Md Kamrul Islam

Paris, France — +33 6 52 24 36 55 — mdkamrul.islam@student-cs.fr
kamrul-portfolio-sigma.vercel.app/ — linkedin.com/in/kamrulkonok — github.com/kamrulkonok

Education

Erasmus Mundus Master Degree in Big Data Management and Analytics (BDMA)

M2 – CentraleSupélec, Université Paris-Saclay — Gif-sur-Yvette, France
Master Sciences de l'Ingénieur — *Sept 2024 – Present*

- Master Thesis: *"Data Sharing Policy Recommendation using LLMs"(ongoing)*

M1 – Universitat Politècnica de Catalunya — Barcelona, Spain
Master Erasmus Mundus in Big Data Management and Analytics — *Feb 2024 – July 2024*

M1 – Université libre de Bruxelles — Brussels, Belgium
Master of Science in Computer Science and Engineering — *Sept 2023 – Jan 2024*

Bachelor of Engineering in Software Engineering — Chengdu, China
Sichuan University — *Mar 2018 – Jan 2022*

- Average Score: 87/100
- Bachelor Thesis: "Brain Tumor Detection and Classification using CNN"

Work Experience

AI Research Intern — Paris, France
Laboratoire Images, Signaux et Systèmes Intelligents, Université Paris-Est Créteil — *May 2025 – Present*

- Building an end-to-end pipeline for extracting security and data-sharing requirements from multimodal inputs to automatically generate BPMN workflows enriched with semantically valid cybersecurity annotations.
- Benchmarked different prompt strategies and RAG system, improving valid security labels extraction rate with GDPR/HIPAA recomendations.
- Developed an evaluation suite including F1 classification against ground-truth security labels, and LLM-as-a-judge scoring for final output quality.

***Technologies**: Python, LangChain, LLMs, RAG, React-JS, Flask, SecBPMN*

Graduate Research Assistant **(Project Link)** — Gif-sur-Yvette, France
Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), CentraleSupélec — *Oct 2024 – Feb 2025*

- Developed an unsupervised framework with AlexNet and Group Equivariant CNNs on large-scale medical dataset.
- Improved clustering accuracy through geometric symmetry encoding, reducing reliance on data augmentation.
- Enabled early disease diagnosis by supporting multi-class classification and segmentation without manual labels.
- Engineered a scalable multi-GPU pipeline for high-performance experiments.

***Technologies**: PyTorch, OpenCV, LaTeX, HPC*

Professional Experience

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

- Developed and optimized Apache Spark data pipelines for large-scale processing and analytics.
- Improved preprocessing workflows, achieving a ∼20% increase in predictive model performance.
- Collaborated in Agile Scrum to ensure best practices in code quality and version control.

***Technologies**: Scikit-learn, Apache Spark, Agile Scrum, Jira, Git*

Technical Skills

Programming Languages & Frameworks: Python, C/Java(OOP), SQL, Cypher, SPARQL
Machine Learning & Deep Learning: LLMs, Pytorch, TensorFlow, Langchain, Scikit-Learn, OpenCV
Big Data & Cloud Platforms: PySpark, Apache Airflow, Microsoft Fabric, HPC
Databases: SQL Server, PostgreSQL, Neo4j, GraphDB
Project Management & Collaboration: Docker, Agile, Scrum, Git

Projects

DigiScan360 *Feb 2024 – June 2024*
Developed a competitive intelligence platform and pitched it as a startup prototype at UPC's entrepreneurship initiative. Project Link
Technical Skills: *PySpark, LLMs, Microsoft Fabric, Azure Data Factory, Power BI, GraphDB, SPARQL*

Anomaly Detection in Diesel Train Cooling Systems *Sept 2023 – Dec 2023*
Developed unsupervised models to detect anomalies in train cooling systems for the Belgian National Railway Company (SNCB). Project Link
Technical Skills: *Pandas, Numpy, Matplotlib, Seaborn, scikit-learn, Tableau, Anomaly Detection*

PostgreSQL Extension for Chess Game Analysis *Sept 2023 – Dec 2023*
Created a custom PostgreSQL extension for storing and analyzing chess games. Project Link
Technical Skills: *PostgreSQL, C, SQL, Linux, Database Systems, Indexing*

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

AI-Based Disease Prediction System *Mar 2021 – June 2021*
Built a web-based symptom checker predicting over 40 diseases with machine learning algorithms. Project Link
Technical Skills: *Python, Django, ReactJS, Machine Learning, REST APIs*

Publications

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

Extra-Curricular Activities

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

- Participated in the fully-funded summer school sponsored by the European Commission, focused on advanced topics in Big Data, AI, and Business Intelligence.
- Presented a research poster titled *"Applying Knowledge Graphs in Retrieval-Augmented Generation (RAG)"*.

Awards & Scholarships

European Union Sept 2023 – Sept 2024
Erasmus Mundus Partner Country Scholarship
Chinese Government Mar 2018 – Mar 2022
The Belt and Road Initiative Scholarship

Languages

English: C1 (Advanced proficiency), **Chinese:** B2 (Intermediate proficiency), **French:** A2 (Elementary Proficiency)

References

Dr. Tiphaine Henry
Research Engineer, CEA LIST *Email: tiphaine.henry@cea.fr*
Relationship: Master Thesis Supervisor

Prof. Nacera Seghouani
Professor, CentraleSupélec, Université Paris-Saclay *Email: Nacera.Seghouani@centralesupelec.fr*
Relationship: Big Data Research Project Supervisor
and Instructor for "Massive Graph Management and Analytics" course