

# Manisha Mukherjee

mmukherj@andrew.cmu.edu  
manishamukherjee.github.io  
*School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA*

## INTERESTS

---

- Artificial Intelligence for Code, Security, Large Language Models, Generative AI, Code Generation, Security

## EDUCATION

---

- Carnegie Mellon University** Pittsburgh, PA  
PhD in Software Engineering *Current*  
– Advisor: Dr. Vincent Hellendoorn
- The Pennsylvania State University** University Park, PA  
M.S. in Computer Science and Engineering  
– Advisor: Dr. Thomas La Porta  
– Thesis: “Determination of real-time traffic flow parameters in different devices based on QoI requirements”
- West Bengal University of Technology** Kolkata, India  
B.Tech in Computer Science and Engineering

## EXPERIENCE

---

- Carnegie Mellon University** Pittsburgh, PA  
Graduate Research Assistant August 2017-Current  
– LLM4Code, augmenting LLMs for generating secure code, developer forum analysis, mining software repositories, code recommendation and code security.
- Adobe Research** San Jose, CA  
Research Intern Summer 2024  
– Implemented retrieval-augmented generation (RAG) techniques using knowledge graphs to enhance AI assistant
- Lawrence Livermore National Laboratory** Livermore, CA  
Research Intern Summer 2023  
– Deep learning for HPC error log classification
- Lawrence Livermore National Laboratory** Livermore, CA  
Research Intern Summer 2022  
– Deep learning for power net load forecasting
- Fujitsu Labs America** Sunnyvale, CA  
Research Intern Summer 2021  
– Mining data from Kaggle and analyzing popularity metrics for Kaggle kernels to guide the enhancement of pipelines produced by Fujitsu’s AutoML core pipeline synthesis
- Idaho National Laboratory** Idaho Falls, ID  
Research Intern, INL Wireless Security Institute Summer 2020  
– Wireless signal classification and threat detection. Designed and built a tool for spectrum monitoring and threat visualization in real-time.
- Fujitsu Labs America** Sunnyvale, CA  
Research Intern Summer 2019

- Semantic code search and code recommendation using Deep Learning.

### Cisco Systems, Inc

Software Engineer, ASR9K group

San Jose, CA

October 2014- August 2017

- Developed a distributed router with an intelligent control plane (ICON) by using SDN technologies to dis-aggregate control plane and data plane.
- Wrote ODL (OpenDayLight) plugins and REST APIs to control several underlying router clusters.

### The Pennsylvania State University

Graduate Research Assistant (Networking and Security Research Center)

University Park, PA

August 2012-August 2014

- Worked on extracting features from traffic videos and determining real-time traffic flow parameters in different devices. Optimized the transfer and processing of this data based on Quality of Information (QoI) requirements.

### Capgemini India Pvt. Ltd.

Senior Software Engineer

Kolkata, India

July 2011-July 2012

## SCHOLARSHIPS AND AWARDS

---

- Sansom Graduate Fellowship in Computer Science 2024
- Presidential Fellowship in SCS 2023
- Carnegie Institute of Technology Dean's Fellowship 2017
- Frank J. Marshall Graduate Fellowship 2018
- Center for Integrated Healthcare Delivery Systems (CIHDS) Scholarship 2012

## TEACHING

---

- **Teaching Assistant** at Carnegie Mellon University Fall 2022  
*Principles of Software Construction Objects, Design, and Concurrency (17-214)*
- **Head Teaching Assistant** at Carnegie Mellon University Fall 2019  
*INI MSIT Project Practicum (14-798)*
- **Teaching Assistant** at Pennsylvania State University Fall 2013  
*Communication Networks (CMPEN 362)*

## SELECTED PUBLICATIONS AND PATENTS

---

- [1] M. Mukherjee and V. J. Hellendoorn, "Sosecure: Safer code generation with rag and stackoverflow discussions", *arXiv preprint arXiv:2503.13654*, 2025.
- [2] M. Mukherjee, S. Kim, X. Chen, D. Luo, T. Yu, and T. Mai, "From documents to dialogue: Building kg-rag enhanced ai assistants", *arXiv preprint arXiv:2502.15237*, 2025.
- [3] M. Mukherjee and V. J. Hellendoorn, "Skill over scale: The case for medium, domain-specific models for se", *Proceedings of the 2025 IEEE/ACM Second International Conference on AI Foundation Models and Software Engineering*, 2024.
- [4] V. J. Hellendoorn, J. Tsay, M. Mukherjee, and M. Hirzel, "Towards automating code review at scale", in *29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '21)*, 2021.
- [5] M. Mukherjee, M. Bahrami, and W. P. Chen, "Source code retrieval", in *US Patent Application 17/085,894*, 2020.
- [6] M. Mukherjee, J. Edwards, H. Kwon, and T. F. La Porta, "Quality of information-aware real-time traffic flow analysis and reporting", in *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, IEEE, 2015, pp. 69–74.
- [7] M. Mukherjee, "Determination of real-time traffic flow parameters in different devices based on qoi requirements", in *MS Thesis*, 2014.