Business Software Alliance

Cybersecurity Professionals First

Launching a Voluntary Public-Private Partnership to Execute a Strategic, Phased Rollout of Frontier AI Cybersecurity Systems Will Make the Digital Ecosystem Stronger and More Resilient

Frontier artificial intelligence (AI) cybersecurity systems create significant opportunities to improve vulnerability discovery and remediation. They also create new risks if their capabilities become widely available before defenders can put them to use.

To maximize the benefits and manage the risks, government, industry, and the cybersecurity community should establish a voluntary public-private partnership to support a strategic, phased rollout of these systems, ensuring trusted cybersecurity professionals receive timely access while promoting broader deployment through security platforms. By coordinating these activities through a structured, transparent, globally aligned, and predictable framework, stakeholders with proven records of stewardship, capability, and impact can strengthen the security and resilience of the broader digital ecosystem.

Frontier AI Cybersecurity Systems Create Opportunities and Risks

Recent advances in frontier AI cybersecurity systems, like GPT-6 Astra and Mythos, are rapidly transforming the cybersecurity landscape. These systems can autonomously discover and exploit vulnerabilities at a speed and scale that previously would have required teams of highly skilled professionals.

For defenders, these capabilities create a significant opportunity. Frontier AI cybersecurity systems can accelerate vulnerability discovery and remediation, improve software security, strengthen incident response, and help organizations identify weaknesses before attackers exploit them. Over time, widespread adoption of these systems and platforms that use them will substantially improve the security and resilience of the digital ecosystem.

At the same time, these systems create new risks. The same capabilities that enable defenders to discover and fix vulnerabilities can also be used by malicious actors to discover and exploit them. In the near term, one of the most important cybersecurity objectives should be ensuring trusted cybersecurity professionals can leverage these capabilities before attackers do.

To date, access to the most capable frontier AI cybersecurity systems has largely been managed through ad hoc decisions by individual developers. This approach reflects the realities of rapid innovation and has enabled developers to move quickly with confidence. That said, it can be difficult for stakeholders to understand, may not account for all relevant interests, and is unlikely to provide the transparency, predictability, and coordination needed as capabilities continue to advance.

A more structured approach is therefore needed, which entails replacing fragmented, case-by-case decision-making with a voluntary public-private partnership. A voluntary public-private partnership can deliver a structured, transparent, globally aligned, and predictable strategic, phased rollout of these systems to credible stakeholders, resulting in a more secure and resilient digital ecosystem.

Goals of a Voluntary Public-Private Partnership

To harness the benefits of frontier AI cybersecurity systems while managing associated risks, the voluntary public-private partnership should pursue two primary objectives:

1. Ensure that organizations with proven records of stewardship, capability, and impact have timely access to frontier AI cybersecurity systems, so they can discover, prioritize, and remediate vulnerabilities; and
2. Promote broad deployment of frontier AI cybersecurity capabilities through security platforms that organizations already rely upon to defend themselves from malicious actors.

By coordinating these activities through a structured, transparent, globally aligned, and predictable framework, stakeholders with proven records of stewardship, capability, and impact can strengthen the security and resilience of the broader digital ecosystem.

Together, these goals would help ensure that the benefits of frontier AI cybersecurity systems are realized as quickly and broadly as possible while reducing opportunities for malicious actors to exploit vulnerabilities before defenders can address them.

To achieve these objectives, the partnership and its decisions should be:

- Structured, with clearly defined processes, roles, and criteria;
- Transparent, so stakeholders understand how decisions are made;
- Globally aligned, enabling coordination across borders; and
- Predictable, providing clear expectations regarding processes and outcomes.

Activities of the Voluntary Public-Private Partnership

As a preliminary matter, the partnership should determine which frontier AI cybersecurity systems and platforms warrant a phased rollout framework. Most AI cybersecurity systems improve existing workflows without materially changing the cybersecurity landscape and therefore do not require a phased rollout. However, some frontier systems may significantly exceed currently available capabilities by autonomously discovering, prioritizing, or exploiting vulnerabilities at scale. The widespread availability of such systems could create both substantial cybersecurity opportunities and increased systemic risk.

The partnership should therefore identify the systems and platforms that fall within scope based on their capabilities and potential impact. A narrowly tailored approach will help ensure that phased rollout mechanisms are applied only where justified, while preserving innovation and enabling the broad deployment of technologies that strengthen cybersecurity.

