Md Kamrul Islam

Dhaka, Bangladesh | +88 018 64 208599 | mdkamrul.islam@student-cs.fr

Portfolio | LinkedIn | GitHub | ORCID

Career Objective

AI Engineer with two years of research and industry experience in large language models, information extraction, deep learning, and medical imaging. Skilled in developing and deploying LLM-based agents, information extraction using LLMs, and Retrieval-Augmented Generation (RAG), LangChain, and PyTorch. Seeking a research assistantship in Natural Language Processing, Large Language Models, Information Extraction, and Deep Learning for real-world applications.

Education

Erasmus Mundus Joint Master's Degree in Big Data Management and Analytics (BDMA) *Sept 2023 – Oct 2025*
CentraleSupélec | Universitat Politècnica de Catalunya | Université Libre de Bruxelles *France | Spain | Belgium*

- **Weighted Average:** 14.5/20 (Grade: A)
- **Award:** Recipient of the Erasmus Mundus Scholarship, European Commission (fully funded).

Mobility Path:

- **Université Libre de Bruxelles (ULB)**, Brussels, Belgium
- **Universitat Politècnica de Catalunya (UPC)**, Barcelona, Spain
- **CentraleSupélec**, Paris, France

Bachelor of Engineering in Software Engineering *Mar 2018 – Dec 2021*
Sichuan University *Chengdu, China*

- **CGPA:** 3.64/4.00 (Weighted Average: 87%)
- **Award:** Recipient of the Belt and Road Initiative Scholarship, Chinese Government, (fully funded).

Work Experience

AI Research Engineer Intern (Project Link) *May 2025 – Oct 2025*
LISSI, Université Paris-Est Créteil (in collaboration with CEA–LIST, France) *Paris, France*

- Proposed a hybrid framework integrating LLM-based semantic reasoning with rule-based validation for automated cybersecurity annotation of BPMN models.
- Improved annotation accuracy and efficiency through advanced prompt engineering and retrieval-augmented generation (RAG) across multiple LLMs.
- Built a benchmark dataset and achieved higher annotation accuracy while reducing annotation time by 95% compared to domain experts.

Tech Stack: *Python, LangChain, LLMs, RAG, Prompt Engineering, React.js, Flask, CI/CD*

Graduate Research Assistant (Project Link) *Oct 2024 – Feb 2025*
Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), CentraleSupélec *Paris, France*

- Developed an unsupervised deep-learning framework combining CNN and Group-Equivariant CNNs (G-CNNs) to analyze large-scale medical imaging datasets for early disease detection.
- Improved clustering accuracy and feature robustness by encoding geometric symmetries, mitigating reliance on data augmentation.
- Implemented multi-GPU distributed training pipelines for scalable, high-performance experiments.

Tech Stack: *PyTorch, OpenCV, HPC, CUDA, LaTeX*

Data Engineer Intern *Dec 2020 – May 2021*
Chengdu Suncape Co., Ltd *Chengdu, China*

- Developed an end-to-end data pipeline for large-scale data ingestion, transformation, and analytics in distributed environments.
- Enhanced data preprocessing and feature-engineering, boosting predictive model accuracy by 20%.
- Collaborated within an Agile Scrum environment to implement best practices in modular design and CI/CD.

Tech Stack: *Apache Spark, Scikit-learn, SQL, Agile Scrum, Jira, Git, CI/CD*

Publications

1. **M. K. Islam**, T. Henry, M. Salnitri, J. Köpke, and S. Souihi, "*A Hybrid LLM-Based Framework for Automated Security Annotations Generation in Business Process Models*," Submitted to the *24th International Conference on Business Process Management (BPM)*, 2026 (under review).
2. C. K. Sah, L. Xiaoli, M. M. Islam and **M. K. Islam**, "*Navigating the AI Frontier: A Critical Literature Review on Integrating Artificial Intelligence into Software Engineering Education*," *2024 36th International Conference on Software Engineering Education and Training (CSEE&T)*, Würzburg, Germany, 2024, pp. 1–5. doi:10.1109/CSEET62301.2024.10663054

Projects

Purchase Intelligence Brief for Amazon (Chrome Extension) (Project Link) *Mar 2026*
Developed an LLM-based agent that analyzes product reviews to generate interpretable recommendations; built for the Major League Hacking (MLH) challenge.
Tech Stack: *JavaScript, Node.js, LLMs APIs, AI Agents, Express.js*

DigiScan360: End-to-End Data Analysis for Competitive Intelligence (Project Link) *Feb 2024 – June 2024*
Developed a competitive intelligence platform and pitched it as a startup prototype at UPC's entrepreneurship initiative.
Tech Stack: *PySpark, LLMs, Microsoft Fabric, Microsoft Azure, Power BI, GraphDB, SPARQL*

Anomaly Detection in Diesel Train Cooling Systems (Project Link) *Sept 2023 – Dec 2023*
Developed unsupervised models to detect anomalies in train cooling systems for the Belgian National Railway Company (SNCB).
Tech Stack: *Python, SQL, Scikit-learn, Tableau*

Brain Tumor Detection and Classification by Using CNN (Project Link) *Sept 2021 – Dec 2021*
Achieved 98% accuracy in brain tumor detection using a custom CNN model.
Tech Stack: *TensorFlow, Keras, Python, OpenCV*

AI-Based Disease Prediction System (Project Link) *Mar 2021 – June 2021*
Built a web-based symptom checker predicting over 40 diseases with machine learning algorithms.
Tech Stack: *Python, Scikit-learn, Django, ReactJS, REST APIs*

Technical Skills

Programming & Framework: Python, C/C++ (OOP), SQL, JavaScript, React.js, Flask
Data Science & ML: PyTorch, LLM APIs, LangChain, LangGraph, MCP
Big Data & Cloud: Apache Spark, Microsoft Azure, Microsoft Fabric, Apache Airflow
Graph & Database Systems: PostgreSQL, Neo4j, GraphDB, ETL Pipelines
Data Visualization: Power BI, Tableau, Streamlit, D3.js
Tools & Collaboration: Linux, Git, Docker, Overleaf Latex

Extracurricular Activities

Twelfth European Big Data Management & Analytics Summer School (eBISS 2024) *Padova, Italy*
University of Padova (Poster Link) (Paper Link) *July 2024*

- Participated in the fully funded summer school sponsored by the European Commission.
- Presented a research poster titled *"Applying Knowledge Graphs in Retrieval-Augmented Generation (RAG)"*.

Languages

Bengali: Native | **English:** C2 (Proficient) | **Chinese:** B2 (Upper-intermediate) | **French:** A2 (Elementary)

References

Tiphaine HENRY, PhD
Researcher, CEA LIST
Paris, France
Email: tiphaine.henry@cea.fr
Master's Thesis Supervisor

Mahmoud SAKR
Professor, Université Libre de Bruxelles
Brussels, Belgium
Email: mahmoud.sakr@ulb.be
Graduate Course Teacher

Esteban ZIMÁNYI
Professor, Université Libre de Bruxelles
Brussels, Belgium
Email: esteban.zimanyi@ulb.be
Master's Program Coordinator