December 2025

ATARC

WHITE PAPER

Cyber–AI Convergence: From Experimentation to Operational Reality

A Chatham House Rule Synthesis of a Multi-Sector Roundtable hosted by the Advanced Technology Academic Research Center (ATARC)

powered by GovExec

Executive Summary:

Cybersecurity and artificial intelligence are no longer separate or loosely connected disciplines. They are converging operationally, organizationally, and strategically in ways that fundamentally reshape how digital systems are defended. This convergence is driven by the accelerating speed and scale of cyber threats, the growing automation of adversary behavior, and the increasing reliance on AI-enabled systems across critical missions.

Participants in a closed-door roundtable spanning government, industry, and research communities broadly agreed on a central reality: machine-speed threats increasingly require machine-assisted defense. At the same time, the rapid adoption of AI introduces new attack surfaces, governance challenges, and systemic risks that are not yet fully understood or consistently managed.

This white paper synthesizes those discussions without attribution, in accordance with Chatham House rule. It captures areas of alignment, points of tension, and practical implications rather than individual viewpoints. A consistent theme emerged throughout the discussion: successful cyber–AI integration will depend less on the sophistication of individual models and more on governance, data discipline, operational integration, and sustained human judgment.

1. Why Cyber and AI Are Converging Now

Cybersecurity has always been an adversarial domain, but participants emphasized that the pace of modern threats represents a qualitative shift. Adversaries increasingly automate reconnaissance, exploitation, credential harvesting, and lateral movement. These activities occur at machine speed and at scales that overwhelm human-centric defensive teams.

At the same time, organizations face chronic shortages of skilled cyber personnel and persistent alert fatigue. Manual processes and human-only analysis cannot keep up with the volume and velocity of modern cyber activity.

Artificial intelligence has therefore moved from an experimental capability to an operational necessity. AI enables pattern recognition across massive telemetry streams, prioritization of alerts, automated response actions, and simulation of attacker behavior. It also introduces new efficiencies in testing, red-teaming, and system monitoring.

Convergence is occurring in both directions. AI is being embedded into cyber defense operations, and cybersecurity organizations are now responsible for protecting AI systems themselves. Participants agreed that this convergence is already underway and will only accelerate.

2. AI for Cybersecurity: Opportunities and Constraints

Participants identified clear value in existing AI-enabled cyber use cases. These include anomaly detection across logs and networks, alert triage and prioritization, automated containment of known threats, red-teaming and vulnerability discovery, and augmentation of security operations centers.

These applications are already improving efficiency and enabling organizations to scale limited human resources. In many cases, AI is reducing mean time to detect and respond, particularly for well-understood threat patterns.

However, participants consistently cautioned against inflated expectations. AI does not eliminate cyber risk; it reshapes it. Poor data quality, narrow training sets, limited explainability, and poorly bounded automation can introduce new failure modes. Several participants noted that the most significant risks arise when AI is deployed without sufficient visibility or oversight.

A recurring observation was that AI performs best when applied to clearly defined problems, supported by high-quality data, and embedded in workflows that preserve human decision authority.

3. Securing AI Systems as a Cyber Mission

As AI becomes more deeply integrated into operational environments, participants emphasized that securing AI systems must be treated as a first-order cyber mission. Risks discussed included data poisoning during training, prompt injection and manipulation, model inversion, abuse of agentic workflows, and vulnerabilities introduced through third-party models and APIs.

Traditional perimeter-based controls and static authorization processes were widely viewed as insufficient for adaptive systems that learn and evolve over time. Participants stressed that securing AI requires continuous monitoring, strong identity and access controls, visibility into data provenance, and explicit constraints on model behavior and autonomy.

Several participants emphasized that secure AI cannot be an afterthought. It must be designed, governed, and operated as foundational infrastructure rather than layered on after deployment.

4. Agentic AI: Promise, Risk, and Control

Agentic AI systems, capable of planning and acting with limited human intervention, generated some of the most substantive discussion. Participants recognized their potential to transform incident response, continuous testing, policy enforcement, and operational coordination.

At the same time, these systems raise significant concerns. Risks include unintended actions at machine speed, privilege escalation, cascading failures, and erosion of human situational awareness. Participants noted that once agentic systems are embedded deeply into workflows, reversing or constraining them becomes difficult.

There was strong alignment that autonomy must be earned incrementally. Agentic systems should operate within clearly defined authority boundaries, include robust logging and auditability, and maintain clear escalation paths to humans. Trust, participants emphasized, must be built through evidence and experience rather than assumed by design.

5. Data as the Central Battleground

Across nearly every topic, data emerged as the decisive factor in cyber–AI convergence. Participants described persistent challenges with fragmented data sources, inconsistent labeling, unclear lineage, and over-collection without defined purpose.

Several participants challenged the assumption that larger models and broader data ingestion are always beneficial. In many cyber contexts, smaller, domain-specific models trained on well-curated datasets were viewed as more effective, more explainable, and easier to govern.

Intentional data strategy, focused on quality, relevance, provenance, and lifecycle management, was widely seen as a prerequisite for trustworthy AI-enabled cyber operations.

6. Governance, Risk, and Assurance

Participants expressed concern that existing governance models have not kept pace with adaptive technologies. Static, compliance-driven frameworks struggle to address systems that learn, change, and interact dynamically.

There was broad support for moving toward risk-based governance approaches emphasizing continuous evaluation, outcome-focused assurance, and real-time visibility into system behavior. Governance was framed not as a barrier to innovation but as an enabler of trust and scale.

