up in this national data-science competition based on customer purchase behavior prediction in a data-scarce setting. Received a $3000 scholarship.

Leadership Experience

- **Lead Organizer & Instructor of DLSprint 2.0 2023** *August 2023*
  BJIT Group, CSE BUET [🔗]
  - Served as an organizer of this open-for-all Computer Vision Competition on Automatic Document Layout Analysis for Bengali.
  - Conducted two workshops as an instructor on common competition practices, computer vision, and image segmentation.
  - Took ownership, being involved in the competition's conception to execution.

Projects

- **Ticketing** – A microservices web app that lets users buy and sell tickets
  - Backend: NodeJS, Typescript, MongoDB, NATS Streaming. Consists of 6 independent services, coordinated by an eventbased architecture.
  - Frontend: NextJS, TailwindCSS
  - Orchestration: Kubernetes
- **Sub-C-Compiler** – A compiler for a subset of the C programming language
  - Tools: Flex (Lexer), GNU Bison (Paser)
  - Languages: C, C++
- **Mobile-Doc** – A telemedicine web app
  - Backend: FastAPI, MongoDB, Redis, BigQuery
  - Frontend: React, MUI
  - Deployment: Netlify
  - Deployment: Pokedoc
- **Suntech** – An e-commerce web app
  - Backend: Flask, OracleDB
  - Frontend: Jinja2, HTML, CSS, VanillaJS
- **Ray-Tracing** – A rendering pipeline implementing ray tracing in OpenGL scenes
  - Language: C++
  - Algorithms: Phong Lighting Model, Recursive Reflection
- **Enhanced xv6 OS** – Copy-On-Write and Memory Page-Replacement in the xv6 OS
  - Language: C
  - System Calls, Process Schedulers, Page Replacement
- **twitt3r** – A Twitter clone using the t3 stack
  - Framework: NextJS
  - Stack: T3 (Typescript, trpc, TailwindCSS)
  - Persistence: PlanetScale, Prisma ORM
  - Deployment: Vercel
  - Deployment: twitt3r
- **DocumentQA** – A langchain LLM document chatbot
  - Language: Python
  - Libraries: langchain, OpenAI
  - UI: Chainlit
  - Database: ChromaDB Vector Database
- **IllusionVQA** – A Challenging Optical Illusion Dataset for Vision Language Models
  - This work is a part of my undergraduate thesis
  - Published at **COLM'24**
  - Paper: IllusionVQA: A Challenging Optical Illusion Dataset for Vision Language Models
  - Website: IllusionVQA

References

1. **Rifat Shahriyar (PhD)**
   Professor, Department of Computer Science and Engineering
   Bangladesh University of Engineering and Technology (BUET)
   Email: rifat@cse.buet.ac.bd
   Relationship: Undergraduate Thesis Supervisor
2. **AKM Ashikur Rahman (PhD)**
   Professor, Department of Computer Science and Engineering
   Bangladesh University of Engineering and Technology (BUET)
   Email: ashikur@cse.buet.ac.bd
   Relationship: Academic Advisor