Atharva Naik

arnaik@andrew.cmu.edu atharva-naik Atharva Naik
https://atharva-naik.github.io/ Google Scholar

Experience

2022 – · · · · **Research Assistant, Carnegie Mellon University** Automated code review generation, Code translation, and Interactive reflection questions for database optimization tutoring.

2021 **Research Intern, Technische Universität Darmstadt** Neural Network Arcthitecture for Faithful Interpretability in NLP.

Research Intern, Adobe RL agent for Creative Human-Human Collaboration.

Research Intern, University of Alberta Neuro-Symbolic Fuzzy Logic-based Reasoning for Explainable Natural Language Inference.

2019-2020 **Student Researcher, Autonomous Ground Vehicle (AGV) Group** Path Planning and Localization for Autonomous Driving.

Education

2022 – 2024 **M.S Computer Science, Carnegie Mellon University** in Language Technologies.
GPA: 4.08/4

2018 – 2022 **B.Tech. Computer Science, Indian Institute of Technology, Kharagpur**
GPA: 9.66/10

Research Publications

Conference Publications

1. **A. Naik**, S. Das, J. Vedurada, and S. Aditya, "Sync: A structurally guided hard negative curriculum for generalizable neural code search," in ***AACL***, 2023.
2. Z. Wu, Z. X. Zhang, **A. Naik**, Z. Mei, M. Firdaus, and L. Mou, "Weakly Supervised Explainable Phrasal Reasoning with Neural Fuzzy Logic," in ***ICLR Poster***, 2023.
3. Y. Xie, **A. Naik**, D. Fried, and C. Rose, "CMTrans: Improving Code Translation with Comparable Corpora and Multiple References," in ***EMNLP Findings***, 2023.
4. S. Bv, J. A. Patel, **A. Naik**, Y. Butala, S. Sharma, and N. Chhaya, "Towards Enabling Synchronous Digital Creative Collaboration: Codifying Conflicts in Co-Coloring," in ***CHI Extended Abstracts***, 2022.
5. B. Santra, S. Roychowdhury, A. Mandal, *et al.*, "Representation Learning for Conversational Data using Discourse Mutual Information Maximization," in ***NAACL***, 2022.
6. Y. Wang, S. Mishra, P. Alipoormolabashi, *et al.*, "Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks," in ***EMNLP***, 2022.
7. R. Mukherjee, **A. Naik**, S. Poddar, S. Dasgupta, and N. Ganguly, "Understanding the Role of Affect Dimensions in Detecting Emotions from Tweets: A Multi-task Approach," in ***SIGIR***, 2021.

Preprints

1. A. Rao, S. Vashistha, **A. Naik**, S. Aditya, and M. Choudhury, "Tricking LLMs into Disobedience: Understanding, Analyzing, and Preventing Jailbreaks," 2023. arXiv: 2305.14965 [cs.CL].

Skills

Coding	Python (expert), C/C++, Bash (familiar), Javascript (novice)
Frameworks	PyTorch, HuggingFace, Fairseq, NLTK, spaCy, Tensorflow, FastAPI, Flask, Django, PyQt5, Jupyterlab, OpenCV, Git

Awards

- 2022 **2nd place** at Deep Learning Labs OpenAI GPT-3 Hackathon
- 2021 **DAAD WISE Scholarship**
- **MITACS Globalink Scholarship**
- **Bronze** at Inter IIT Technology Meet: Member of IIT Kharagpur contingent
- 2019 **2nd** place at Intelligent Ground Vehicle Competition (IGVC)
- 2018 All India Rank **1248** in JEE Advanced and **1618** in JEE mains among 1M candidates
- **Kishore Vaigynaik Protsahan Yojana (KVPY) Scholarship**

Other Projects

Interactive Reflection Prompts for Collaborative Programming in Education (Ongoing)

- Used LLMs to generate alternative solutions for teaching database optimization using SQL.
- Explored the educational value of strategic and contextual interactive reflection questions in a collaborative SQL programming activity with 35 students through pre/post tests.

Reference Free Evaluation of Automated Code Review Generation (Ongoing)

- Revisited evaluation and modeling assumptions for automated code review generation.
- Proposed pragmatics-inspired, model-based reference-free evaluation metrics to measure relevance, informativeness, correctness, and readability of reviews.
- Leveraged contrastive learning to model relevance and abstractness lexicons for informativeness.
- Proposed metrics align better with human judgment compared to current metrics like BLEU

Multilingual Multimodal Contrastive Learning for Unified Representations (Ongoing)

- Using multimodal, multilingual contrastive learning to learn unified representations for code and multiple natural languages for jointly modeling multilingual code generation and translation.
- Data augmentation for code using semantics preserving syntax level transformations

NBQA: Retrieval Augmented Jupyter Notebook Completion

- Leveraged prompting-based query rewriting and semantic code retrieval to augment code generation for Jupyter notebooks.
- Achieved improvements in next-cell code prediction over CodeBLEU and API call coverage