Monoshiz Mahbub Khan

Personal website ◇ LinkedIn profile
monoshizmk@gmail.com

EDUCATION

PhD in Computing and Information Sciences
Rochester Institute of Technology
August 2021 - Present
(Expected: May 2026)
CGPA: 3.93 (out of 4.00)

BSc in Computer Science and Engineering
University of Dhaka
January 2016 - January 2020
CGPA: 3.55 (out of 4.00)

PUBLICATIONS

- Khan, M. M., & Yu, Z. (2024). Approaching Code Search for Python as a Translation Retrieval Problem with Dual Encoders. *Empirical Software Engineering*, 30(1), 1-28. DOI: 10.1007/s10664-w024-10580-3
- Khan et al. Efficient Story Point Estimation With Comparative Learning. arXiv preprint arXiv:2507.14642 (2025).

WORK EXPERIENCE

- **Intern**   **June 2024 - August 2024**
  ABB
  Mannheim, Germany
  – Developed an end-to-end Named Entity Recognition (NER) pipeline using a range of techniques, including traditional NLP methods, ML models, deep learning models, LLMs and generative AI tools
  – Final NER model showed an improvement in F-1 score of 0.36 over the initial NER model
  – Used various tools including PyTorch, spaCy, scikit-learn, Hugging Face, MLflow
  – Conducted under the supervision of Nika Strem as part of the DAAD RISE Professional Program 2024

- **Graduate Research Assistant**   **Fall 2021 - Fall 2023, Fall 2024, Fall 2025**
  Lab of Human-In-the-Loop Software Engineering
  Rochester Institute of Technology
  Supervisor: Dr. Zhe Yu
  – Conducted research on **code search** and **comparative learning**, using deep learning, NLP and ML methods
  – Also explored research topics involving LLMs, image processing, software engineering and generative AI
  – Mentored Masters students on thesis projects: guided experimental design, advised research direction, and provided feedback on thesis writing

- **Graduate Teaching Assistant**
  Rochester Institute of Technology
  – **IDAI-720: Research Methods for Artificial Intelligence**   **Spring 2024**
    ∗ Instructors: Dr. Zhe Yu & Dr. Esa Rantanen
    ∗ Graded assignments and final projects, and hosted office hours
  – **IDAI-710: Fundamentals of Machine Learning**   **Spring 2025, Spring 2026**
    ∗ Instructor: Dr. James Heard
    ∗ Graded assignments, hosted office hours, conducting review classes

DOCTORAL RESEARCH EXPERIENCE

Code Search   2021 - 2024

Research project focusing on retrieving programming language artifacts related to some natural language queries from a pool of possible programming language artifacts, using dual encoder models. Model is built in Python using TensorFlow and Keras modules. This model showed an average improvement of 10.03% over state-of-the-art methods in terms of MRR scores. The research was conducted under the guidance of Dr. Zhe Yu. This work has been published in Empirical Software Engineering Journal (EMSE) and presented at FSE 2025 in the journal-first track.

Comparative learning   2023 - Present

Research project focusing on modeling learning comparative judgments for Agile story point estimation. Experiments involved using GPT2, SBERT, and other deep learning and large language model (LLM) based structures and traditional machine learning methods. The framework was built using TensorFlow modules. The proposed model showed an average increase of 21.84% in Spearman's rank correlation coefficient scores over state-of-the-art models. This project has also involved conducting human subject studies to complement the findings of the deep learning experiments. The research has been conducted under the guidance of Dr. Zhe Yu.

COLLABORATIVE RESEARCH EXPERIENCE

Explainable image classification   2024
Image processing and Explainable AI-based research project focusing on explaining a pre-trained VGG model's classification decisions on face image data. This work involved fine-tuning a pre-trained VGG model on SCUT face image data for classification, and using the model's gradients on the images to explain why the model made those decisions.

Comparative learning for face image attractiveness   2024
Research project focused on modeling comparative learning on face image data. This work involved using the comparative judgment framework with a pre-trained VGG model as the encoder to predict a ranked preference order for the images.

Comparative learning for image captioning   2024 - Present
Research project focused on modeling comparative learning on image and associated caption data. This work involved using the comparative judgment framework on this multi-modal data to predict whether a paired image and text caption are likely to be connected.

Outdated comment detection for repository commits   2024
Research project focused on detecting whether the comment associated with repository commits are up-to-date or outdated after new commits. This work involved the use of various deep learning structures, including dual encoders.

Bangla Text Summarization   2019 - 2020
A research project on constructing a dataset for the task of abstractive text summarization in Bangla, and constructing a deep learning based model capable of using said dataset. The model was written in Python using TensorFlow modules. The research was conducted as the final year research project at University of Dhaka under the supervison of Dr. Muhammad Asif Hossain Khan, collaborating with another final year student.

MENTORSHIP & SUPERVISION EXPERIENCE

- **Graduate mentor, REU Site: Trustworthy AI Workshop 2025**
  Faculty mentor: Dr. Zhe Yu
  Mentored a visiting student on a research project focused on outdated comment detection in repository commits. Mentorship responsibilities included -

  – Guiding the overall direction of the research project
  – Contributing to experimental design and methodology
  – Providing detailed feedback on literature review, experiment execution, and report writing

- **Informal Graduate Mentor, Human-in-the-Loop Software Engineering Lab, RIT**
  Mentored two Masters students on thesis projects involving comparative learning on image-caption data using multi-modal inputs. Support included -

  – Guiding overall research direction and experiment planning
  – Offering technical assistance (e.g., code-level help and architecture design)
  – Providing regular feedback in weekly meetings
  – Currently supporting extension of this work into a peer-reviewed article submission

COURSE PROJECTS AND SKILLS LEARNED

- Kabaddi (2016) [C++]
- Trapped (2017) [JAVA, Network programming]
- Musyc (2018) [JAVA, SQLite, Android, Full-stack]
- EasyML (2018) [Python, JavaScript, Flask, Full-stack]
- Pharmassistant (2018) [Python, JavaScript, Flask, NoSQL (mongoDB), Full-stack]
- CSEDU Project Hub (2019) [Python, JavaScript, Flask, Full-stack]
- BackPack (2022) [JAVA Spring, AngularJS, Full-stack, Agile, Scrum]

TECHNICAL STRENGTHS

Programming languages	Python, Java, R, C, C++, JavaScript
Machine Learning & AI	TensorFlow, Keras, PyTorch, scikit-learn, LLM fine-tuning, RAG
MLOps & Data Engineering	MLflow, Airflow, PySpark, Docker
Frameworks & Databases	Flask, Spring, Angular, SQL (Oracle, SQLite), NoSQL (MongoDB)
Tools & Methodologies	Git (GitHub, Azure DevOps), LaTeX, Agile, Scrum

SCHOLARSHIPS
- DAAD Rise Professional Program 2024

M. M. Khan
....
Signature