ANIRUDH KHATRY

anirudhkhatry@gmail.com — **linkedin**.com/in/anirudh-khatry — **github**.com/anirudhkhatry — **anirudhkhatry.com**

EDUCATION

Veermata Jijabai Technological Institute, Mumbai August, 2017— May, 2021
Bachelor of Technology (B. Tech.) Overall GPA: 8.97/10
Major: Information Technology

INTERESTS

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

RESEARCH EXPERIENCE

Microsoft Redmond, US (Remote)
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**.
- Devised two state-of-the-art strategies TST^R (**EMNLP-Findings '23**) and COOPER (Under submission to ACL ARR) for optimal dynamic prompt construction aiding in-context learning for natural language to code tasks.
- Developed Alternate Task Technique (ATT) (Under submission ACL ARR), 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) (Submitted to ICLR '24) 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

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

S. Parthasarathy, L. Pattanaik, **A. Khatry**, A. Iyer, A. Radhakrishna, S. Rajamani, and M. Raza. *Landmarks and Regions: A Robust Approach to Data Extraction.* In Proceedings of the 43rd ACM SIGPLAN International Conference on Programming Language Design and Implementation (**PLDI '22**). [Link]

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

A. Khatry, Y. Bajpai, P. Gupta, S. Gulwani and A. Tiwari. *Augmented Embeddings for Custom Retrievals.* **Under Submission to ICLR, 2024**. [Link]

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. Alternate Task Technique for Natural Language to Code in Low-Resource Languages. **Under Submission to ACL ARR**. [Link]
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. From Words to Code: Harnessing Data for Program Synthesis from Natural Language. Microsoft Machine Learning, AI & Data Science Conference (**MLADS '23**). [Link]

PROFESSIONAL EXPERIENCE

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

- Collaborated with 30 change-makers to develop a war-crime detection tool using social media channels.
- 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

Invited talk at Singapore Management University to Prof. Dr. David Lo's Software Analytics Research (SOAR) Group.
Presenter at Technical Advisory Board, Microsoft, 2022.
Presenter for Azure Machine Learning Workshop, Microsoft, 2022.
Presenter at the Model United Nations Workshop at Institute of Chemical Technology (ICT), 2020.

VOLUNTARY AND LEADERSHIP SERVICES

Student Volunteer for POPL, 2024.
Member of Program Committee at NuCLeaR Workshop at AAAI, 2024.
Member of Program Committee at TRL Workshop at NeurIPS, 2023.
Member of Program Committee at SRW Workshop at ACL, 2023.
Cloud Volunteer for Google's Developer Student Club, VJTI, 2021.
Director General, VJTI Model United Nations, 2021.
Debate Judge for Hysteria, Don Bosco College, 2020.
Public Relations Head, VJTI Cultural Committee, 2020.
Research Head, VJTI Model United Nations, 2020.