

How does AI-driven ransomware at the state level reshape national security accountability and long-term geopolitical stability by altering the cost-benefit calculus of asymmetric warfare?

September 25, 2026 | SnugLab Research

Executive Summary

AI-driven ransomware at the state level fundamentally reshapes national security accountability and long-term geopolitical stability by dramatically altering the cost-benefit calculus of asymmetric warfare. It lowers barriers to entry for hostile state actors, enabling them to inflict significant economic damage on critical infrastructure with enhanced plausible deniability, thereby straining traditional attribution and deterrence frameworks [2, 4, 8, 9]. This shift incentivizes prolonged economic attrition over decisive military victories and accelerates conflict tempo, increasing the risk of algorithmic miscalculation and eroding diplomatic trust through AI-generated disinformation, which collectively destabilizes global security [2, 7, 16].

Key Findings

AI Lowers Barriers and Incentivizes Economic Attrition

AI-driven ransomware significantly reduces the financial and technical barriers for minor powers and non-state actors to deploy sophisticated cyber exploitation tools, making critical infrastructure attacks more cost-effective and sustainable than kinetic strikes [6, 8, 11, 12]. AI automates multi-stage attacks, compressing execution timelines from days to mere seconds, and enables the development of polymorphic malware that dynamically mutates to evade detection, lowering the long-term costs of sustained campaigns [9, 10]. This empowers smaller countries, such as North Korea and Iran, to target the economic interests of major powers for political ends with minimal resources [8].

This technological shift incentivizes a move toward prolonged economic attrition in hybrid conflicts. By masking political intent as profit-driven crime, state actors can sustain low-level aggression and asymmetric coercion without the high costs and immediate escalation risks associated with traditional military actions [4, 8, 14]. For example, North

Korea uses ransomware and cryptocurrency hacks to fund its regime and bypass sanctions [14, 17]. Chinese state-linked groups have deployed ransomware to obscure espionage activities, such as the Winnti (APT41) group using ColdLock ransomware against Taiwan's energy and semiconductor sectors to hide intelligence gathering [14, 17]. Russia has outsourced cyber-espionage to criminal groups targeting Ukraine and used cyberattacks to disrupt Ukrainian energy and water services [14]. Iranian-linked actors have disguised destructive wiper attacks as ransomware incidents to mask intent and avoid kinetic escalation [17]. These incidents demonstrate a measurable shift in the cost-benefit calculus, favoring deniable cyber disruption over overt kinetic strikes [5, 15].

Erosion of National Security Accountability Through Attribution Challenges

The operational capacity of machine learning systems is fundamentally strained by AI-generated malware, which dynamically mutates and weakens signature-based detection and static behavioral models [9]. This significantly increases the difficulty in attributing AI-driven attacks, blurring the lines between state-sponsored and financially motivated cybercrime and preventing multinationally coordinated punitive sanctions [2, 14]. For instance, Large Language Models generate Tactics, Techniques, and Procedures (TTPs) datasets with a low average Jaccard similarity of 0.39 compared to human-generated data, complicating behavioral attribution [19]. Autonomous AI agents further strain foundational assumptions about digital identity, as their actions do not map cleanly onto individual human operators [20].

This erosion of accountability is evident in specific incidents where international justice mechanisms have been targeted. The International Criminal Court (ICC) disclosed a cyber intrusion targeting its sensitive information systems in September 2023 [18]. Russian cyber operations have also targeted bodies investigating war crimes in Ukraine to disrupt investigations or discredit findings [18]. Generative AI further degrades these protocols by using deepfake technology and synthetic media to undermine the integrity of digital evidence and public trust, which can fabricate witness testimonies or implicate innocent individuals in legal proceedings [1, 7]. Operationally, AI bypasses established response frameworks through unprecedented speed, with the fastest confirmed intrusion breakout time falling to 27 seconds, rendering traditional incident response playbooks untenable [21].

Accelerated Conflict Tempo and Eroded Diplomatic Trust Jeopardize

Geopolitical Stability

Both the acceleration of conflict tempo and the qualitative erosion of diplomatic trust through AI-generated disinformation are primary, interconnected risks to long-term geopolitical stability [2, 4, 6, 7, 16]. AI-enabled algorithmic systems can execute tasks in seconds that traditionally take hours, pushing the pace of cyber attacks and responses beyond human cognitive capabilities [7]. This rapid execution increases the likelihood of algorithmic miscalculation, where flawed AI systems generate errors in intelligence or misinterpret operational contexts, creating pathways for unintended escalation [7, 16].

