Earlence Fernandes — Research Statement — Systems Security for Evolving Computers

Computing systems are evolving at a rapid pace, fueled by technology trends that make them intelligent, distributed, and embedded in everyday objects. A grand challenge in computer security is enabling society to gain the benefits of these rapidly evolving technologies without the inherent risks. My research shows that this challenge can be addressed in principled ways using a systems security approach. This involves two steps: (1) applying the security mindset to deeply understand the new ways in which these evolved computer systems are vulnerable; and (2) constructing principled systems-level defenses that offer end-to-end guarantees. This approach served our community well during the first wave of evolution when computers became efficient, portable, and inter-connected. I believe these principles will serve us well during this second wave of evolution, where computers are becoming intelligent and embedded into our daily lives. To this end, my research:

(1) Identifies important threats in new technologies before they become widespread. This is the essence of the systems security approach — determine an appropriate threat model for new technologies before they become ubiquitous so that we can fix their issues by design. In the area of computers becoming ingrained in our daily lives, we were the first to identify threats in Internet-of-Things systems where we demonstrated large-scale device-agnostic attacks (S&P'16); we showed that wireless control systems in bicycles used in professional sporting events such as the Tour de France can be hijacked (WOOT'24), leading to a new form of wireless doping that threatens the integrity of sport. In the area of computers becoming intelligent, we demonstrated physical world attacks on self-driving car AI (CVPR'18, '21); we showed that Large Language Models are susceptible to a range of inference-time attacks that force them to pursue malicious tasks (S&P'25). For communication platforms that became essential during the pandemic, we disclosed vulnerabilities in Slack and Microsoft Teams (USENIX Security'22). We also showed that on-device scanning for illicit images inevitably undermines end-to-end encryption (NDSS'24). This entire line of work contributes to the science of security by determining what are appropriate threat models in evolving computing systems and it has been recognized with bug bounties, a best paper award, has been cited as a motivating factor for the multi-million DARPA GARD and SABER programs, has improved the software security standard in bicycle components in all professional racing, and has impacted the general public through widespread press coverage and an exhibit at the Science Museum in London, where I was quoted.

(2) Builds defenses with end-to-end guarantees. Based on my threat modeling work above, I build real systems that address these threats end-to-end. A long-running line of work here focuses on the fundamental problem of sharing access to data and devices on the Internet. This problem is at the root of security issues in diverse areas ranging from the Internet-of-Things to emerging machine learning-based systems. Specifically, I investigate how to enforce the principle of least privilege in a way that strikes a reasonable balance between functionality, performance and security. We've built several systems that use techniques from applied cryptography (S&P'21), programming languages (USENIX Security'22, '24) and trusted hardware (NDSS'24) to explore different notions of least privilege while keeping the systems practical. This line of work has received a best paper award, an NSF CAREER award, a Google Research Scholar Award, an Amazon Research award, has led to U.S. patents and has impacted product design at Samsung.

Future Efforts. In line with this vision, I believe a fundamental change in computing will be AI agents that reason, plan and interact with physical and digital environments through observation and tool-based manipulation. Current efforts to secure these systems are inadequate because of their narrow focus on models in isolation. My effort is twofold: (1) building strong optimization attacks that show how current ML-based defenses are inadequate; (2) building isolation and access control systems with formal guarantees.

Broader Impact and Community Building. I believe it is important to facilitate broader impact that goes beyond writing papers. For example, my work on algorithmic prompt injections on Mistral's LeChat agent and Google's Gemini fine-tuning system led to fixes by the vendors. On the policy front, I have advised a U.S. member of congress on IoT security issues and have briefed NASEM, FTC and the JASON defense advisory group on AI security issues. I also advise the Union Cycliste Internationale (the administrative organization for professional cycling) to combat technology-enabled fraud in sport. From a community perspective, I recently (May 2025) co-organized the SAGAI workshop that involved researchers from frontier AI labs. The agenda was to discuss the systems approach to securing agentic AI. We will be releasing a report shortly that lists broad approaches and challenges in this space. From a public education perspective, much of my work has been covered in the media (e.g., Wired, Forbes, Science and Nature periodicals). Finally, PhD students from my lab have gone on to top positions in industry and AI research labs and take with them the systems approach to securing evolving computers.