408-839-1978
linkedin.com/in/marcusma
marcus.ma@att.net

Marcus Ma

Programming Language and Framework Proficiencies

- **Languages**: Python, Java, Javascript, SQL, C++, PHP, C#, Git, MongoDB, MATLAB, R, Bash
- **Libraries and Frameworks**: pytorch, scikit-learn, .NET, AWS, Docker, TensorFlow, React, JavaFX

Education

Georgia Institute of Technology

MS in Computer Science

AUGUST 2023 – PRESENT

- Relevant Coursework: Game AI, Advanced Natural Language Processing
- Master's Thesis: leveraging novel NLP techniques and LLM's for the research area of cross-genre authorship attribution and verification

BS in Computer Science, Intelligence and Theory Concentration

AUGUST 2020 – MAY 2023

- 4.0 GPA (graduated summa cum laude, Dean's List every semester)
- Relevant Coursework: Machine Learning, Computer Vision, Natural Language Processing, Algorithms Honors, Database Systems, Data Structures, Advanced Linear Algebra, Number Theory, Multivariable Calculus, Computational Machine Learning, Combinatorics

Teaching Experience

Teaching Assistant for CS 4650: Natural Language Processing

JAN 2023 - PRESENT

- Assist in leading classes of 75+ students through NLP, covering topics from encoders to present-day LLM's
- Create and maintain one of three programming assignments where students create an AI chatbot via LSTM's
- Hold weekly office hours and grade 200+ assignments

Relevant Work History

Amazon Web Services

Software Developer Intern: Optical Networking Automation Team

MAY 2023 – AUG 2023

- Spearheaded the transition of optical device networking communication from internal Amazon servers to the more scalable and maintained Native AWS environment
- Created a DNS device hostname resolution script that was deployed to hundreds of existing optical devices in the AWS fleet
- Added a robust testing and logging system that monitors the hostname resolution script that automatically initiates rollback in case of failure

Amazon

Software Developer Intern: Postgres SQL Team

MAY 2022 – AUG 2022

- Worked in Postgres source code to implement an optimization on the SQL Join command that used semi-join hash filtering to remove unnecessary table rows early in the execution tree
- Improved the speed of certain SQL join commands by up to 36% and integrated filter decisions for the merge join strategy into Postgres query planner, improving AWS Postgres' overall TPC-H database benchmark performance by ~5%
- Created an internal SQL test suite to verify improvement and deployed these tests to the entire AWS codebase, with future work looking to contribute semi-join filtering optimization to an upcoming Postgres community release

Publications and Conference Presentations

- **Ma, M.**, Kim, C., Hall, K., Kim, J. It Takes Two to Avoid Pregnancy: Addressing Conflicting Perceptions of Birth Control Responsibility in Romantic Relationships. *Proc. ACM Human-Computer Interaction*. 7, CSCW2, Article 282 (October 2023).
- **Ma, M.** and Glass, L. (2022, Oct 13-15). BedWord: Computer-generated, human-edited TextGrids for faster sociolinguistic transcription. NWAV 50 Conference, Palo Alto, CA.
- Seeburger, D., Xu, N., Godwin, C., **Ma, M.,** Keilholz, S., Schumacher, E. (2022, Apr 24-27). Identifying the Neural Mechanisms of Zone State Performance using Time-varying Functional Connectivity Methods. CNS 2022 Conference, San Francisco, CA.

Research

NLP-X Lab (Natural Language Processing)

Undergraduate Research Assistant

JANUARY 2022 - PRESENT

- Researching the performance of current transformer-based authorship attribution (AA) models in low-data, many-author situations
- Created the largest publicly available cross-genre authorship dataset across several different social media sites, including Twitter, Reddit, StackOverflow, and Tumblr
- Currently building a few-shot cross-domain authorship model that identifies stylistics agnostically relative to source medium

Co-Well Lab (Human-Computer Interaction)

Contraception Collaboration between Couples – Accepted to CSCW 2023

AUGUST 2021 – MAR 2023

- Investigated the expectations placed on women in relationships for taking birth control pills consistently, with findings indicating responsibility lies entirely on women
- Developed an app prototype designed to split the responsibility of taking pills more equitably between partners and interviewed nine couples to investigate the intersection of contraception responsibility and technology, findings published in CSCW 2023

DAAD RISE Germany Research Internship (Artificial Intelligence)

RISE Scholar Recipient and Researcher

JULY 2022 – OCT 2022

- Chosen as one of 320 scholarship recipients out of 1365 applicants to research in Germany at Forschungszentrum Jülich
- Investigated how to use McCormick convex-concave relaxations to improve verifiable robustness of neural networks, with hopes of later applying these findings to optimize resource consumption of green energy process systems
- McCormick relaxations were shown to be ~40% tighter than the alternative Interval Bound Propagation on several internal tests

CoNTRoL Lab (Neuroscience)

Undergraduate Research Assistant

AUGUST 2021 – MAY 2022

- Investigated the Quasi-Periodic Pattern relationship between different attention networks in the brain
- Wrote a script that parsed raw four-dimensional MRI scan data into detection of QPP's, with specific attention paid to the DMN and TPN brain network regions
- Ran data pipelining scripts, QPP detection algorithms, and MRI scan visualizations in Matlab and Bash on Georgia Tech's neuroscience Linux cluster

Language and Politics in the New South (Sociolinguistics)

Vertically Integrated Projects Member and Ambassador

JAN 2020 – APR 2023

- Investigated the relationship between accent and political leanings, based on interviews from ~100 Georgia Tech undergrads
- Interviewed, transcribed, and analyzed ~20 participants and visualized findings with R scripts
- Created an automated transcription tool, Bed Word, that leverages existing speech-to-text NLP models to make auto-generated TextGrid files, improving human transcription time by ~200%
- Integrated Bed Word with an existing linguistics website hub, which to date has been used to automatically transcribe 500+ audio hours and 300+ transcriptions for the wider linguistics community