

Emory University

PhD Computer Science and Informatics

During a discussion with my advisor, Professor Wang, on integrating Large Language Models (LLMs) into healthcare research at the University of Texas Rio Grande Valley, I realized that adoption in this sector lags behind tech and marketing industries. The demand for high accuracy and correctness in healthcare is a critical factor, limiting the speed at which LLMs are being adopted. My research goal is to address these challenges by developing domain-specific LLMs enriched with comprehensive medical knowledge to enhance reliability and trustworthiness. By overcoming these obstacles, LLMs can more effectively process vast amounts of medical data, support decision-making leading to new drug discoveries, and automate routine tasks like documentation. Beyond healthcare, I am also deeply interested in exploring the broader implications of LLMs and other machine learning models across diverse domains. This includes areas such as federated learning for secure and privacy-preserving model training, developing robust and interpretable AI systems to address ethical challenges, and applying ML techniques to tackle complex optimization problems in industries like finance, education, and climate science. I am particularly excited by the opportunity to explore intersections between fundamental ML research and real-world applications, tailoring solutions to meet unique domain-specific requirements. My background in Artificial Intelligence, Machine Learning, and Deep Learning positions me well to expand into these versatile research directions. I am eager to collaborate with faculty at Emory University whose pioneering work spans both theoretical advancements and impactful applications, contributing to innovative research that drives progress across industries while improving societal outcomes.

I have been working on my Master's thesis since March 2024, titled *"Implementing and Evaluating an AI-Powered Visual Decision Support System to Improve Antibiotic Usage Among Physicians, Featuring a Built-in Alert System."* This project, supervised by Professor Xuan Wang and Professor Tamer Oraby, uses real-world longitudinal healthcare data and is developed in Python and Tableau. It is designed to alert physicians when antibiotics are overprescribed, promoting responsible antibiotic stewardship. Additionally, I am integrating a Large Language Model to assist physicians unfamiliar with technical jargon, enabling them to navigate the dashboard and understand relevant terms. I have implemented k-means clustering and anomaly detection to identify patients with unusual prescription patterns. We are also developing a predictive model to forecast future trends in antibiotic usage; this model is currently under evaluation. My abstract for this research was accepted at the 2024 Decision Sciences Institute Annual Conference in Arizona, where I presented my work on November 25, 2024.

In September 2022, I completed my undergraduate thesis titled "Classification and Prediction of Heart Diseases Using Machine Learning Algorithms," which was later published on ArXiv on the 5th of September, 2024. Using the UCI heart disease dataset, I applied Machine Learning models such as logistic regression, k-nearest neighbors, support vector machine, and a multilayer perceptron. GridSearchCV optimization revealed k-nearest neighbors as the best-performing model with an accuracy of 89%. While I was pleased with the model's effectiveness for this specific problem, I realized that scaling up to a broader range of medical conditions would require more advanced data processing techniques. This insight has motivated me to explore Large Language Models to handle the complexity and diversity of larger medical datasets, aiming to enhance predictive capabilities across various healthcare applications. This project was

a rewarding introduction to practical applications in data science, and its publication has allowed me to share my findings with the broader research community.

In addition to research, I teach MATH 1342: Elementary Statistics Method at the University of Texas Rio Grande Valley, instructing a class of 16 undergraduates since August 2024. I have achieved an average student evaluation score of 4.8. I collaborated with the School of Mathematical and Statistical Sciences at the University of Texas Rio Grande Valley, serving as a Graduate Research Assistant and mentor for participants in the Research Experience for Undergraduates (REU) program. Together with the REU students, I wrote a manuscript for publication on “*Using Convolutional Neural Networks To Predict The Order of Fractional Partial Differential Equations*”, and applied advanced deep learning techniques for numerical approximation and analysis. We generated data through numerical solutions of Fractional Partial Differential Equations using Monte Carlo simulations, dividing it into training, validation, and testing subsets for model evaluation. The CNN model achieved high predictive accuracy, with correlation coefficients between predicted and actual values of 0.975, 0.968, and 0.955 for the three subsets, respectively, as well as low Root Mean Squared Error (RMSE) values. This project not only allowed me to delve deeper into deep learning applications in numerical analysis but also enabled me to guide and inspire undergraduate students in research.

Beyond academia, I am deeply engaged in technical and leadership roles. As a local organizer for the Teens in AI Ghana Hackathon in February 2024, I coordinated online events aimed at introducing young students to AI principles. This experience underscored the impact of technology education in empowering youth and strengthened my commitment to leadership within the data science community. I am an active member of Black in AI and AnitaB.org since September 2024, and the Association for Computing Machinery since July 2024, which allows me to stay connected with industry advancements and contribute to ongoing discussions shaping the field of computing and AI. Recently, I served as an abstract reviewer for the NeurIPS 2024 Women in Machine Learning workshop in September 2024, where I assessed novel research contributions in machine learning and made recommendations based on workshop criteria. I am also the Graduate Senator and the Alternate Representative on the Student Government Association and University of Texas System Advisory Council respectively, advocating for student interests and enhancing campus life.

My desire to pursue a Ph.D. in Computer Science and Informatics at Emory University stems from my dedication to advancing machine learning, artificial intelligence, and data science, particularly within the realm of Large language models. I am passionate about shaping the technological landscape of the future, where Large Language Models will serve as pivotal tools across industries. Emory University offers an ideal environment to pursue this vision, with its robust research community and exceptional faculty.

I am particularly inspired by the work of Professors Abeed Sarker, Fei Liu, and Jinho Choi, whose research in natural language processing (NLP) and biomedical informatics aligns with my academic goals. Professor Sarker's focus on applying NLP to healthcare data resonates deeply with my aspiration to leverage LLMs for improving healthcare delivery. Professor Liu's expertise in NLP methodologies inspires me to explore innovative ways to enhance the capabilities of LLMs in understanding and generating human language. Similarly, Professor Choi's contributions to data mining and machine learning motivate me to investigate scalable and efficient LLM systems. My ultimate goal is to contribute to the development of domain-specific LLMs, ensuring they are both reliable and ethically sound, particularly in high-stakes applications

like healthcare. At the same time, I am excited about the prospect of collaborating with other faculty members whose work spans theoretical and applied domains, such as optimization, reinforcement learning, natural language processing, and beyond. Emory University's breadth of expertise offers endless opportunities for interdisciplinary exploration, and I look forward to engaging with diverse researchers who can broaden my perspectives and shape my approach to addressing complex problems. By contributing to Emory University's vibrant research community, I aim to work alongside peers and mentors who share my dedication to advancing technology for societal benefit.

I am confident that my background in mathematics, statistics, and programming, coupled with my passion for data science, makes me a strong candidate for this program. I look forward to the opportunity to collaborate with faculty and peers at Emory University as I pursue my goal of driving meaningful technological advancements in healthcare and beyond.