Personal, Relevant Background, and Future Goals Statement

Summary of Personal, Relevant Background, and Future Goals Statement: My interest in the security and reliability of Artificial Intelligence (AI) began with a realization I made while conducting undergraduate research: while the widespread adoption of AI is a net positive for society, many systems are not designed with security and privacy in mind, leading to persistent flaws. AI permeates almost every aspect of our modern lives, making it vital to design and implement safe and reliable AI systems. This pressing research focus led me to pursue a PhD in Computer Science at the University of Wisconsin-Madison, where my goal is to conduct foundational research in provably secure and private AI systems.

Intellectual Merit

Early Research in Applied Machine Learning and AI Robustness

My path to research began at a small, rural high school in northern Michigan; in fact, my graduating class had only 15 people, with my twin sister and me as co-valedictorians. Motivated by a deep curiosity for computer science and mathematics but deeply cognizant of the cost of undergraduate education, I accepted a full-tuition Centralis scholarship from the Central Michigan University Honors Program. This decision was transformative, enabling me to graduate debt-free and, most importantly, providing me with the opportunity to work one-on-one with faculty on exciting research.

As a Centralis scholar in the Honors Program, I intentionally sought out interdisciplinary research opportunities that used AI and Machine Learning (ML) to solve real-world problems. After signing a major in Computer Science and completing core classes, I joined Dr. Jesse Eickholt's applied ML lab to work on automated fishery surveillance, a project that combined my deep interest in AI with my passion for nature. I began work on this research by developing tools to aid the Great Lakes Fishery Commission (GLFC), our partner organization, in statistically analyzing invasive fish movements. This work accelerated data collection efforts by the GLFC and led to a formal hazard analysis of the risks posed by invasive fish.

Completing this research prompted me to reflect on the importance of designing safe and reliable AI systems, as I recognized that even small flaws in our models could result in catastrophic failures to detect invasive species. Driven by this concern, I proposed and led the development of deep ML computer vision systems to continuously monitor bodies of water for signs of invasive fish: real-life intrusion detection systems for fish. Additionally, I collaborated with scientists from the GLFC and the United States Geological Survey (USGS) to develop a novel ML system to support selective fish passage and invasive species detection. Overall, this body of research resulted in **6 high-quality, open-source computer vision datasets**, **1 refereed research note**, **1 peer-reviewed conference publication**, and **3 peer-reviewed journal articles**, including one on which I was first author. The GLFC and USGS intend to use the models and systems developed in these projects to further automate invasive species detection and prevention, highlighting the real-world applications of this research.

Transition to Research in AI Security and Privacy

A critical problem with deploying these fish recognition systems *in situ* was the prevalence of natural adversarial noise – such as changes in light, water turbidity, or weather – that caused serious model misclassifications. Investigating the impact of adversarial noise in affecting AI models naturally led me to explore the field of adversarial machine learning and AI security.

To formalize this focus, I became one of the first Cybersecurity majors at Central Michigan University. Under the guidance of Dr. Qi Liao, I conceptualized and implemented two independent research projects in AI security. The first focused on fooling a machine learning spam detector using adaptive adversarial perturbations of word embeddings. My custom perturbation algorithm succeeded in increasing the spam evasion rate of a target classifier by 35%, and I published and presented these results as the first author at

the 2023 IEEE AIBThings conference. This research clearly showed me how the vulnerabilities I observed in environmental systems could be intentionally exploited by adversaries.

Driven to find other examples of unsafe AI, my second project investigated the use of large language models (LLMs) for system penetration testing. I successfully demonstrated that locally hosted LLMs running in resource-constrained environments could be used for effective, low-cost vulnerability analysis by both responsible security practitioners and adversaries. I then published and presented these results as the first author at the peer-reviewed 2024 IEEE CSR conference. My key observation from this work was the inherent unpredictability of unsupervised LLMs running in a loop, leading me to doubt whether AI should be completely trusted to act independently for agentic tasks. This insight underscored the necessity of developing provably secure and private AI systems.

To gain the expertise necessary to study provable security at a high level, I added a third major in Mathematics. Under Dr. Debraj Chakrabarti, I completed NSF-funded summer research on developing an explicit formula for Bergman kernels of n-dimensional monomial polyhedra, a special topic in complex analysis. Here, I worked with a team of undergraduate, graduate, and faculty mathematicians to develop a rigorous proof of our formula. I also implemented scientific Python programs to computationally validate our logical proofs. This research was published as a peer-reviewed article in the *Journal of Mathematical Analysis and Applications*. Additionally, I presented our results at the Young Mathematician's Conference at the Ohio State University. This experience provided me with the mathematical rigor necessary to pursue research in cryptography and provable security.

