

Reshmi Ghosh

Websites: <https://www.linkedin.com/in/reshmi-ghosh/>

Personal Webpage: <https://reshmighosh.github.io/>

Email : reshmighosh@microsoft.com

Mobile : (412)499-5751

Education

Carnegie Mellon University

Pittsburgh, PA

Doctor of Philosophy(**Dean's List); **Focus: Deep & Machine Learning for Climate Change;**

Oct 2021

M.S. Engineering & Public Policy (2nd master's, completed with Ph.D.); **Focus: Machine Learning for Policy**

M.S. Advanced Infrastructure Systems; **Focus: Machine Learning for Infrastructure Systems;**

Dec 2017

University of Mumbai

Mumbai, India

B.Tech Engineering; Minor: Numerical Methods

May 2016

Professional Experience

Microsoft Corp.

Cambridge, MA

Applied Scientist II (Ph.D.), Microsoft AI Development Acceleration Program

Oct 2021 - Present

- Integration Team(**ongoing****): Office ML (Office Product Group), Product: **Microsoft Forms**

1. Integrating ChatGPT + GPT 3.5.1 models to Microsoft Forms using AugLoop API to assist users in creating Surveys/Quizzes in cold-start and interactive experience using a chatbot.
2. Designing innovative ways to mitigate hallucinations (at low & high temperature settings, Top P) and developing custom pre-processors with PyTorch (+ small embedding models) to help GPTx models query from niche contexts.
3. Developing novel evaluation paradigms to test generative outputs from the LLMs on factuality, relevancy, fluency, and accuracy.

- Integration Team FY23H1: Office of Applied Research (OAR) + ODSP Cortex Team + MSAI, Product: **Viva Topics**

1. Researched, developed and implemented novel probabilistic graph analysis frameworks in PySpark to understand the complex interactions in organization-level information searching and sharing mechanisms using efficiently queried TBs of IDEAs Outlook, Teams Chat & Meetings datasets, & **Viva Topics** telemetry.
2. Designed value metrics from aforementioned graph-based analysis frameworks to monitor **310K MAU**(in-production by Summer 2023) and help Viva Topics trial customers convert to premium buyers.
3. Drafted documents & assisted in Privacy Review process, to get approval for managing highly-sensitive-customer data. Also architected a Data Flow Diagram design to extract and analyze customer data responsibly for understanding search behavior of users.

- Integration Team FY22H2: Office Products Group (OPG) - Docs ML Team, Products: **PowerPoint, Word, Excel**

1. Developed and assisted in shipping a PyTorch-based GRU model to extend Adaptive Floatie's functionalities from 3 to 5, leading to a 12% increase in MAU (**to 4.5 Million**) worldwide.
2. Supported OXO's effort to simplify commanding in O365 applications (PowerPoint & Word) by developing database querying scripts in SCOPE & C# to analyze Gigabytes of usage data. Investigated frequently used command probabilistic distributions in Python by consuming the extracted usage data.
3. Researched and shipped a novel 'vectorized reward & penalty' based loss function in PyTorch in dogfood to test the impact of **positional-bias from users** for improved command predictions and model performance.

- Integration Team FY22H1: Azure Edge & Platform - Signal Quality (SKY) Team, Product: **Azure services**

1. Leveraged an end-to-end pipeline in Python and PySpark by MSRA to monitor anomalies in Azure services by consuming TBs of Service Level Indicator data in Synapse.
2. Analyzed customer impacting Azure service outages by contributing to development of a python package to determine statistical correlation metrics between customer responses and SLI service anomalies.

Other Microsoft Contributions

- **Tech Lead, University of Massachusetts-Microsoft Capstone 2023 (NLP)**: Leading a research project with graduate students from University of Massachusetts, Amherst, to understand what types of linguistic and syntactic features are transferred during the transfer learning process in GLUE tasks
- **Researcher, Collaboration with Office of Applied Research 2023 (HCI)**: Researching the complex human behavior in organizations of searching, editing, finding, and sharing information using AI-based knowledge management tool (Viva Topics) to understand the nuances of modern workplace setting
- **Researcher & Advisor, University of Massachusetts-Microsoft Capstone 2022 (NLP)**: Research and developed analysis for Topic Segmentation methods for unstructured datasets using BiLSTM, BERT, & RoBERTa models along with graduate students from University of Massachusetts, Amherst. The research led to a short paper published during NeurIPS 2022, & long paper to be published at IntelliSys 2023

Publications

- **IntelliSys September 2023** Topic Segmentation for Conversational Data; [Reshmi Ghosh](#), Harjeet Singh Kajal, Sharanya Kamath, Dhuri Shrivastava, Samyadeep Basu, Soundaranjan Srinivasan (Microsoft collaboration)
- **NeurIPS December 2022** Topic Segmentation in the Wild - Topic Modeling for Semi-structured & Unstructured Data; [Reshmi Ghosh](#), Harjeet Singh Kajal, Sharanya Kamath, Dhuri Shrivastava, Samyadeep Basu, Soundaranjan Srinivasan (Microsoft collaboration). [link](#)
- **ICML July 2021** Reconstruction of long-term historical demand data, [Reshmi Ghosh](#), Michael Craig, H. Scott Matthews, Constantine Samaras [link](#)

Programming Skills

Languages: Python(PyTorch, TensorFlow NumPy, Pandas, scikit-learn, NLTK, Matplotlib, PySpark), KQL, SCOPE, SQL, Matlab, C#*

Technologies: OpenAI, Augmentation Loop, Azure ML, Synapse, COSMOS, Azure Databricks, Visual Studio, AWS, Git, MySQL, PostgreSQL *Familiar

Relevant Coursework

Data Structures & Algorithms, Ph.D. Deep Learning, Practical Data Science, Machine Learning, Data Analytics, Business Intelligence, Decision Analytics, Risk Analysis, Python for Developers, Data Warehousing, A/B Testing (Udacity), Applied Data Analysis (Statistics), Reinforcement Learning

Invited Speaker

- WiDS Western Massachusetts (Environmental and Climate Justice using AI), *March 2023*
- WiMLDS Bay Area (Use of Contextual Bandits for Online Products), *July 2022*
- ODSC East 2022 Speaker for Women in Ignite Session Panel *April 2022*
- Guest Lecturer (Deep Reinforcement Learning), School of Computer Science, Carnegie Mellon University *April 2022*

Achievements

- Semi-Finalist CTO Open AI CODEX Challenge, Microsoft *Jan 2022*
- Dean's Fellowship recipient, Carnegie Mellon University *Aug 2018*
- Carnegie Mellon Merit Scholarship *March 2017*
- CMU Civil and Environmental Engineering department scholarship *Jan 2016*

Leadership

- Organizer, Women in Machine Learning and Data Science, Boston chapter
- Oxford University Artificial Intelligence Society Mentor 2022
- Conference Reviewer, Microsoft Machine Learning & Data Science Conference
- Scholarship Committee Reviewer, Women @ Microsoft