Md Kamrul Islam

mdkamrul.islam@student-cs.fr — +33 6 52 24 36 55 — kamrulkonok.github.io/portfolio —
linkedin.com/in/kamrulkonok — github.com/kamrulkonok

Research Interests

• Machine Learning • Deep Learning • Large Language Model (LLMs) • MultiModal Learning • Reinforcement Learning • Self-supervised Learning • Multilingual NLP • Retrieval-Augmented Generation (RAG).

Education

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

Semesters 3 & 4 – CentraleSupélec, Université Paris-Saclay — Gif-sur-Yvette, France
Master Sciences de l'Ingénieur — *Sept 2024 – Present*

Semester 2 – Universitat Politècnica de Catalunya — Barcelona, Spain
Master Erasmus Mundus in Big Data Management and Analytics — *Feb 2024 – July 2024*

Semester 1 – Université libre de Bruxelles — Brussels, Belgium
Master of Science in Computer Science and Engineering — *Sept 2023 – Jan 2024*

Sichuan University — Chengdu, China
Bachelor of Engineering in Software Engineering (Average grade: 87/100) — *Mar 2018 – Dec 2021*

Work Experience

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

- Building an LLM-assisted solution that extracts security and data-sharing requirements from multimodal inputs to automatically generate BPMN workflows enriched with semantically valid SecBPMN annotations and context-specific data sharing protection mechanisms.

***Technologies**: Python, React-JS, LangChain, LLM, Retrieval-Augmented Generation, SecBPMN, Knowledge Graphs*

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

- Project Title: "Enhancing Self-Supervised Learning for Image Clustering Using Geometric Deep Learning"
- Developed a novel deep clustering architecture that integrates Group Equivariant CNNs to enhance self-supervised learning for medical image analysis using the NIH chest X-ray datasets.
- Demonstrated that encoding geometric symmetries directly in the network architecture eliminates the need for explicit data augmentation while improving clustering performance and generalization.
- Built and optimized a scalable training pipeline using PyTorch's DistributedDataParallel and automatic mixed precision technologies, enabling efficient model training across multi-GPU systems.

***Technical Skills:** Pytorch, OpenCV, Latex, HPC*

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

- Developed and optimized data pipelines using *Apache Spark* to process and analyze large data sets.
- Improved data preprocessing workflows to support algorithm accuracy, achieving a *10% increase* in predictive model performance.
- Collaborated with cross-functional teams using *Agile Scrum* methodology to ensure best practices in code quality and version control.

***Technical Skills:** SciKit-Learn, Apache Spark, Agile Scrum, Jira, Git*

Technical Skills

Programming Languages & Frameworks: Python, C/C++(OOP), SQL, Cypher, SPARQL, JavaScript, React-JS
Machine Learning & Deep Learning: LLMs, Pytorch, TensorFlow, Langchain, Scikit-Learn, OpenCV
Big Data & Cloud Platforms: PySpark, Hadoop, Apache Airflow, Microsoft Azure, Microsoft Fabric, HPC
Databases: SQL Server, PostgreSQL, Neo4j, GraphDB
Data Analytics & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Power BI, Tableau, Streamlit, D3.js
Project Management & Collaboration: Agile, Scrum, Git, Jira

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 SQL Server, 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*

Extra-Curricular Activities

Twelfth European Big Data Management & Analytics Summer School (eBISS 2024) Padova, Italy

Poster Link | *Paper Link* | *University of Padova* *July 2024*

- Selected to participate in a 5-day 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) Systems"*.

Awards & Scholarships

European Union Sept 2023 – Sept 2024

Erasmus Mundus Partner Country Scholarship

- Awarded a fully-funded scholarship (one of the 24 selected awardees from more than 800 applicants), covering tuition, living expenses, and academic support for the Erasmus Mundus Joint Master Degree program.

Chinese Government Mar 2018 – Mar 2022

The Belt and Road Initiative Scholarship

- Received a comprehensive 4-year scholarship covering tuition, accommodation, and academic resources during undergraduate studies.

Languages

English: C1 (Advanced proficiency), **Chinese (Mandarin):** B2 (Intermediate proficiency), **French:** A1 (Elementary proficiency)