Narrative Statement – ACSAC Conferenceship Application

My name is Haohan Yuan, and I am a second-year Ph.D. student in Computer Science at the ICS Department, University of Hawaii at Manoa, advised by Prof. Haopeng Zhang. My research focuses on natural language processing (NLP), specifically focusing on the challenge of making large language models (LLMs) trustworthy, faithful, and genuinely useful in real-world applications.

My research experience has equipped me with a strong foundation for this goal. For example, my work published at NAACL 2025, a top NLP conference, explored how to the reliability of text summarization systems when dealing with documents of unknown topics or from variant domains. This and other cross-disciplinary projects have trained me to connect methodologies from different fields.

My recent interest in security arose from working with large language model APIs. I observed that newer models, such as GPT-5 and Gemini-2.5, enforce stricter checks than earlier ones like GPT-4o. Text that was once processed may now be flagged as sensitive or harmful. This raised deeper questions for me about safety: how to design summarization systems that remain robust against adversarial or biased inputs, how to prevent leakage of sensitive or private information, and how to ensure models do not generate unsafe or misleading outputs. I am especially interested in techniques for adversarial robustness, privacy-preserving training, detecting and mitigating model misuse, and frameworks for evaluating trustworthiness of outputs. These are areas where I currently lack expertise, and I believe attending ACSAC is the best way for me to learn from experts and see how they tackle these problems.

Currently, my budget is tight, and I rely on my assistantship, so I wouldn't be able to afford this conference on my own. Getting this conferenceship would mean a great deal. It would allow me to learn so much about security and bring that knowledge back to my own NLP research and my lab group.

Thank you for considering my application. I hope to have the chance to join the ACSAC community, learn from the experts in this field, and find new ways to bridge the gap between NLP and security.

Best regards,

Haohan Yuan