ANIRUDH KHATRY

akhatry@cs.utexas.edu | anirudhkhatry@gmail.com
linkedin.com/in/anirudh-khatry | **github**.com/anirudhkhatry | **anirudhkhatry.com**

EDUCATION

University of Texas at Austin — Started 2024
Doctor of Philosophy (Ph.D.) in Computer Science.

Veermata Jijabai Technological Institute, Mumbai, India — 2017-2021
Bachelor of Technology (B. Tech.) in Information Technology. — GPA: 8.97/10

INTERESTS

Program Synthesis, Information Retrieval, Machine Learning for Programming Languages and Software Engineering, NLP.

RESEARCH EXPERIENCE

Microsoft — Bangalore, India
Research Fellow, Program Synthesis (PROSE) Team — August 2022 - Present

- Conceptualized and built the natural language to code feature for the **Power Query M** language, used for wrangling tables in **Excel**, **Fabric** and **PowerBI**.
- Collaborated towards building the **Copilot** experience as a part of the Power Query experience in **Fabric** and **Excel**.
- Devised two state-of-the-art strategies TST^R (**EMNLP-Findings '23**) and COOPER (Under submission) for optimal dynamic prompt construction aiding in-context learning for natural language to code tasks.
- Developed Alternate Task Technique (ATT) (Under submission), a generalized framework to post process LLM outputs using alternate tasks that improved performance on low resource languages, like Power Query M, by **13%** and bridges the gap between performance for low and high resource languages.
- Developed Adapted Dense Retrieval (ADDER) (Under submission) framework for Information Retrieval tasks using dense embedding for efficient code retrieval in low-resource settings.

Microsoft Research — Bangalore, India
Research Intern — August, 2021 – July, 2022

- Assisted with the development of Landmark-based Robust Synthesis (LRSyn) framework (**PLDI '22**), that extracted data from semistructured formats from images like forms and HTML documents.
- Led early discussions with the Machine-to-Human (M2H) team at Microsoft for productionizing LRSyn in **Bing Ads**.
- Built a tool for data extraction from forms for the Finance India team at Microsoft that reduced the invoice turnaround time by **50%**.

Samsung Research — Bangalore, India
Research Intern — May, 2020 - July, 2020

- Worked with the On-Device AI team to improve system performance using Reinforcement Learning.
- Built a State-Of-The-Art Multi-Agent Deep Q-network leveraging prioritized experience replay (PER) and time-bound dynamic reward functions.
- Designed a multi-agent multi-target simulation environment for benchmarking performance.

PUBLICATIONS

Semantically Aligned Question and Code Generation for Automated Insight Generation.
A. Singha, B. Chopra, **A. Khatry**, S. Gulwani, A. Henley, V. Le, C. Parnin, M. Singh, and G. Verbruggen.
LLM4Code at Internation Conference on Software Engineering **(LLM4Code, ICSE '24)**. [Link]
Best Paper Award 🏆

TST^R: Target Similarity Tuning Meets the Real World.
A. Khatry, S. Gulwani, P. Gupta, V. Le, A. Singha, M. Singh, and G. Verbruggen.
Findings of Conference on Empirical Methods in Natural Language Processing, 2023.
(**EMNLP-Findings '23**). [Link]

Landmarks and Regions: A Robust Approach to Data Extraction.
S. Parthasarathy, L. Pattanaik, **A. Khatry**, A. Iyer, A. Radhakrishna, S. Rajamani, and M. Raza.
International Conference on Programming Language Design and Implementation (**PLDI '22**). [Link]

From Words to Code: Harnessing Data for Program Synthesis from Natural Language.
A. Khatry, J. Cahoon, J. Henkel, S. Deep, V. Emani, A. Floratou, S. Gulwani, V. Le, M. Raza, S. Shi, M. Singh and A. Tiwari.
Microsoft Machine Learning, AI & Data Science Conference (**MLADS '23**). [Link]

PREPRINTS

An Empirical Study of Validating Synthetic Data for Formula Generation
U. Singh, J. Cambronero, S. Gulwani, A. Kanade, **A. Khatry**, V. Le, M. Singh, and G. Verbruggen. [Link]

COOPER: Learning what to teach language models for code generation.
A. Khatry, S. Gulwani, V. Le, M. Singh, and G. Verbruggen. [Link]

Augmented Embeddings for Custom Retrievals.
A. Khatry, Y. Bajpai, P. Gupta, S. Gulwani and A. Tiwari. [Link]

Alternate Task Technique for Natural Language to Code in Low-Resource Languages.
A. Khatry, J. Cahoon, J. Henkel, S. Deep, V. Emani, A. Floratou, S. Gulwani, V. Le, M. Raza, S. Shi, M. Singh and A. Tiwari. [Link]

PROFESSIONAL EXPERIENCE

Human Rights First — Remote
Machine Learning Engineer — May, 2021 – July, 2021

- Developed a war-crime detection tool using social media channels along with 30 change-makers from around the world.
- Fine-tuned a **Distil-RoBERTa** model for binary classification of war crimes that obtained **80%** accuracy in war crime detection from social media channels.
- Spearheaded the development of a novel two-stage prediction pipeline for multi-label classification of war crimes.

Pexabyte Technology Solutions — Remote
Programming Analyst Intern — May, 2019 – July, 2019

- Designed and developed a **Enterprise Resource Planning (ERP)** tool utilizing MySQL as the back-end database.
- Collaborated closely with product managers to gather requirements, understand business needs, and translate them into functional specifications for the ERP tool.
- Implemented indexing to reduce the query execution time by **25%**.
- Provided training and documentation for end-users and support teams, ensuring a smooth transition and ongoing maintenance of the ERP tool.

SKILLS

Computer Languages: Proficient: Python, C#, SQL. **Familiar:** Java, C++, Javascript.
Software and Tools: PyTorch, MongoDB, AzureML, MySQL, WPF Applications.

PRESENTATIONS AND TALKS

Copilot for Data Integration, Microsoft, 2024.
NL to Code in low resource settings, Software Analytics Research Group, Singapore Management University, 2023.
Low Resource Data Extraction, Technical Advisory Board, Microsoft, 2022.
Azure Machine Learning Workshop, Microsoft, 2022.

VOLUNTARY SERVICES

Program Committee:

OOPSLA 2024
ASE Industry Showcase (ASE, 2024)
NuCLeaR Workshop (AAAI, 2024)
SRW Workshop (ACL, 2023 and 2024)
TRL Workshop (NeurIPS, 2023)

Student Volunteer: POPL, 2024.
Cloud Volunteer: Google's Developer Student Club, VJTI, 2021.