In recognition of these achievements – which include **8 peer-reviewed scholarly works (on 4 of which I am first author)**, **4 conference presentations**, and the creation of valuable open-source tools and datasets – I was awarded the **2025 President's Award for Undergraduate Research and Creative Accomplishments** (given to 3 out of over 10,000 students annually) and the **2025 Academic Excellence Award from the Honors Program** (given to one outstanding student from the graduating Honors cohort). Additionally, I maintained high academic standards and graduated with the **Board of Trustee's Award**, which recognizes that I maintained a perfect 4.0 throughout my undergraduate degree. I was also awarded the **Richtmeyer-Foust Award** for my achievements as a mathematics major.

Future Goals

My path from research in ML-supported fishery monitoring to AI security has given me a unique perspective to pursue creative, interdisciplinary solutions to pressing challenges in AI, and I am excited to continue to work on these research opportunities as a PhD student. My initiative at Central Michigan University to move beyond coursework and actively conduct research prepared me for the rigor of the University of Wisconsin-Madison, and I am very excited to be at the forefront of research in computer science at this university.

Short-Term Research (First Three Years): Fundamentally, I chose to pursue a PhD in Computer Science at UW-Madison to conduct cutting-edge research on provably secure and private AI systems. **My research objective is to develop new ML architectures that will address security vulnerabilities and privacy flaws without reducing system effectiveness.** I have already begun research with Dr. Grigoris Chrysos on leveraging polynomial network architectures to support fully homomorphic encryption (FHE) in ML systems. FHE allows computations on encrypted data, enabling AI systems to process confidential information (e.g., medical records, financial data) without ever decrypting it, providing a provable guarantee of data privacy and security. The NSF GRFP will help me advance the state of AI privacy and security by providing the support to focus on this research, allowing me to develop the necessary libraries and tools to make FHE implementations in existing ML services simple, inexpensive, and transparent.

Long-term Research and Career Goals: My long-term research goal is to investigate how agentic AI systems can be made provably secure using privacy-enhancing technologies like FHE and differential privacy. Additionally, my ultimate career objective is to work as a security research scientist in industry, as I want to be on the inside of cutting-edge AI developments and ensure that the tremendous promise of this technology is realized safely and responsibly. The NSF GRFP will help me meet these goals by accelerating my research and providing a network of other fellows with whom I can collaborate.

Broader Impacts

My proposed research in AI privacy and security directly addresses a foundational challenge to consumer trust in AI. To maximize the broader impacts of my research, I will achieve the following goals as a researcher supported by the NSF GRFP.

Tools and Dissemination: I will design tools, libraries, and architectures that make it simple and inexpensive to implement safe and private AI systems using FHE. Minimizing these obstacles to security and privacy in AI implementations will incentivize companies to build safety into their AI systems from the start. Additionally, I intend to collaborate with industry leaders to ensure that these tools are disseminated beyond the lab and into real-world applications. AI has become the cornerstone of our modern lives, so guaranteeing that AI systems are built with safety and privacy in mind is critical to ensuring that individuals can trust this new foundational technology.

Outreach and Education: As a graduate scholar for the Wisconsin AI Safety Initiative, I will organize and hold technical workshops for the UW-Madison School of Computer, Data & Information Sciences. These seminars will increase the visibility of privacy-enhancing technologies in AI and emphasize how practitioners can use technologies such as FHE to create more secure, private, and safe AI systems. In collaboration with Dr. Chrysos, I will also organize workshops at premier national conferences such as NeurIPS and IEEE S&P that will introduce our work to the broader AI and security research communities. I plan to make my research open source to encourage broad access to this work, as I believe that minimizing the obstructions to free and open research encourages vibrant growth and collaboration.

I was supported and inspired by numerous professors and computer scientists over the course of completing my undergraduate degree, and it is my goal to continue this tradition of inspiring the next generation of STEM researchers. As such, I intend to hold presentations of my research at the UW-Madison Engineering EXPO, a yearly outreach event aimed at encouraging students in Wisconsin to pursue STEM and introducing the public to the scientific advancements of the university. It is my goal to use this event – which draws thousands of visitors per year – to give students an engaging introduction to cryptography, cybersecurity, and ML using hands-on demonstrations and real-world examples.

National Security: My research has direct applications to national security tasks, which require the highest security and confidentiality guarantees. I intend to adapt my research for national security by collaborating with governmental agencies, as I believe that securing and expanding our country's AI infrastructure is key to protecting our cybersecurity as well.

By actively communicating and disseminating the results of my work to student, academic, industry, and government research communities, I will foster greater understanding about the risks posed by unmitigated AI flaws and provide concrete, provably sound solutions.