Statement of Purpose

My passion for Computer Science stemmed from an undergraduate project, Generative Adversarial Learning on Graphs, where I integrated adversarial learning methods, innovating in perturbation analysis and graph homogeneity evaluation to balance robustness and privacy. This experience affirmed my belief that building AI models required not only efficiency but also defense mechanisms against potential threats to ensure interpretability and transparency in real-world cases. I believe AI holds the potential to redefine human lives, yet instances of deepfake manipulation, algorithmic bias, and privacy violations prevent them from creating more real-world social impacts. In my PhD studies, my research goal is to bridge the gap between cutting-edge AI capabilities and their responsible applications, specifically by building **trustworthy AI systems** that integrate innovative frameworks, augmented data, and enhanced interpretability.

During my first year of master's studies (2023-2024) in the Computer Engineering program at Columbia University in the City of New York (GPA: 4.0/4.3), under the guidance of **Prof. Junfeng Yang** and **Prof. Xuan Di**, my research has primarily centered on AI-generated video detection and robust machine learning. Over the last year, I have undertaken four research projects, resulting in two first-author publications at top conferences, including Computer Vision and Pattern Recognition Conference (**CVPR 2024**) GenAI Workshop and Intelligent Transportation Systems Conference (**ITSC 2024**), with the remaining two manuscripts currently under submission to **CVPR 2025** and International Conference on Very Large Data Bases (**VLDB 2025**). **I am the first author of all these papers.** As a result of my great enthusiasm and strong motivation for interdisciplinary research and study, I am marked as one of **2024 Spring MS Honors Students** in CU and my research also be featured as **Columbia Engineering Research Highlight**.

Bearing this goal in mind, I worked with Prof. Junfeng Yang to explore the **Robust Detection on AI-Generated Video** in the Software Systems Laboratory at Columbia University. Specifically, I proposed **the world's first detection framework**, extending DIffusion Reconstruction Error (DIRE) to the video domain by incorporating temporal dimension information. For a reconstruction model like a video diffusion model with forward and reverse diffusion process, I measured the distribution gap to distinguish the AI-Generated videos and achieve up to 93.7% accuracy on videos generated by the state-of-the-art generative models. Nevertheless, we found inherent limitations in deep learning-based methods such as their inability to discover new patterns beyond what is encoded in the training data. This issue reinforced my desire to develop a more generalizable detection framework for LAVID, a Large Vision Language Models(LVLMs) based detection framework that could call external tools to extract additional information from the video to facilitate its determination. My method significantly improves the performance of out-of-distribution (OOD) data, making LVLMs detection AI-generated videos a reality. My work resulted in a submission to **CVPR 2025,** and a preliminary paper published in the **GenAI workshop** in **CVPR 2024**. My research also was marked as **Columbia Engineering Research Highlight**.

My research on video detection intrigued me that the reliability of AI models not only depends

on the underlying algorithm design, but also requires the model to have the ability to adapt to complex data distribution. This realization prompted me to explore cross-domain applications for improving the robustness of AI models under the supervision of **Prof. Xuan Di** at DitecT Lab at Columbia University. I started my research on **Out-Of-Distribution Generalization on Graphs**. I designed the Causal Adjacency Learning (CAL) method to identify robust causal relationships in spatiotemporal graphs. I demonstrated that capturing these causal relations significantly improves prediction performance on OOD test samples, even when causal learning is not directly incorporated into downstream tasks. The paper on this research was published at **ITSC 2024**, a top transportation conference. After further reflecting on the challenges of OOD generalization for graph data, I identified that a fundamental reason for the lack of generalization in current GNN models is the insufficient diversity of training data. Motivated by this, I proposed a novel data augmentation method, Counterfactual Augmentation. By applying the opposite gradient to the target class during the sampling process of a diffusion model, I generated counterfactual samples that blend features from the original and target classes. The paper based on this research was submitted to **VLDB 2025**.

In addition to framework innovations and data augmentation techniques, my multidisciplinary research has deepened my understanding of AI and computer science. At BAIR Lab, UC Berkeley, I applied causal inference to Mixture of Experts (MoE) systems, identifying critical layers in large language models (LLMs) and proposing an efficient layer-fusion method with minimal interference. At Illinois Institute of Technology, I developed robust node injection attacks in graph neural networks, enhancing adversarial resilience against defenses. At the National University of Singapore, I designed a spatio-temporal GNN to predict stock market trends using sentiment and corporate data. Beyond research on AI, I excelled in hardware-driven projects, such as building a MIPS CPU and a Bubble Bobble game system on an embedded platform with ARM CPUs and FPGA. These experiences have refined my technical expertise across software and hardware while inspiring my pursuit of advancing trustworthy AI through improved robustness, efficiency, and interpretability.

As a conscientious and critical thinker with a solid computer science background, I am driven to satiate my curiosity for building trustworthy AI systems. University of Pennsylvania University offers an exceptional academic environment where I can further my expertise and make meaningful contributions to advancing model robustness, ensuring data privacy, and promoting fairness. I hope to work with **Prof. Swati Gupta** on developing advance machine learning model for fairness and transparency, as her recent work on improving fairness of online decision making. I would additionally like to work with **Prof. Negin Golrezaei** on creating more resilient, equitable, and sustainable digital ecosystems. Besides, I hope to work with **Prof. Stefanie Jegelka** on robustness and for scaling machine learning algorithms. Last but not least, I would like to work with **Prof. Cathy Wu** on developing reliable, explainable and robust learning algorithms for safe and sustainable mobility. Ultimately, I hope my Ph.D. will shape me into an impactful leader in the field of trusted AI and a responsible researcher devoted to benefitting the world through social impact.