

# Md Zabirul Islam

📍 Troy, New York | 📩 [zabirul.islam1996@gmail.com](mailto:zabirul.islam1996@gmail.com) | 📞 (518) 805-7992 | 💬 [/zabirul-islam/](https://www.linkedin.com/in/zabirul-islam/)



## PROFESSIONAL SUMMARY

Ph.D. researcher in Computer Science specializing in **large language models (LLMs)**, **multimodal foundation models**, and **generative AI**, with **15+ peer-reviewed publications** and **1,150+ citations**. Research centers on large-scale neural architectures for medical decision support, multimodal vision–language modeling with interactive explanations, and AI-enabled clinician and STEM education.

Experienced in designing **reliable, interpretable, and deployable AI pipelines** that integrate medical images, clinical text, and structured representations. Proficient in **PyTorch**, **Hugging Face Transformers**, and **vLLM**, with hands-on expertise in model fine-tuning, evaluation, and efficient inference. Current work emphasizes **scalable and auditable AI systems** for healthcare and advanced education, supporting responsible clinical decision-making and workforce training. Eligible for F-1 OPT/CPT internships.

## EDUCATION

### Rensselaer Polytechnic Institute (RPI)

Troy, NY

*Doctor of Philosophy in Computer Science*

Jan. 2024 – Present

- Ph.D. Advisor: Prof. Ge Wang

- Focus: AI in Education; Large Language Models (LLMs); Multimodal Learning and Vision–Language Models; Medical Imaging and Clinical AI.

### Rensselaer Polytechnic Institute (RPI)

Troy, NY

*Master of Science in Computer Science (CGPA: 3.93/4.00)*

Dec. 2025

### Khulna University of Engineering and Technology (KUET)

Khulna, Bangladesh

*Bachelor of Science in Computer Science and Engineering*

Mar. 2020

## EXPERIENCE

### Graduate Research Assistant

Jun. 2025 – Present

*Rensselaer Polytechnic Institute*

Troy, NY

- Conducting Ph.D. research under Prof. Ge Wang on **large artificial neural networks for medical imaging research and education**, with emphasis on multimodal foundation models, semantic reliability, and interactive AI-driven learning systems.
- Fine-tuned large language models (**LLaMA-3.1 8B**, **Qwen-3 8B**, **Mistral 7B**) for **structured radiology report generation**, achieving F1 scores of 98.05%, 98.10%, and 97.25% respectively, while reducing end-to-end inference time from 9 hours to 2.5 hours on 5,000 reports via dynamic template-constrained decoding.
- Developed the *ALIVE* interactive avatar-based lecture Q&A system to support **scalable, AI-assisted medical imaging education**, enabling **pause-triggered, lecture-grounded question answering** through text and voice with timestamp-aligned retrieval, LLM reasoning, and synthesized instructor-style explanations.
- Designed a modular end-to-end multimodal pipeline (ASR → retrieval → LLM inference → TTS → avatar synthesis) with support for three deployment modes: (i) fully local on-device inference, (ii) browser-based systems with API-hosted models, and (iii) self-hosted **vLLM** servers for scalable multi-user access.
- Integrated **Whisper** for speech recognition, offline TTS pipelines, and **SadTalker**-based neural talking-head synthesis to deliver low-latency, lip-synced avatar responses in educational settings.

### Graduate Teaching Assistant

Jan. 2024 – May 2025

*Rensselaer Polytechnic Institute*

Troy, NY

- Assisted in *CSCI 6360 – Parallel Computing and Parallel Programming* and *CSCI 2300 – Introduction to Algorithms* under Prof. Christopher D. Carothers, supporting instruction across undergraduate and graduate levels.
- Guided students in **CUDA** and **MPI** programming on the **AiMOS supercomputer**, emphasizing parallelization and performance optimization.
- Supported laboratory sessions on parallel algorithm design and performance profiling using **TAU** and MPI profilers in high-performance computing environments.

- Held weekly office hours and problem-solving sessions; supported grading, student mentoring, and assessment design for algorithmic coursework.
- Reinforced core algorithmic topics including asymptotic analysis, dynamic programming, and graph algorithms (DFS/BFS, MST, shortest paths).
- Prepared solution walkthroughs and rubric-aligned feedback using **Submitty** and **Gradescope**, clarifying proofs of correctness and complexity analysis.