Concurrently, AI-driven disinformation qualitatively erodes the shared facts and trust necessary for effective signaling. State actors exploit AI for psychological warfare, deploying highly targeted misinformation and deepfake campaigns to fabricate false diplomatic crises, provoke international conflicts, and manipulate political discourse [2, 4, 14]. For example, Iranian-linked actors deployed generative AI to amplify leaked information following a hack-and-leak campaign targeting journalists in July 2025 [15]. Russian AI-driven bot attacks contributed to the annulment of Romania's 2024 presidential election [2]. By undermining democratic institutions and sowing distrust, these campaigns destabilize crisis management, making it difficult for adversaries to distinguish between genuine strategic threats and AI-generated provocations [4]. The combination of compressed decision windows and degraded information environments compounds the risk of miscalculation, threatening long-term geopolitical stability [2, 7].

Deterrence Models Are Strained, Necessitating Norm Revision

Historical precedents of successful cyber deterrence are strained by AI-driven ransomware, which lowers entry thresholds and increases attack variance, suggesting a need for revised international norms rather than a complete invalidation of existing models [6, 8, 11]. AI acts as a force multiplier, dramatically lowering barriers to entry for hostile nation-states and non-state actors, allowing them to inflict significant economic damage with relative ease and deniability [2, 4, 5, 9, 10]. While AI complicates attribution, making it harder to discern the origin and intent of attacks [2], some evidence suggests that machine learning and data fusion can improve detection and attribution, potentially increasing the credibility of deterrence [3]. State-sponsored actors remain more susceptible to deterrence due to the potential repercussions on the nation-state [13]. Therefore, while deterrence is still possible through improved forensic capabilities and targeted responses, the lower thresholds and higher variance of AI attacks strain these models, necessitating a fundamental revision of international norms regarding critical

infrastructure protection [2, 4, 6, 7, 16].

Implications

The reshaping of national security accountability and geopolitical stability by AI-driven ransomware has several critical implications. The reduced cost and increased deniability of cyberattacks mean that states, particularly those with limited conventional military power, will increasingly favor economic attrition in hybrid conflicts, leading to prolonged periods of low-level aggression rather than decisive engagements. This shift demands a re-evaluation of national defense strategies to prioritize cyber resilience and proactive defense over traditional deterrence based on kinetic retaliation. The accelerated tempo of AI-enabled attacks, combined with the erosion of trust from AI-generated disinformation, necessitates urgent development of robust international frameworks for cyber attribution and crisis management that can operate at machine speed while preserving human oversight and diplomatic channels. Without clearer international norms and enhanced public-private collaboration for critical infrastructure protection, the risk of miscalculation and unintended escalation in the cyber domain will continue to grow, posing a significant threat to global stability.

Limitations and Caveats

The available research provides a strong qualitative understanding of how AI-driven ransomware alters the cost-benefit calculus and impacts national security. However, direct quantitative comparisons between the cost-benefit ratios of state-level AI ransomware campaigns and traditional kinetic strikes for specific strategic objectives (e.g., disrupting energy grids vs. data exfiltration) are limited. The research lacks specific deployment cost figures for ransomware, exact duration metrics for kinetic alternatives, and a numerical analysis of their respective political fallout. Furthermore, the provided context does not contain data on the specific technical capabilities, current false-positive rates, or detection latency metrics of leading commercial AI cybersecurity platforms (e.g., CrowdStrike, SentinelOne, Microsoft Defender for Endpoint) in detecting polymorphic AI ransomware. This limits the ability to assess the current defensive posture against these evolving threats with precise metrics.

