Milind Agarwal

magarwa@gmu.edu   Website — LinkedIn — GitHub
Updated May 5, 2025

SUMMARY

Creative and detail-oriented professional with PhD in Artificial Intelligence (NLP) and strong industry and academic experience in software development, analytics, and frontier technologies. Skilled in translating complex problems into elegant, scalable solutions.

EDUCATION

Johns Hopkins University   *2017-2021*
BS in Computer Science, Applied Mathematics & Statistics   GPA: 3.8
MSE in Computer Science   GPA: 3.9

George Mason University
Ph.D. in Computer Science   *Expected May 2025*
Research Advisor: Antonis Anastasopoulos   GPA: 4.0

INDUSTRY EXPERIENCE

Researcher (Reinforcement Learning), Microsoft Research   2020
- Developed an in-house RL-policy visualization library for complex analysis, enabling risk assessment and behavior forecasting by researchers and product teams.
- Implemented pruning methods to build faster learners based on our hierarchical policy visualizations.
- At NeurIPS 2020, presented a demo of our exploration, policy training, feature importance, and action distribution functions.

Juni Fellow, Juni Learning (edTech startup)   2021-2022
- Delivered personalized 1:1 instruction in algorithms, data structures, USACO & ML to 20+ students (ages 8–18).
- Recruited and onboarded instructors, assessing pedagogy and technical depth to uphold Juni's teaching standards.

ACADEMIC EXPERIENCE

Research Assistant (AI/NLP), George Mason University   2022-*present*
- Aligned and adapted multilingual LLMs to specific data-constrained settings using RLHF (reinforcement learning through human feedback) and efficient finetuning methods like LoRA (low-rank adaptation).
- Developed novel multilingual modeling techniques for improving the efficiency and scalability of ML methods on tasks like sequence prediction, translation, optical character recognition etc.
- Used computer vision and multimodal LLM methods for automatic digitization of the Smithsonian Archives for American Indigenous languages (Stanford SILICON Awardee)

Research Engineer (ML for Healthcare), Johns Hopkins University   2018-20
- Systematically analyzed healthcare provider policies and information from local stores and pharmacy websites to model real-time critical updates. This fed into our Whatsapp and web-based chatbot to answer questions around COVID-19.
- Worked collaboratively with physicians to design and implement an ML pipeline, WebSeq, to analyze personal, clinical, and genetic data. My efficient pipeline reduced processing times from 1 month to just 1 day, leading to an exponential decrease in case backlog.
- Developed innovative unsupervised ML models to analyze complex breast cancer healthcare data and cellular blood data from thousands of patients in the United States and the United Kingdom.

SKILLS	**Programming Languages:** Python, R, SQL
	AI: LLMs, Deep Learning (PyTorch), Reinforcement Learning, Computer Vision, NLP
	Tools: Docker, Celery, Airflow, Flask, Django, LaTeX
	Cloud / HPC: AWS, GPU Clusters, SLURM & Schedulers
	Data: Pandas, NumPy, Scikit-learn, Visualization (Matplotlib, Seaborn)

RESEARCH PUBLICATIONS

S. Blouir*, C. Watkins*, **M. Agarwal***, P. Acharya* et al. SLP Sidekick: An Open-Source, Multilingual Speech Therapy Platform. *To appear in Interspeech 2025 Demo Track*
∗ Equal Contribution

M. Agarwal, J. Otten, A. Anastasopoulos. Script-Agnosticism and its Impact on Language Identification for Dravidian Languages. *NAACL 2025* [Paper]

M. Agarwal, A. Anastasopoulos. AILLA-OCR: A First Textual and Structural Post-OCR Dataset for 8 Indigenous Languages of Latin America. *Comput-EL 2025*

M. Agarwal, D. Rosenblum, A. Anastasopoulos. Developing a Mixed-Methods Pipeline for Community-Oriented Digitization of Kwak'wala Legacy Texts. *Comput-EL 2025*

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

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

M. Agarwal, K. Ghimire, J. Cogan, Undiagnosed Disease Network, J. Markle. WebSeq: A Genomic Data Analytics Platform for Monogenic Disease Discovery. Journal of Bioinformatics and Systems Biology 6 (2023): 01-09. [Paper]

S. Ahmadi, **M. Agarwal**, and A. Anastasopoulos. PALI: A Language Identification Benchmark for Perso-Arabic Scripts. *VarDial 2023 @ EACL 2023.* [Paper]

A. Poliak et al. Collecting Verified COVID-19 QA Pairs. *NLP-COVID Workshop 2020*

AWARDS	**Stanford SILICON Practitioner Award**	2024-25
	Doctoral Research Scholar Award	2024-25
	Summer PhD Research Grant	2023
	Dorothy F. Sheppard Award for Outstanding Service to JHU Residential Life	2021
	Best Project Award, Intuitive Surgical	2020
	Self-supervised Contrastive Image Classification with Image Sentences	
	Microsoft Open Source Fest Award	2020
	Contextual Bandits Data Visualization using VowPal Wabbit	
	Joseph C. Pistritto Research Fellowship	2019-20
	Visualization research for multi-view data (Battle Lab)	
	Provost Undergraduate Research Award	2019
	Developing WebSeq to find genetic causes of rare autoimmune disorders (Markle Lab)	