

The Great AI Defense Paradox: Why Stronger Digital Walls Are Triggering Faster Wars

Ezra Matiwas Wesenie
ezramatiwas1@gmail.com

1. Introduction

The Problem: Cyber ops historically stayed in a quiet "grey zone" of human-speed espionage.
Significance: Low-level hacks didn't cause escalation because human diplomats had time to negotiate.
Research Question: Do machine-speed AI defenses bring peace, or do they trigger preemptive aggression? : Modern advancements in artificial intelligence have completely upended this dynamic. AI-driven cyber defense systems can now monitor networks, analyze threat signatures, patch software vulnerabilities, and neutralize intrusions in mere milliseconds, effectively closing the historical "Automation Gap".

2. Methods

The research looks at how automated hackers use advanced search techniques (like computer chess algorithms) to figure out how to sneak into complex networks.
This study constructs its foundation by blending advanced international relations theory with computer science. It utilizes Security Dilemma 2.0 (analyzing how inherently defensive technologies can mistakenly appear offensive), the Deception-Detection Dichotomy (examining why defense holds structural advantages over offense), and the Subversive Trilemma.
Frameworks: Security Dilemma 2.0, Deception-Detection Dichotomy, and Subversive Trilemma.
Modeling: Evaluates Monte Carlo Tree Search (MCTS) and LLM-driven planning agents.
Cases: Compares U.S. Persistent Engagement (Epic Fury) vs. Singapore Digital Fortress (Cyber Guardian).

3. Results

Defense Advantage: AI excels at real-time threat detection; offense remains brittle and restricted. Defense Has the Upper Hand: AI is amazing at spotting patterns and locking down networks, which makes defending computers much easier than attacking them. Offensive tools, on the other hand, are messy and prone to mistakes.
The Fait Accompli Trap: Slow subversion is instantly blocked; revisionist states resort to high-stakes first-strikes.
Escalation Risk: Compressed kill-chains risk dangerous overlaps with nuclear command systems. Because automated firewalls instantly block slow, sneaky spying, cheating countries realize their normal tricks don't work anymore. If they want to achieve their goals, they feel forced to launch a massive, sudden "first strike" before the AI defense blocks them.

4. Discussion

Strategic Backfire: Unbreakable digital borders create use-it-or-lose-it panic before diplomacy can work. Making your digital borders impossible to break sounds good, but it actually causes a panic. Rival countries think, "If we don't attack now while we still can, we never will." This pushes leaders into rushed military reactions before diplomats can stop them.
Sovereignty: Smaller nations face coercion by superpower AI blocs amidst an enforcement gap. Who Gets Left Behind: Smaller or developing countries face a terrible choice: stay vulnerable to hackers, or hand over their digital safety to superpower "AI tech blocs." Plus, international laws don't know how to handle rogue autonomous code yet.

5. Conclusion

The Takeaway: Super-fast AI defense accidentally ruins the "thinking time" that kept the world peaceful during the old days of hacking. It turns cyberspace from a safe playground into a hair-trigger battleground.
How We Fix It: To stop a digital disaster, world powers need to set up instant computer hotlines, enforce strict human-controlled rules for AI, and help developing nations build their own safe AI defenses.
Why It Matters for the Future: As AI continues to evolve faster than human lawmakers can keep up, the choices we make today about automated warfare will decide whether the future of technology remains safe or spirals out of control.

6. References

Maschmeyer, L. (2025/26). *International Security*.
Maschmeyer, L. (2021). *International Security*.
Harknett, R. J. (2020). *TNSR*.
Johnson, J. (2020). *The Washington Quarterly*.