Statement of Purpose

Due to the prevalence of Internet technologies, cloud systems, and smartphone apps, we recurrently hear *buzzwords* like reliability, software testing, and cybersecurity. As a self-motivated software engineering student at *Xi'an Jiaotong* University (XJTU) focusing on secure computer systems, I have observed severe cybersecurity breaches and vulnerabilities in software applications, such as data security and retention, authentication bypass, personal information disclosure, legal compliance, *etc*. After completing multiple AI-based cybersecurity projects, I am convinced that integrating AI with cybersecurity is the future trend. Thus, this scenario offers a great opportunity for aspirant students like me to pursue an enduring career in cybersecurity. According to *Cybercrime Magazine*, 3.5 million cybersecurity professionals will be required by 2025, allowing me to become a potential contributor. Therefore, an advanced knowledge of information security, networks, and software systems is worthy of postgraduate study for my professional growth, and the MS program in Cyber Security Engineering at USC offers an ideal platform for my aspirations.

My initial exposure to AI-based security came when I joined Dr. *Peng Zhang's* cybersecurity research group at the Chinese Academy of Sciences (CAS) as an intern and investigated URL-based phishing website detection. Based on the research outcomes, my article has been accepted at CONF-SEML 2024. Under Dr. *Zhang's* guidance, I reviewed 50+ highly cited recent articles to explore suitable methods for my research. Through an intensive exploration, I determined that although SOTA models report impressive detection results, their performance is usually accompanied by massive resource requirements and long training and inference times. Thus, leveraging CNNs and attention mechanisms, I developed an efficient model to extract local features from various receptive fields and effectively optimized training and detection times. Next, I put the output of the convolutional layers into an RNN-based GRU model to shorten the time with efficient contextual information extraction. Finally, a multi-head attention layer was applied for weighted enhancement, followed by a fully connected layer for result output. Consequently, I accomplished a significant speed-up of 34.93% with an accuracy of 98.3%. In addition, I incorporated the model into a browser plugin for widespread use.

Boosted by my research, I joined Prof. *Wang Weiping's* team for summer research to create a model for generating a knowledge graph from countless unstructured Cyber Threat Intelligence (CTI) reports. It implies extracting Structured Threat Information eXpression (STIX) entities from each report and specifying the correlations between various STIX entities, such as STIX Relationship Objects (SROs). The primary challenge was accurately determining SRO between distant entities, like one being in the beginning and another at the end of the text. As traditional rule-based techniques for CTI report analysis cannot adequately tackle this problem, I turned to a deep learning model, Sentence-BERT (SBERT), to capture the contextual information and semantic meaning of sentences. I customized this model accordingly and achieved the task with 82.8% precision as opposed to 72.1% precision by a non-AI, rule-based algorithm.

Apart from my research, I developed a solid foundation for my advanced studies in cybersecurity by taking courses like data structures, operating systems, and networks at XJTU. Furthermore, I participated in various projects to develop secure system architectures and proactively prevent cyberattacks. For instance, I developed a secure communication software based on the RSA/AES algorithm as a capstone project in *Computer Networks*. This software can endure man-in-the-middle, replay, and denial-of-service attacks targeting *WebSocket* communication. These courses and projects expanded my horizons to identify potential security vulnerabilities.

The MS in Cyber Security Engineering at USC is a comprehensive program that inculcates in-depth information security techniques and tools to address contemporary challenges in this field. The

well-organized curriculum includes courses like *Trusted System Design, Analysis, and Development* (DSCI 525) to help me design secure systems using my software engineering skills. Similarly, core courses like *Security Systems* (CSCI 530) and *Security and Privacy* (DSCI 529) will teach me the threat models and mitigation techniques. I am also interested in *Directed Research* (DSCI 590) to learn contemporary research methods. Through hands-on experience, I hope to gain extensive research opportunities in this MS program and utilize them in my future plan to pursue a Ph.D. I firmly believe that the prestigious faculty, scholarly ecosystem, and internship opportunities in this MS program will allow me to fulfill my career aspirations.