Importantly, the partnership should also recognize that frontier status is inherently temporary. As capabilities evolve and equivalent—or near-equivalent—systems become available, a phased rollout may no longer be necessary to maximize cybersecurity outcomes. Accordingly, the partnership should establish a clear mechanism for promptly removing systems from its scope once they no longer present the opportunities and risks that justify a phased rollout. This dynamic approach will ensure the partnership remains focused on truly frontier AI cybersecurity systems while avoiding unnecessary constraints on innovation or delaying the broader deployment of cybersecurity technologies.

By clearly identifying which systems warrant special consideration, the partnership can avoid unnecessary constraints on innovation while focusing its efforts where they will have the greatest impact on cybersecurity and resilience.

For frontier AI cybersecurity systems that warrant a strategic, phased rollout, the partnership should determine whether and when organizations receive access based on a holistic assessment of their stewardship, capability, and potential impact. This assessment should consider:

- Stewardship factors, such as trustworthiness and transparency in ownership or control;
- Capability factors, such as technical competence, organizational capacity, and collaborative capacity; and
- Impact factors, such as mission, systemic reach, and public benefit.

Together, these considerations can help ensure that early access is provided to those who can be trusted to use these systems responsibly, have the capabilities necessary to use them effectively, and are positioned to meaningfully improve cybersecurity and resilience.

Everyone Benefits From a Partnership That Delivers a Strategic, Phased Rollout of Frontier AI Cybersecurity Systems

By moving to a voluntary public-private partnership, stakeholders can ensure that the benefits of these systems are realized broadly while responsibly managing the risks associated with their deployment and helping to ensure that access decisions advance both cybersecurity and public-interest objectives.

By getting these powerful systems in the hands of those best situated to use them, this partnership will:

- Build trust through clear, consistent, and explainable access criteria and decision-making processes;
- Strengthen defensive capabilities by accelerating vulnerability discovery, prioritization, and remediation;
- Reduce systemic risk by enabling vulnerabilities to be identified and addressed before they can be broadly exploited;
- Provide developers with predictable governance frameworks and greater clarity around risk management; and
- Support international alignment among like-minded partners and reduce fragmentation in national approaches.

Frontier AI cybersecurity systems can accelerate vulnerability discovery and remediation, improve software security, strengthen incident response, and help organizations identify weaknesses before attackers exploit them.

Stakeholders Must Work Together to Create This Public-Private Partnership and Improve Cybersecurity

The opportunity is enormous. Effectively deployed, frontier AI cybersecurity systems will significantly improve vulnerability discovery and remediation and ultimately strengthen collective security and resilience. But these outcomes are not guaranteed. It is now incumbent upon government and industry to make sure that the most capable defenders obtain access to the most powerful tools first.

Launching a voluntary public-private partnership will require collaboration between governments, AI developers, and the cybersecurity community. No single stakeholder can manage the risks alone, but together, stakeholders can all realize the benefits of frontier AI cybersecurity systems.

BSA Framework for a Voluntary Public-Private Partnership to Manage a Strategic Phased Rollout of Frontier AI Cybersecurity Systems

We recognize that the partnership determinations cannot be reduced to a purely quantitative formula. They require a holistic, multi-factor assessment that weighs a range of relevant criteria and considers the totality of the circumstances of any given case. Nevertheless, identifying clear criteria for how the partnership operates and for determining whether and when organizations should be provided with access will promote better outcomes and build trust in the system.

Program Design (PD): The voluntary program is administered predictably.

- **Purpose (PD.PU):** The partnership's objectives and priorities for a phased release of frontier AI cybersecurity systems are established and understood.
  - **PD.PU-1:** The partnership should enable a phased release of these systems, including prioritizing access to organizations and sectors where it is likely to produce the greatest benefits for cybersecurity and resilience.
  - **PD.PU-2:** The partnership should ultimately enable access for both organizations improving their own cyber defenses and for organizations improving the defenses of their customers.