Several participants noted that without credible assurance mechanisms, organizations face a false choice between over-restricting AI and deploying it unsafely.

7. Workforce and Organizational Implications

The convergence of cyber and AI does not require universal expertise in both domains. Instead, participants emphasized interdisciplinary teams where skills complement one another.

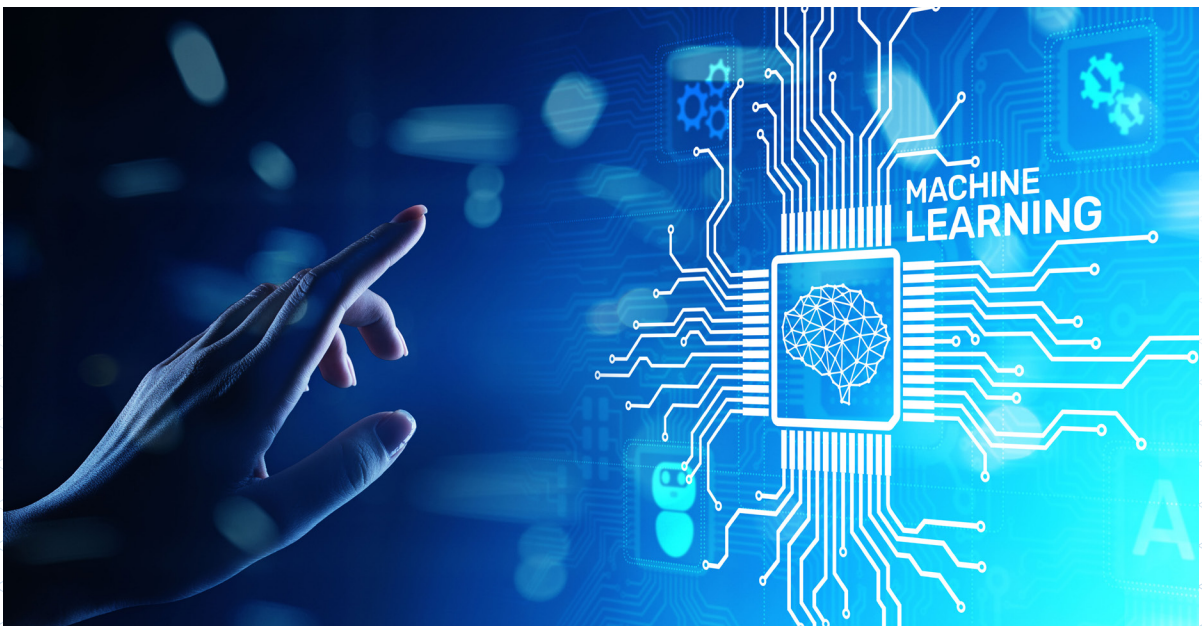
Cyber professionals need AI literacy, while AI practitioners need operational and security context. Human judgment remains essential for ambiguity, ethical considerations, escalation decisions, and accountability. AI was consistently described as a force multiplier for people, not a replacement.

Participants also highlighted the importance of organizational structures that support collaboration rather than siloed ownership of AI or cyber capabilities.

8. Measurement, Standards, and Evaluation

Participants identified a lack of shared metrics for evaluating AI trustworthiness, resilience, and performance in cyber contexts. Existing benchmarks often fail to reflect adversarial conditions or operational realities.

There was strong interest in continuous testing, red-teaming, and operationally relevant evaluation frameworks that move beyond static, lab-only assessments. Participants emphasized that measurement is essential not only for performance, but for trust and governance.



9. Near-Term Opportunities and Practical Steps

Despite the challenges, participants identified practical steps organizations can take now. These include deploying AI narrowly in high-confidence cyber use cases, improving data hygiene and visibility, piloting constrained agentic systems, modernizing governance toward continuous assurance, and establishing cross-functional cyber–AI working groups.

Progress was consistently framed as iterative and evidence-driven rather than transformational.

10. Strategic Recommendations

From the discussion, several strategic imperatives emerged:
• **Organizations should treat cyber–AI convergence as an operational reality rather than a future concern.**
• **Governance, data discipline, and integration should be prioritized before scaling autonomy.**
• **AI should be deployed incrementally with clear constraints and oversight.**
• **AI systems should be secured as critical infrastructure. Workforce models should emphasize interdisciplinary collaboration.**
• **Governance should evolve from static compliance to continuous risk management.**

11. From Kickoff to Sustained Collaboration

The roundtable was intended as a kickoff rather than a one-time event. Participants expressed strong interest in continued engagement through focused working groups exploring specific challenges in greater depth.

Suggested areas include securing AI systems, operational AI-enabled cyber defense, governance of agentic systems, data strategy and trust, continuous authorization models, measurement and evaluation frameworks, workforce design, interoperability challenges, policy and legal considerations, and controlled experimentation environments.

The upcoming ATARC Cyber-AI Convergence Working Group will provide a pathway from shared dialogue to sustained collaboration and practical outcomes.

12. Conclusion

Cyber–AI convergence represents one of the most consequential shifts in how organizations defend digital systems. The opportunity is substantial, but so are the risks of haste, hype, and insufficient governance.

Participants agreed that technology alone will not determine success. Disciplined integration, high-quality data, thoughtful governance, and sustained human judgment will ultimately define whether AI strengthens or undermines cybersecurity.

Organizations that approach AI as a powerful but constrained partner, capable of speed and scale, but embedded within clear boundaries and accountability, will be best positioned to navigate this transition securely and responsibly.