Artificial intelligence (AI) is rapidly transforming cybersecurity, enabling both defenders and adversaries to automate, adapt, and escalate their capabilities. On one hand, AI enhances threat detection through real-time monitoring, behavioral analytics, and predictive modeling. On the other, it empowers malicious actors to develop stealthy and scalable attacks—most notably within Ransomware-as-a-Service (RaaS) ecosystems—by automating reconnaissance, payload delivery, and evasion tactics. This dual-use dynamic has created a volatile cybersecurity landscape, especially in high-risk environments like healthcare, where digital infrastructures are both mission-critical and vulnerable.

This thesis portfolio addresses the general problem: **How can institutions respond effectively to rapidly evolving AI-powered cybersecurity threats in environments where both attackers and defenders use AI?** This overarching question connects two distinct projects: a technical investigation into AI-driven ransomware detection systems, and an STS analysis of hospital cybersecurity responses to AI-based threats. Together, these projects highlight both the technological possibilities and sociotechnical complexities of defending against intelligent, adaptive cyberattacks.

Technical Project Summary: An AI-Based Framework for Ransomware Detection

The technical report explores the design of a machine learning-based ransomware detection system built to address the shortcomings of traditional, signature-based security tools. The rise of AI-enabled ransomware—capable of evading static detection mechanisms through polymorphism, fileless execution, and behavioral mimicry—requires more dynamic and adaptive defenses.

The proposed system features three main components. First, it collects and preprocesses data from system logs, file activity, and network behavior to extract key indicators of compromise. Second, it uses both supervised and unsupervised machine learning models (e.g., Random Forests, SVMs, and Isolation Forests) to detect known patterns and identify anomalies that may indicate emerging ransomware variants. Finally, it incorporates real-time mitigation protocols, automatically isolating affected endpoints, halting malicious activity, and alerting administrators.

While this design has not yet been implemented in a real-world environment, it draws on established research and demonstrates potential to improve detection speed, reduce false positives, and adapt to new ransomware behaviors using reinforcement learning. Challenges include data quality, system scalability, and the risk of adversarial attacks, but the model provides a promising foundation for next-generation cyber defense.

STS Research Summary: Hospital Security in the Era of AI Arms Races

The STS research paper investigates how hospitals are adapting their cybersecurity practices amid the dual-use proliferation of AI. Grounded in Mutual Shaping Theory and Actor-Network Theory (ANT), the study conceptualizes hospital cybersecurity as a sociotechnical process—shaped by institutional policies, regulatory frameworks, workforce capacity, and adversarial innovation.

The research asks: **How are advancements in AI reshaping hospital approaches to security in a landscape where both attackers and institutions use AI?** Through thematic analysis of scholarly literature, government policy, and two case studies—the 2018 SingHealth breach and a 2024 AI-powered ransomware attack on an Indian healthcare provider—the study identifies several key trends.

Hospitals increasingly rely on AI for endpoint detection, anomaly monitoring, and risk management. However, these tools introduce new vulnerabilities, including data poisoning, adversarial learning, and model manipulation. The findings suggest that attackers now test and exploit AI defenses just as institutions deploy them, accelerating a reciprocal arms race.

In response, hospitals are developing layered governance models that integrate technical defense with regulatory compliance and ethical oversight. One example is the adoption of the NIST AI Risk Management Framework, which emphasizes transparency, adaptability, and continuous reassessment. These frameworks reflect a broader shift away from purely technical solutions toward integrated sociotechnical strategies.

Conclusion: Toward Adaptive and Accountable Cybersecurity

Together, these projects underscore the need for adaptive, interdisciplinary approaches to AI-driven cybersecurity. The technical report outlines how machine learning can detect and mitigate ransomware more effectively than legacy systems, while the STS research highlights the institutional and ethical dimensions that shape how such technologies are deployed and governed.

AI-based threats are not static—they evolve in tandem with the defenses built to stop them. As such, institutions must anticipate, not just react to, emerging vulnerabilities. Sustainable cybersecurity requires tools that evolve in real time, governance frameworks that support flexibility and accountability, and a holistic understanding of the interplay between technology, institutions, and adversaries.

By examining both the technical and sociotechnical aspects of AI in cybersecurity, this portfolio contributes to the ongoing effort to secure digital infrastructures in a world where the boundaries between human oversight and machine autonomy are continually shifting.