Sources

- [1] [gov] Ai Apt Campaigns And Urgent Threats To Critical Infrastructu - cyber.nj.gov - <https://www.cyber.nj.gov/threat-landscape/nation-state-threat-analysis-reports/ai-apt-campaigns-and-urgent-threats-to-critical-infrastructure>
- [2] [gov] Acpd Official Meeting Minutes May 6 2026 - state.gov - <https://www.state.gov/acpd-official-meeting-minutes-may-6-2026>
- [3] [edu] Deterrence Through Ai Enabled Detection And Attribution - kissinger.sais.jhu.edu - <https://kissinger.sais.jhu.edu/programs-and-projects/kissinger-center-papers/deterrence-through-ai-enabled-detection-and-attribution/>
- [4] [edu] Weaponized Ai A New Era Of Threats - ash.harvard.edu - <https://ash.harvard.edu/articles/weaponized-ai-a-new-era-of-threats/>
- [5] [edu] Ai Driven Cybersecurity Trends 2025 - captechu.edu - <https://www.captechu.edu/blog/ai-driven-cybersecurity-trends-2025>
- [6] Artificial Intelligence Terrorism And International Relation - cirsd.org - <https://cirsd.org/opinions/artificial-intelligence-terrorism-and-international-relations/>
- [7] UNIDIR AI International Security Understanding Risks Paving - unidir.org - https://unidir.org/wp-content/uploads/2023/10/UNIDIR_AI-international-security_understanding_risks_paving_the_path_for_confidence_building_measures.pdf
- [8] Cybersecurity Era Asymmetric Economic Warfare - justsecurity.org - <https://www.justsecurity.org/19150/cybersecurity-era-asymmetric-economic-warfare/>
- [9] [blog] How Ai Is Changing Ransomware And Why Its Faster Smarter And - morphisec.com - <https://www.morphisec.com/blog/how-ai-is-changing-ransomware-and-why-its-faster-smarter-and-harder-to-detect/>
- [10] The Rise Of Ai Ransomware How Artificial Intelligence Is Tra - mcgowanprofessional.com - <https://mcgowanprofessional.com/the-rise-of-ai-ransomware-how-artificial-intelligence-is-transforming-cyber-crime/>
- [11] [peer-reviewed] tandfonline.com - AUTHORS UNAVAILABLE - <https://www.tandfonline.com/doi/abs/10.1080/14751798.2019.1600800?needAccess=true>
- [12] Asymmetric Ai Warfare In The Era Of Cheap Intelligence - federalnewsnetwork.com - <https://federalnewsnetwork.com/commentary/2025/03/asymmetric-ai-warfare-in-the-era-of-cheap-intelligence/>
- [13] [peer-reviewed] 396161327 Cyber Conflict And International Relations A Compr - researchgate.net - AUTHORS UNAVAILABLE - https://www.researchgate.net/publication/396161327_Cyber_Conflict_and_International_Relations_A_Comprehensive_Analysis_of_Cyber_Deterrence_Strategies_in_Contemporary_Geopolitics
- [14] [blog] Top Nation State Cyber Attack - infosecurityeurope.com - <https://www.infosecurityeurope.com/en-gb/blog/threat-vectors/top-nation-state-cyber-attack.html>
- [15] 2025 Cyber Incident Trends What Businesses Must Know - exclusive-networks.com - <https://www.exclusive-networks.com/se/resources/knowledge-base/articles/2025-cyber-incident-trends-what-businesses-must-know>
- [16] Strategic Analysis Of Cyber Conflicts A Game Theoretic Model - securityanddefence.pl - <https://securityanddefence.pl/Strategic-analysis-of-cyber-conflicts-A-game-theoretic-modelling-of-global-cyber,203979,0,2.html>
- [17] [peer-reviewed] Capa.70063 - onlinelibrary.wiley.com - AUTHORS UNAVAILABLE - <https://onlinelibrary.wiley.com/doi/10.1111/capa.70063>
- [18] [peer-reviewed] Retaliating against cyber-attacks: a decision-taking framework for policy-makers and enforcers of international and cybersecurity law - Authors: Leventopoulos, Sozon; Pipyros, Kosmas; Gritzalis, Dimitris - Journal: International Cybersecurity Law Review - <https://link.springer.com/article/10.1365/s43439-024-00113-5>
- [19] Impact Of Ai On Cyber Threat - ncsc.gov.uk - <https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat>
- [20] [peer-reviewed] dl.acm.org - AUTHORS UNAVAILABLE - <https://dl.acm.org/doi/10.1016/j.cose.2025.104606>
- [21] Ai Agents Agency In The Internet Ecosystem - securityandtechnology.org -

<https://securityandtechnology.org/virtual-library/white-paper/ai-agents-agency-in-the-internet-ecosystem/>