

Aisha Khatun

Toronto, ON, Canada
aysha.kamal7@gmail.com • linkedin.com/in/tanny411

EDUCATION

Masters of Mathematics in Computer Science

University of Waterloo (Advisor - Dan Brown)

Grade - 96%

Sep 2022 - Aug 2024

Core Courses: Probabilistic Inference and Machine Learning, Advanced Artificial Intelligence, Knowledge Graphs

Bachelor of Science in Computer Science and Engineering

Shahjalal University of Science and Technology (Advisor - Md. Saiful Islam)

Grade - 97%

Jan 2016 - Mar 2020

Core Courses: Data Structure and Algorithms, Object Oriented Programming, Software Engineering and Design Patterns, Neural Networks, and Deep Learning

SKILLS

Languages: Python, Kotlin, Java, SQL, SPARQL (Proficient); Javascript, C++, Scala, PHP (Comfortable)

Tools: Numpy, Pandas, Spark, Hadoop, Airflow, Scikit-learn, PyTorch, HuggingFace, Plotly, D3.js, Tableau, Power BI

Tech: HTML/CSS, ReactJS, Flask, MongoDB, PostgreSQL, Git, GitLab, Gerrit, Phabricator, Docker, Streamlit

PROFESSIONAL EXPERIENCE

Data Engineer

Wikimedia Foundation, Canada

Dec 2025 - Present

- Working on maintaining and modernizing the Wikimedia Foundation data platform. Exposing analytics APIs. Creating and maintaining ETL processes for research, experiment platform, movement insights and all other teams within the foundation.
- Skills/Tech: Scala, Python, ETL, Airflow, Kafka, DataLake, Project Design and Management.

Software Development Engineer

Amazon, Canada

Dec 2024 - Dec 2025

- Working on durable, highly available, low-latency, distributed log service that supports critical AWS database systems. Responsible for end-to-end service development, improvement, deployment, and on-call rotations.
- Launched a new feature to automatically tag and manage high volume customers which included writing design docs, developing durable, low-latency APIs while ensuring smooth customer experience.
- Skills/Tech: Kotlin, Java, Unit and Beta Testing, Low-level API design, Project Management.

Data Scientist

Wikimedia Foundation, USA

Feb 2023 - June 2024

- Enhanced Wikipedia **link recommendation system** by implementing a universal language tokenizer, achieving over 90% precision in non-Latin languages and expanding support to 10 additional languages.
- Reduced deployment strain by consolidating 300+ language-dependent models into a single high-performing **language-agnostic XGBoost model**.
- Increased the quality of Wikipedia articles by creating an **automated copy-editing pipeline** that detects 80% more misspellings than hand-curated lists in 100+ language Wikipedias.
- Skills/Tech: Python, PySpark, Sklearn, XGBoost, Machine Learning, ETL, Hadoop, Airflow, Unit Test.

Data Analyst

Wikimedia Foundation, USA

Apr 2021 - Aug 2022

- Advised business decisions to minimize data import speed and query timeouts in Wikidata Query Service by **analyzing terabytes of data** from Wikidata and user queries.
- Implemented a **pipeline** to periodically calculate and save metrics, effectively **monitoring** graph and query **statistics**.
- Skills/Tech: Python, Scala, Spark, SQL, SPARQL, Airflow, Docker, Big Data, Data Wrangling, ETL, Unit Test.

Machine Learning Engineer

Therap BD Ltd, Bangladesh

Mar 2020 - Sep 2020

- Improved in-office attendance application with **high-accuracy face detection** in image and video footage.
- Streamlined care home operations by developing an **OCR system** that extracts measurements from pulse oximeter images for swift COVID-19 detection.
- Boosted care home safety and accessibility by implementing a **machine learning based fall detection system** using inertial sensor readings from smartwatches.
- Skills/Tech: PyTorch, Machine Learning, Computer Vision, AWS Sagemaker, Google Cloud Platform, Docker, Research, Data Visualization, Reporting.

PROJECTS AND INTERNSHIPS

Data Scientist, Wikimedia Foundation, Internship

- Identified unique Lua modules across 300+ language Wikipedias through **data analysis and machine learning based source code similarity**.
- Skills/Tech: Machine Learning, Data Analysis, Big Data, Reporting.

Evaluation of LLM Consistency and Robustness, Master's Thesis Project [↗]

- Conducted an in-depth analysis of **misinformation, bias, and robustness** to prompt variations in 37 LLMs on sensitive topics, including conspiracy theories and stereotypes.
- **Publication:** Reliability Check: An Analysis of GPT-3's Response to Sensitive Topics and Prompt Wording
- Skills/Tech: Python, HuggingFace, Vertex AI, Amazon Bedrock, PyTorch, Tableau, Data Analysis, and Visualization.

Authorship Attribution, Bachelor's Thesis Project [↗]

- Developed baseline Bengali **language models**, curated the largest dataset in Bengali Literature, **designed deep learning architectures**, and analyzed the effects of tokenization on **downstream tasks**.
- **Publication:** Authorship Attribution in Bangla Literature via Transfer Learning using ULMFiT (**Best Poster Award**)
- Skills/Tech: Fast.ai, PyTorch, Keras, Deep Learning, Natural Language Processing, Data Curation.

Knowledge Graph - Language Model Integration, Research Course Project [↗]

- Improved factual accuracy by modifying the **deep learning based architecture** of a BERT-based language model and incorporating Wikidata entity embeddings. Performed **ablation studies** with text and entity evaluation.
- Skills/Tech: PyTorch, SPARQL, Data Generation, Natural Language Processing, Deep Learning, Language Models.

Image Captioning, Course Project

- Integrated CV and NLP techniques to develop an image captioning **deep neural network from scratch**, leveraging pre-trained models to achieve state-of-the-art performance across several large datasets.
- Skills/Tech: PyTorch, Computer Vision, Natural Language Processing, Deep Learning.

ACHIEVEMENTS

- **Barbara Hayes-Roth Award** for Women in Math and Computer Science, University of Waterloo, 2023
- **Graduate Research Scholarship**, University of Waterloo, 2022-2024
- **International Master's Excellence Award**, University of Waterloo, 2022-2024

- **Secure and Private AI Scholarship from Facebook**, one of the top 5% candidates worldwide, 2019
- **Top 10** positions in multiple National and Inter-University Girls **Programming Contests**, 2017-2019
- **Ranked 34th in ACM ICPC ASIA Regional Programming Contest**, Dhaka Site 2018

VOLUNTEER WORK

- **Research Co-Lead at Tech+UW**, leading the research team to identify campus trends on demographics, club interactions, and the impact of Tech+ programs and events.
- **Coordinated GSOC and Outreachy internships** for the Wikimedia Foundation, ensuring intern engagement and retention by reporting progress, hosting AMAs, and managing communication.
- **Mentored Waterloo Undergraduate students** in a Generative AI project for the Directed Reading Program, focusing on conducting literature reviews and applying LLMs for computational creativity tasks.
- **Led an interactive AI workshop** at WiCSCon, guiding participants through Data Science and Machine Learning algorithms with practical use cases.