- **Globally Aligned (PD.GA):** The partnership is globally aligned to promote consistent access to frontier AI cybersecurity systems and coordinated approaches across jurisdictions.
  - **PD.GA-1:** The partnership should be open to organizations headquartered in like-minded jurisdictions that meet the applicable criteria.
  - **PD.GA-2:** The partnership should promote alignment among participating governments on approaches to the phased release of frontier AI cybersecurity systems.
- **Transparent (PD.TR):** The partnership's processes, criteria, decisions, and timelines are clear and understandable to stakeholders.
  - **PD.TR-1:** The partnership should provide clear information about how organizations can participate and apply for pre-release access to systems.
  - **PD.TR-2:** The partnership should provide a clear explanation of the process and criteria it applies when making decisions.
  - **PD.TR-3:** The partnership should provide a clear explanation for denying access.
  - **PD.TR-4:** The partnership should provide clear information about the expected duration of the pre-release phase for new systems.
- **Accessible (PD.AC):** The partnership enables qualified organizations to benefit from frontier AI cybersecurity systems based on their needs, potential impact, and ability to improve cybersecurity.
  - **PD.AC-1:** The partnership should consider how to remove barriers, including cost, that prevent otherwise qualified organizations from leveraging frontier AI cybersecurity systems.
  - **PD.AC-2:** The partnership should support high-impact cybersecurity uses, including efforts to secure open-source software as well as coordinate and deduplicate vulnerability discovery and remediation.

Stewardship (ST): Organizations receiving access demonstrate they can be trusted to responsibly use frontier AI cybersecurity systems.

- **Trustworthiness (ST.TU):** The organization demonstrates a record of responsible conduct and practices to provide confidence it will use frontier AI cybersecurity systems to improve cyber defenses.
  - **ST.TU-1:** Does the organization have a history of responsible conduct and compliance with applicable laws, particularly those related to data protection, cybersecurity, and AI governance?
  - **ST.TU-2:** Does the organization use best practices and internationally recognized standards, including for vulnerability disclosure and remediation?

- **ST.TU-3:** Does the organization possess relevant internationally recognized certifications?
- **ST.TU-4:** Does the organization have relationships, operations, ownership, or other ties to a country of concern that could create risks associated with its access to frontier AI cybersecurity systems?
- **ST.TU-5:** Does the organization provide a high degree of confidence that it can limit unintended access to these systems and that they will be used for their intended purposes?

» **Transparency (ST.TA):** The organization's ownership, control, and material influence are sufficiently transparent to inform access decisions.

- **ST.TA-1:** Does the organization provide sufficient transparency into its ownership and control for the partnership to identify who owns, controls, or materially influences the organization?

Capability (CA): Organizations receiving access have the technical competence, organizational capacity, and collaborative capacity necessary to effectively use frontier AI cybersecurity systems to improve cyber defenses.

» **Technical Competence (CA-TC):** The organization has the technical expertise and experience necessary to effectively use frontier AI cybersecurity systems.

- **CA.TC-1:** Does the organization have demonstrated expertise and experience using advanced cybersecurity tools and technologies to improve cyber defenses?

» **Organizational Capacity (CA.OC):** The organization has the resources and operational capacity necessary to effectively deploy frontier AI cybersecurity systems.

- **CA-OC-1:** Does the organization have the scale, resources, governance, and operational maturity necessary to use frontier AI cybersecurity systems to meaningfully improve cyber defenses?

» **Collaboration Capacity (CA.CC):** The organization has trusted relationships and demonstrated experience working with relevant stakeholders to coordinate vulnerability disclosure and remediation.

- **CA.CC-1:** Does the organization have demonstrated experience working with software developers, open-source software communities, vulnerability researchers, customers, and other stakeholders to coordinate vulnerability disclosure and remediation across organizational and national boundaries?

Impact (IM): Organizations receiving access are positioned to use frontier AI cybersecurity systems to meaningfully improve cybersecurity and resilience.

- **Mission (IM.MI):** The organization's mission and activities contribute to improving cyber defenses.
  - **IM-MI-1:** Does the organization have a demonstrated track record of contributing to defensive cybersecurity outcomes, including by securing its products, services, systems, or data; providing cybersecurity services; conducting vulnerability research; improving software security; responding to incidents; protecting critical infrastructure; or through other activities?
- **Systemic Reach (IM.SR):** The organization is positioned to use frontier AI cybersecurity systems to improve cybersecurity at significant scale.
  - **IM.SR-1:** Does the organization secure critical infrastructure, widely used technologies, or other systems with significant cybersecurity implications?
  - **IM-SR-2:** Do significant numbers or types of individuals, businesses, critical infrastructure operators, or other organizations rely on the organization's products or services?
- **Public Benefit (IM.PB):** The organization's use of frontier AI cybersecurity systems is likely to generate cybersecurity benefits beyond the organization itself.
  - **IM.PB-1:** Does the organization's position make it likely that its access to these systems would generate broader cybersecurity benefits, including through vulnerability disclosure and remediation, securing open-source software, or otherwise improving the security and resilience of the broader digital ecosystem?