## PROJECTS

---

<b>Development of Multimodal Foundation Model for Lung Cancer Diagnostics</b>	Sep. 2024 – Present
<ul style="list-style-type: none"> <li>– Developing a <b>domain-specific LLM</b> integrating <b>radiology reports</b> and <b>medical imaging</b> for structured reasoning and diagnostic prediction.</li> <li>– Implementing <b>cross-attention</b> for multimodal fusion between <b>CT image encoders</b> and GPT-style decoders.</li> <li>– Designing evaluation pipelines using <b>IuC</b> and <b>F1 metrics</b>, with instruction tuning and ablation studies on clinical datasets.</li> <li>– Employing <b>schema-guided constrained decoding</b> to ensure structured outputs and reduce hallucination, with <b>RAG</b> for retrieval consistency.</li> <li>– Leveraging <b>TorchTune</b> and <b>EleutherAI LM-Eval</b> for scalable benchmarking and reproducible experimentation.</li> </ul>	
<b>ALIVE: Interactive Video Lecture Q&amp;A System with Human Avatar Integration</b>	June 2024 – Present
<ul style="list-style-type: none"> <li>– Designed and implemented a fully local, content-aware interactive lecture system enabling students to pause videos and receive real-time, lecture-grounded explanations via text or avatar-delivered responses.</li> <li>– Developed a <b>timestamp-aligned retrieval pipeline</b> combining semantic similarity and temporal proximity using sentence embeddings and FAISS to ground LLM responses in relevant lecture segments.</li> <li>– Integrated a <b>human-avatar response pipeline</b> using offline TTS and SadTalker-based neural talking-head synthesis for lip-synced, instructor-style video explanations.</li> <li>– Implemented <b>pause-triggered multimodal interaction</b> (chat and voice) with local ASR (Whisper), LLM inference, REST-based backend orchestration, and browser-based frontend playback.</li> <li>– Introduced segmented avatar synthesis with progressive preloading to reduce perceived latency while maintaining visual continuity during long explanations.</li> <li>– Evaluated the system on a complete <b>medical imaging course</b> (24+ lectures), analyzing retrieval correctness, grounding behavior, and end-to-end latency under fully local deployment.</li> </ul>	
<b>SlideChain: Blockchain-Backed Semantic Provenance for Lecture Understanding</b>	June 25 - Present
<ul style="list-style-type: none"> <li>– Proposed a blockchain-backed provenance framework to ensure <b>verifiable integrity, auditability, and reproducibility</b> of multimodal semantic extraction in educational AI systems.</li> <li>– Developed a unified multimodel semantic extraction pipeline that generates structured <b>concepts and relational triples</b> from lecture slides using four state-of-the-art vision-language models.</li> <li>– Designed a <b>slide-level provenance schema</b> and normalization pipeline enabling cross-model semantic comparison, disagreement analysis, and deterministic reproducibility across runs.</li> <li>– Implemented an <b>EVM-compatible smart contract</b> that anchors cryptographic (Keccak-256) commitments of slide-level semantic records on-chain, providing tamper-evident and time-stamped provenance.</li> <li>– Conducted the first large-scale empirical study of <b>semantic instability in VLMs</b> on real STEM lecture material, analyzing cross-model disagreement, Jaccard similarity, and lecture-level variability.</li> <li>– Evaluated scalability and performance through full-course registration (1,117 slides), demonstrating constant gas usage, linear cost scaling, ~1 slide/sec throughput, and 100% tamper-detection accuracy.</li> </ul>	
<b>TrialDura: Hierarchical Attention Transformer for Clinical Trial Duration Prediction</b>	Oct. 2024
<ul style="list-style-type: none"> <li>– Proposed a <b>hierarchical attention transformer</b> model for predicting clinical trial durations using multimodal data inputs.</li> <li>– Integrated <b>BioBERT embeddings</b> for diseases, drugs, and trial phases to improve semantic understanding.</li> <li>– Achieved <b>MAE 1.04 years</b> and <b>RMSE 1.39 years</b>, improving over prior baselines, with interpretable attention heatmaps for explainability.</li> </ul>	
<b>AI-Driven COVID-19 Diagnostic Systems</b>	Jun. 2020 – Dec. 2022

- Developed **CNN-LSTM hybrid models** for COVID-19 detection from **chest X-rays**, achieving **99.4% accuracy** and **99.9% AUC** on benchmark datasets with Grad-CAM-based interpretability.
- Built a **CNN-RNN architecture** leveraging **transfer learning** to classify COVID-19 cases from chest X-rays, validated on a dataset of 6,939 samples using Grad-CAM visualizations.
- Conducted **cross-country statistical analyses** of pandemic-related factors using epidemiological data from **100 countries** to study lockdown effectiveness.

