My Research Views. Despite amazing leaps in the performance of language models in generating coherent text and appearing to "understand" prompts to make useful outputs, these models often mimic patterns from data without truly grasping the underlying knowledge and struggle with tasks that require complex reasoning, causality, and common sense. This makes them inefficient in generalizing to new, unseen tasks and, more so, in a multilingual setting. We, humans, perceive the real world through text and, more importantly, through multimodal interaction by combining all our senses that profoundly and fundamentally shape our understanding, learning, and communication. It is, therefore, crucial to integrate the various modalities and languages into a single powerful model that is robust to adversaries and keeps evolving through continual learning. I plan to build upon the 'giant' advancements of large language models, developing artificially intelligent agents that not only process but truly comprehend real-world data while mirroring the human capability to systematically reason, explain, and generalize with commonsense intelligence. Through rigorous research and collaboration with leading experts in the field of Computer Science at Rutgers University, I aim to contribute to the development of more intuitive, truly intelligent, and efficient AI models that can seamlessly interpret, interact with, and learn from the multifaceted world around us.

Research Questions. Building upon these insights, I am driven by the following questions:
How can causal inference and knowledge representation in language models integrate real-world causality for enhanced reasoning capabilities?
How can we scale multiple modalities and languages into a single, extensible framework?
How can we enhance contextual understanding and compositional generalization capabilities in models?
How can we evaluate the reasoning capabilities of the models beyond standard benchmarks?
What approaches can be adopted to ensure explainability and reliability in language models, especially in understanding semantics and discerning causality?
What mechanisms can develop for models to reason with commonsense intelligence?
How can models be designed to withstand malicious inputs, misinformation, and bias?
What are the computational and data challenges in building such integrated models, and how can they be overcome?

From Theory to Research: My research experience in low-resource languages. My curiosity in understanding how closely intelligent machines can emulate human thinking and understanding led me to join the CUET NLP Lab. My research focused on developing large-scale corpora and devising language frameworks for advancement in low-resource languages, particularly Bangla. It included a comparative analysis of trade-offs in efficiency and accuracy of 16 POS tagging models for Bangla, resulting in a publication at ICO 2020. My undergraduate thesis involved creating a hierarchical classification system for Bangla news using an MLP-based framework, achieving high accuracy rates (98.18% and 90.63% for primary and subcategories) and creating a dataset of 76,343 news articles, with partial findings published at ICO 2021. Motivated by the findings, I further scaled up the dataset and explored state-of-the-art algorithms, confirming that ensembles of BERT models excel in performance. One of the key findings of our work is that monolingual models outperform multilingual ones and capture the intricacies of a particular language by adapting to the unique characteristics and complexities of that particular language. Furthermore, the optimal ensemble varies by task, suggesting that no one-size-fits-all model exists and that task-specific customization is crucial for achieving the best results. The work is currently in review for submission to an international journal.

Expanding Horizons: Venturing into Multimodal Learning. In 2020, driven by my interest in multimodal learning, I worked with one of my classmates on developing a sequential multimodal framework for real-time sign language detection for Bangla. We constructed a comprehensive dataset comprising 12.5k images spanning 49 classes and manually annotated the frames (images and videos). We utilized YOLOv4 to develop a real-time object detection model and a multimodal framework. I concentrated on two key aspects: firstly, the development of a comprehensive dataset, and secondly, the implementation and integration of language and speech components with the visual data. This involved transforming a series of images or video streams into coherent words and sentences. To enhance the clarity and structure of the generated sentences, we proposed three distinct signs to represent punctuation. The project, due to its novelty, ranked as one of the top 25 projects in AI for Bangla 1.0, a competition organized by the Bangladesh Computer Council.

Bridging Academia and Industry: Professional experience in language models. At Workera.ai, my involvement in integrating LLMs into the assessment pipeline highlighted the need for robust, accurate language models that avoid hallucination and can generate objective, analytical test questions. I observed significant issues with existing models, including repetitive patterns and a lack of fidelity to instructions, leading to a failure to balance creativity with precision. This experience deepened my understanding of language model performance nuances and reinforced my interest in creating robust, adaptable, and context-aware language models capable of generating plausible and faithful outputs.

Advancing Multimodal and Multilingual Understanding: My Journey with Vision Language Models. Being awarded the Fatima Al-Fihri Predoctoral Fellowship, I am collaborating with the UC Santa Barbara NLP Group. Inspired by the binding capabilities of text-to-image (T2I) models, my research focuses on multilingual concept coverage in T2I models. Our work has identified a crucial gap in existing evaluation methods, which often fail to capture the intricacies and potential inaccuracies yielding inconsistent or unfaithful scores and are thus questionable. To address this, we have introduced a metametric to evaluate existing evaluation metrics and a meta-benchmark, T2IScoreScore, containing natural and synthetic images generated using prompts with varying faithfulness. The work aims to develop a consistent and faithful metric for evaluating action and attribute binding in synthetic images.

Future plan. After completing my Ph.D., I plan to be a professor to contribute to the field of computer science by preparing the next generation of scholars and conducting meaningful research. I have been fortunate to be mentored by several people who guided me through my academic and professional career. Mentoring students and peers has been a central part of my undergraduate and professional life as I used to lead a group of 4 students at our university's computer club and assessment developers at my current workplace. Education lies at the heart of scientific progress. Resonating on Margaret Mead's quote on teaching, I want to teach the next generation how to think, not what to think.

Choosing the Rutgers University: Aligning Goals with Opportunities. The Computer Science Ph.D. program at Rutgers University has a strong reputation in AI and provides a perfect blend of courses and research labs (RuCCs, CBIM). My research interests are closely aligned with the expertise of several esteemed professors. I am interested in working with Dr. Sharon Levy on exploring mechanisms to identify bias and ensure fairness and trustworthiness in language models in a multilingual and multimodal setting. With Dr. Matthew Stone, I am eager to work on exploring knowledge representation and reasoning in language models. I am also interested in working on enhancing generalization capabilities and exploring reasoning and learning capabilities in large language models with Dr. Hao Wang. Dr. Karl Stratos and his work on text understanding and knowledge processing interest me in working with him on semantic representation and context-aware language models.

My Fitness to the Rutgers University. Drawing on my unique yet robust background in CSE, coupled with extensive hands-on experience in NLP, machine learning, and AI, I bring a unique blend of academic excellence, practical industry insights, and a commitment to innovation. Resonating with the mission of Rutgers University to provide for the educational needs of New Jersey, I plan to utilize my mentorship capabilities to engage in mentoring undergraduate and fellow graduate students actively to share my academic and professional experience, fostering a collaborative learning environment. I also believe in bringing opportunities to underrepresented communities through outreach programs through the organizational and leadership skills I have earned from various experiences. Under the guidance of Rutgers University leadership, I intend to expand my leadership skills by actively participating in outreach initiatives, including organizing workshops and volunteering to inspire young minds and promote diversity in STEM fields. Most importantly, I aim to collaborate with faculty members and fellow researchers on interdisciplinary, cutting-edge research that addresses real-world challenges. Through my research, I aspire to discover technologies and solutions that have practical applications and can be translated into real-world benefits. I don't believe in simply seeking opportunities; I am determined to create them. Armed with unwavering determination, resilience, self-discipline, diligent effort, and the right academic environment, which the Computer Science Ph.D. program at Rutgers University provides, I am confident I will make a lasting impact on the forefront of computer science and leave an indelible mark on the university's academic and research.