My curiosity for interpretation drives me to pursue a PhD. To me, understanding a system means more than observing its outputs. It means reasoning about its internal structure, explaining its behavior, and communicating that understanding to others. I am drawn to probabilistic frameworks and how, when arranged correctly, they can model complex phenomena. Building classical machine learning models from the bottom up gave me transparency in their structure and certainty in their behavior. However, as architectures grew complex, their internal logic became increasingly opaque. Faced with unexpected model outputs, I saw that strong performance on traditional evaluation metrics can mask fragile reasoning. This opacity stands out to me as the reason behind the hesitation around deploying modern AI in high-risk environments.

I aim to pursue a PhD to continue studying machine intelligence and contribute to the field of interpretable and explainable AI. Through my research, I want to explore three core questions: (1) how we can evaluate models to validate reasoning, (2) how we can interpret complex architectures to improve learning dynamics, and (3) how we can create representations that preserve semantic meaning.

These questions surfaced during my early work on image classification. In Spring 2024, I joined Dr. Shahrukh Humayoun at the Human-Computer Interaction Lab (SFSU) to study hierarchical image classification involving a large number of classes. We used a dataset of 1,000 animal species drawn from ImageNet. We hypothesized that organizing classes into a hierarchy, like biological taxonomies, could help the model distinguish species based on morphological traits such as body shape and skin patterns. I structured the pipeline grouping species from broad families down to specific classes. To analyze misclassification trends and determine what drives model decisions, I designed an experiment to trace inference paths through the hierarchy tree. For misclassified examples, I identified the nodes where predictions diverged from ground truth. Using feature importance techniques such as LIME and SHAP, I highlighted regions in the image influencing the decision. We observed that errors were often driven by environmental correlations rather than the animal itself. Extending this analysis to correctly classified samples showed that the model often relied on the same correlations, illustrating how data nuances can mislead even advanced models.

In Fall 2024, recognizing my strong interest in language modeling, Dr. Anagha Kulkarni invited me to join her lab as a Research Assistant, where I also serve as the Teaching Assistant. For my master's thesis, I investigate whether language models can understand community assets and navigational challenges in STEM classrooms. We aim to identify cultural capital themes such as aspiration, resistance, and perseverance within student reflections to support STEM retention. These themes are highly context-dependent, raising a foundational question: Can language models truly understand semantic meaning, or are they simply predicting the next token based on syntactic patterns?

Baseline classification using BERT-based transformers achieved results, but embedding analyses revealed a lack of narrative awareness. Processing sentences independently caused a loss of context, and pretraining on general corpora left the model perplexed by the classroom writing style. To address this, I designed AWARE, a modeling approach structured around three forms of awareness. I established Domain Awareness through continued pretraining, aligning the model's vocabulary with student writings. To resolve the loss of global context, I engineered Context Awareness by reconstructing full essays and using pooling with a BiLSTM to create sentence embeddings that retain broader narrative context. Finally, I incorporated Cross-Theme Awareness to address overlapping themes. By reframing the task as multi-label classification and representing labels as 12-dimensional probability vectors, I enabled the model to capture co-occurring signals.

These decisions yielded a 2.1-point improvement in Macro F1, with gains for rare themes. My work was awarded the SF BUILD Agents of Change fellowship for three consecutive semesters. The pipeline is now used as an annotation tool for researchers. This research also culminated in my first manuscript, "AWARE Beyond Sentence Boundaries," submitted to ICMLA 2025.

Parallel to my thesis, I joined Xuman.AI as the Founding AI Engineer in Fall 2025. I worked on enabling voice agents to interact in a more human-like manner rather than rigid turn-based question-answer behavior. This work

addressed limitations in conversational ability, emotional expression, and information retention. These agents integrated multimodal inputs such as text, audio, and images to model human personalities and support low-latency interaction. Given the massive size of modern generative models, direct interpretation of internal mechanisms was infeasible. I turned to evaluation to ensure reliability. I designed evaluation frameworks for interactive agents, including LLM-as-judge mechanisms, to assess factual accuracy, safety, and behavioral consistency. This experience highlighted the lack of robust tools for interpreting and validating dynamic, generative behaviors.

My commitment to understanding complex systems extends to how I communicate these concepts to others. As a Graduate Teaching Assistant for Dr. Kulkarni's Natural Language Processing course, I expanded my duties beyond grading to leading full lectures on topics such as regression, n-grams, and language modeling techniques. I serve as an instructor for a Data Science and Machine Learning certificate program for Genentech professionals, where I have delivered five courses introducing machine learning methods for biotechnology applications. Outside the classroom, I also hold a student leader position focused on fostering inclusion, diversity, and providing aid to the campus community. I look forward to continuing this engagement at Georgia Tech, where I aspire to actively contribute to the academic community as both a researcher and an educator.

I am applying to the Machine Learning PhD program at Georgia Tech because it aligns directly with the questions that have guided all my work. It is exciting to see such extensive work being done here on the very questions of evaluation and semantics that matter to me.

I am eager to learn from Prof. Xu's work on evaluating language models, particularly in settings where the robustness and reliability of generated text are critical. My thesis work on AWARE approached this problem through context preservation in narrative understanding, while my work at Xuman.AI examined the same question through behavioral evaluation of language model responses. These experiences motivate my interest in adopting Prof. Xu's evaluation frameworks to determine whether models capture semantic meaning or rely on superficial signals across domains and users. My research interests also align with Prof. Ritter's work on robust language understanding and information extraction with minimal supervision. My work has frequently encountered domain shift and annotation constraints, and I want to learn strategies for building systems that generalize beyond curated benchmarks while using fewer labels.

Building on these interests, a PhD at Georgia Tech would give me the right training, mentorship, and research environment needed to pursue these questions deeply and grow as a researcher and educator committed to reliable and interpretable AI.