Aisha Khatun

+1 226 899 8911 | aysha.kamal7@gmail.com | linkedin.com/in/tanny411
github.com/tanny411 │ Google Scholar | tanny411.github.io │ Kitchener, ON, Canada

---

SKILLS

Languages: Python, SQL, SPARQL, C++ (Proficient); Javascript, Java, Scala, PHP (Comfortable)
Tools: Keras, PyTorch, Fastai, Tensorflow 2, HuggingFace, Numpy, Pandas, Spark, PySpark, Hadoop, Airflow, Plotly, Tableau, Power BI, MongoDB, PostgreSQL, Git, GitLab, BitBucket, Gerrit, Phabricator
Web Tech: HTML/CSS, JQuery, NodeJS, ReactJS, ExpressJS, Flask

---

PROFESSIONAL EXPERIENCE

Wikimedia Foundation, USA (Remote-Contract)   Feb 2023 - Present
Research Data Scientist
- Addressing deployment bottlenecks and improving the Wikipedia link recommendation system by creating a language-agnostic model to replace the 300+ individual language-dependent models. Used Python, Sklearn, and Wikipedia2Vec. (Project, Code)
- Increased the standard of articles through automated copy-editing in Wikipedia by curating and detecting commonly misspelled words in 100+ languages. Used Python to extract and analyze Wiktionary data and to detect misspellings in Wikipedia. (Report, Code)

University of Waterloo, Canada   Sep 2022 - Present
Graduate Research Student
- Analyzed the capabilities and limitations of LLMs (Large Language Models) in assessing and updating sensitive statements (e.g. stereotypes, conspiracies, etc). Used HuggingFace, Python, and Tableau. (Paper, Paper)
- Identified appropriate LLMs by creating a dashboard based on fact-checking, bias detection and instruction-following abilities across 37 open and closed source models. (Project)
- Currently formulating story worlds and seamless entertaining stories by tweaking the information retention and critical assessment ability of LLMs based on aforementioned research findings.

Wikimedia Foundation, USA (Remote-Contract)   Apr 2021 - Aug 2022
Data Analyst
- Informed business decisions to reduce data import speed and query timeouts in Wikidata Query Service by analyzing Wikidata to find large and most frequently queried subgraphs. Analysis was done using PySpark and SQL (Reports)
- Enabled regular Wikidata graph and query statistics monitoring by creating a pipeline using Airflow, Hadoop, and Spark that performs data analysis, calculates, and saves metrics regularly. (Code)

Wikimedia Foundation, USA (Remote)   Dec 2020 - Mar 2021
Outreachy Intern (Data Science)
- Reduced redundancy in Wikipedia Lua modules by creating a pipeline to identifying unique modules for centralization in Abstract Wikipedia through data analysis and source code similarity analysis. Used Python, SQL, Flask, and unsupervised machine learning. (GitHub, MetaWiki, Phab)

Therap BD Ltd, Bangladesh   Mar 2020 - Sep 2020
Machine Learning Engineer
- Improved in-office attendance application with high-accuracy face detection in image and video footage.
- Optimized care home efficiency by implementing an OCR system that extracts measurements from pulse oximeter images for swift COVID-19 detection.
- Increased care home safety and accessibility by developing an ML-based fall detection system using inertial sensor readings from smartwatches.

SUST NLP Lab, Bangladesh   Nov 2018 - Jun 2020
Research Assistant
- Advanced Bengali NLP's foundation by curating several pre-training datasets and training and fine-tuning multiple versatile language models with various tokenization methods.
- Improved SOTA in Bengali Authorship Attribution by curating a large authorship attribution dataset, developing deep learning architectures, applying transfer learning, and analyzing the effects of tokenization and pre-training datasets on downstream tasks. (GitHub, Paper)

EDUCATION

Masters of Mathematics in Computer Science (Thesis)   Aug 2022 - April 2024
University of Waterloo, Waterloo, Canada

Bachelor of Science in Computer Science and Engineering   Jan 2016 - Mar 2020
Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh

PUBLICATIONS

- Aisha Khatun and Daniel G. Brown. *Reliability Check: An Analysis of GPT-3's Response to Sensitive Topics and Prompt Wording.* TrustNLP Workshop, ACL 2023.
- Aisha Khatun, Anisur Rahman, Md Saiful Islam, Hemayet Ahmed Chowdhury, and Ayesha Tasnim. *Authorship Attribution in Bangla Literature (AABL) via Transfer Learning using ULMFiT*. ACM Journal, ACM TALLIP 2022.
- Aisha Khatun, Anisur Rahman, Hemayet Ahmed Chowdhury, Md. Saiful Islam, and Ayesha Tasnim. *A Subword Level Language Model for Bangla Language*. Springer Conference, IJCCI 2020.
- Aisha Khatun, Anisur Rahman, Md. Saiful Islam, and Marium-E-Jannat. *Authorship Attribution in Bangla literature using Character-level CNN*. IEEE Conference, ICCIT 2019.

PROJECTS

- **Image Captioning** - Merged CV and NLP techniques to develop an image captioning deep neural network from scratch. Used Pytorch to leverage pre-trained models for text and image inputs.
- **Online Study Portal** - Developed a social networking web app to make group interaction easier and organized using the LAMP stack. Features include authentication, group file storage and organization, instant messaging, file searching, notifications, posting, and commenting.
- **Personal Tracker** - A MERN stack web app to track various personal activities from health, expenses, and habits to life goals. Includes a dashboard and choice of activity to track.
- **Others**: Facial keypoint detection (GitHub), Flight delay detection (GitHub), Food review classification (GitHub), FMNIST classification (GitHub).

ACHIEVEMENTS AND VOLUNTEER WORK

- Barbara Hayes-Roth Award for Women in Math and Computer Science, 2023
- One of 54 Outreachy Interns from 1000+ applicants through Open Source contributions, 2020
- Secure and Private AI scholarship from Facebook for the Privacy-Preserving AI course, 2019.
- Runner up Best Research Poster Award, ICBSLP 2019
- National Scholarship awarded by the Education Board, Government of Bangladesh.
- Received awards in several Programming Contests, notably the National Girls Programming Contest (2017 and 2018) and ACM ICPC ASIA Regional Programming Contest, Dhaka Site 2018.
- Directed Reading Program (DRP) mentor for undergraduates in ML research project, 2024
- Grad Panelist in SPARCS workshop. Discussed research and opportunities in CS with underrepresented students in Grades 9-10 across Canada, June 2023. [Slides]
- Conducted a hands-on AI workshop at UWaterloo for the WiCS Conference 2023, Jan 2023.
- Google Summer of Code and Outreachy coordinator for Wikimedia Foundation to organize events and help applicants, mentors, and interns in all steps pertaining to the internship, 2021.
- Panelist for 2 sessions and project presenter at WikidataCon, Oct 2021 [Sessions, Video]
- Conducted a series of IEEE Beginner's Machine Learning Workshop, in 2020. [GitHub]

BOOK ANOTHER APPOINTMENT AFTER INCORPORATING THESE FEEDBACK AND SEND A JOB POST TO HELP
ANALYZE THE RESUME IN CONTEXT