## TECHNICAL SKILLS

---

**Machine Learning & AI:** Machine Learning (ML), Artificial Intelligence (AI), LLMs (LLaMA, GPT, Mistral), Transformer architectures, instruction-tuned models, RAG, diffusion models, multimodal fusion, explainable AI (SHAP, Grad-CAM), RLHF.

**Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, vLLM, TorchTune, LangChain, OpenAI Triton, NVIDIA Riva, Triton Inference Server.

**NLP:** Schema-guided generation, information extraction, NER, template-constrained decoding, prompt engineering, evaluation (BLEU, ROUGE, F1, IuC).

**Computer Vision & Imaging:** CNNs, Vision Transformers (ViT), CT/MRI preprocessing, DICOM, image registration, segmentation, 3D reconstruction, cross-modal fusion.

**MLOps & Deployment:** Docker, Kubernetes, FastAPI, RESTful APIs, model serving, CI/CD pipelines, MLflow, inference optimization, API integration, scalability.

**GPU/HPC:** CUDA, GPU-accelerated training, Tensor Cores, MPI, parallelization, performance optimization, HPC environments (AiMOS, DGX Spark).

**Programming:** Python, C, C++, Java, Bash, SQL, R, MATLAB.

**Cloud Platforms & Tools:** AWS (EC2, S3), Google Cloud (GCP), Azure ML Studio, NVIDIA Omniverse ACE, Visual Studio, PyCharm, GitHub Actions, Linux/Ubuntu.

**Soft Skills:** Cross-functional collaboration, technical communication, project leadership, data pipeline automation, problem solving.

## PUBLICATIONS

---

**Md. Zabirul Islam**, Md. Milon Islam, Amanullah Asraf, “A Combined Deep CNN-LSTM Network for COVID-19 Detection Using X-ray Images,” *Informatics in Medicine Unlocked*, 2020.

Md. Milon Islam, **Md. Zabirul Islam**, Amanullah Asraf, Mabrook S. Al-Rakhami, Ali Hassan Sodhro, Weiping Ding, “Diagnosis of COVID-19 from X-rays Using Combined CNN-RNN Architecture with Transfer Learning,” *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, issue 4, p. 100088, 2022.

Amanullah Asraf, **Md. Zabirul Islam**, Md. Rezwanul Haque, Md. Milon Islam, “Deep learning applications to combat novel coronavirus (COVID-19) pandemic,” *SN Computer Science*, vol. 1, issue 6, p. 363, 2020.

Ling Yue, Sixue Xing, Jonathan Li, **Md. Zabirul Islam**, Bolun Xia, Jintai Chen, Tianfan Fu, “TrialDura: Hierarchical Attention Transformer for Interpretable Clinical Trial Duration Prediction,” *Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '24)*, Article 85, 2024.

Chuang Niu<sup>+</sup>, Md Sayed Tanveer<sup>+</sup>, **Md. Zabirul Islam**<sup>+</sup>, *et al.*, “Development and Validation of a Large Language Model for Generating Fully-Structured Radiology Reports,” *arXiv preprint arXiv:2404.13235*, 2024.

**Md. Zabirul Islam**, Md Motaleb Hossen Manik, Ge Wang, “ALIVE: An Avatar-Lecture Interactive Video Engine with Content-Aware Retrieval for Real-Time Interaction,” *arXiv preprint arXiv:2512.20858*, 2025.

Md Motaleb Hossen Manik, **Md. Zabirul Islam**, Ge Wang, “SlideChain: Semantic Provenance for Lecture Understanding via Blockchain Registration,” *arXiv preprint arXiv:2512.21684*, 2025.

**Md. Zabirul Islam**, Ge Wang, “Avatars in the Educational Metaverse,” *Visual Computing for Industry, Biomedicine, and Art*, vol. 8, article 15, 2025.

**Full list:** <https://scholar.google.com/citations?user=wFrXDwoAAAAJ&hl=en>