Md Kamrul Islam

Paris, France — +33 6 52 24 36 55 — mdkamrul.islam@student-cs.fr
https://kamrulkonok.github.io — linkedin.com/in/kamrulkonok — github.com/kamrulkonok

Research Interest

Foundation Models, Retrieval-Augmented Generation, Information Extraction, Deep Learning, Computer Vision, Medical Imaging, Knowledge Graphs, and Big Data Analytics.

Work Experience

Research Engineer Intern — *Créteil, France*
LISSI, Université Paris-Est Créteil (in collaboration with CEA–LIST, France) — *May 2025 – Oct 2026*

- Proposed a novel hybrid framework integrating LLM-based semantic reasoning and rule-based validation for automated cybersecurity-annotated BPMN model generation.
- Optimized annotation accuracy and efficiency through advanced prompt engineering and retrieval-augmented generation (RAG) across multiple LLMs.
- Constructed a benchmark dataset and evaluation framework supporting quantitative and qualitative metrics.
- Achieved higher annotation accuracy (F1 = 0.516) 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 — *Gif-sur-Yvette, France*
Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), CentraleSupélec — *Oct 2024 – Feb 2025*

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

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

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

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

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

Education

Erasmus Mundus Master's Degree in Big Data Management and Analytics (BDMA) — *Paris, France*
CentraleSupélec, Université Paris-Saclay — *Sept 2023 – Oct 2025*

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

Technical Skills

Core Programming: Python, C/C++ (OOP), SQL, Cypher, SPARQL, JavaScript, TypeScript, React.js, Flask

Data Science & ML: PyTorch, Scikit-learn, Hugging Face Transformers, LangChain, LLMs, OpenCV, CUDA

Big Data & Cloud: PySpark, Hadoop, Apache Airflow, Azure, Vertex AI, Fabric, HPC Environments

Graph & Database Systems: PostgreSQL, Neo4j, GraphDB, ETL Pipelines

Analytics & Visualization: Pandas, NumPy, Power BI, Tableau, Streamlit, D3.js

Tools & Collaboration: Git, Docker, REST APIs, CI/CD, Agile Scrum, Jira

Projects

DigiScan360 (Project Link) — *Feb 2024 – Jun 2024*

- Developed a competitive intelligence platform and pitched it as a startup prototype at UPC's entrepreneurship initiative.
- **Tech Stack:** PySpark, LLMs, Microsoft Fabric, Azure Data Factory, Power BI, GraphDB

Anomaly Detection in Diesel Train Cooling Systems (Project Link) — *Sep 2023 – Dec 2023*

- Developed unsupervised models for anomaly detection in train cooling systems for SNCB (Belgian National Railway).
- **Tech Stack:** Python, SQL, Scikit-learn, Pandas, Numpy, Matplotlib, APIs, Tableau

Brain Tumor Detection and Classification Using CNN (Project Link) — *Sep 2021 – Dec 2021*

- Achieved 98% accuracy in brain tumor detection using a custom CNN model on MRI images.
- **Tech Stack:** TensorFlow, Keras, Python, OpenCV

AI-Based Disease Prediction System (Project Link) — *Mar 2021 – Jun 2021*

- Built a web-based symptom checker predicting over 40 diseases with machine learning algorithms.
- **Tech Stack:** Python, Scikit-learn, Django, ReactJS, REST APIs

Languages

English: C2 (Professional Working Proficiency) — **Chinese (Mandarin):** B2 (Upper-Intermediate) — **French:** A2 (Beginner)