

# MILIND AGARWAL

magarwa@gmu.edu  
(410) 303 7659  
milind-agarwal.github.io

## EDUCATION

**Johns Hopkins University** — BS in Computer Science, 2021  
**Johns Hopkins University** — MSE in Computer Science, 2021  
**George Mason University** — PhD in Computer Science, Aug 2025

## RELEVANT EXPERIENCE

**PhD Research Assistant** (Natural Language Processing)

George Mason University — 2022–Present

- Improving language identification, optical character recognition (OCR), and language modeling (LLMs) for low-resource languages
- Collaborators: Meta, University of British Columbia, Technion

**Research Assistant** (Computational Biology)

Johns Hopkins University — 2018-21

- Developing visualization and automation techniques for multi-modal and multi-view healthcare data

**Microsoft RLOS Researcher** (Reinforcement Learning)

Microsoft Research — 2020

- Creating in-house suite of analysis and visualization algorithms to improve VowPal Wabbit reinforcement learning policies

## SELECT PUBLICATIONS

M. Agarwal, J. Otten, A. Anastasopoulos. SAGE: Script-Agnostic Language Identification. *Preprint on arXiv* 2024

**Best Paper Award** at MASC-SLL'24, Johns Hopkins University

M. Agarwal, A. Anastasopoulos. A Concise Survey of OCR for Low-Resource Languages. *AmericasNLP @ NAACL* 2024

QueerInAI et al. Queer In AI: A Case Study in Community-Led Participatory AI. *FAccT* 2023

**Best Paper Award** at ACM FAccT Conference

M. Agarwal, M. Alam, and A. Anastasopoulos. LIMIT: Language Identification, Misidentification, and Translation using Hierarchical Models in 350+ Languages. *EMNLP* 2023

## RESEARCH AWARDS

**Stanford SILICON Scholar**, 2024-25  
*Digitization of Kwak'wala language texts and speech corpus creation*

**Mason Research Scholar**, 2024-25  
*Modeling structure and layout to improve information extraction*

**Microsoft RLOS Award**, 2020  
*Analyzing and improving in-house reinforcement learning policies*

**Hopkins Pistrutto Fellow**, 2019-20  
*Visualization of complex multi-view and multi-modal healthcare data*

**Hopkins PURA Fellow**, 2019-20  
*Automating discovery of rare genetic disorders through large-scale data*

## SERVICE

**Queer in AI**, 2022-present  
Conference Social Organizer

**Reviewer**, 2022-present  
NLP Venues: ARR, ACL, EMNLP etc.

## LANGUAGES

Python, C/C++, Java  
SQL, R, Tableau  
HTML/CSS, Javascript