Improving National Resilience Against AI Incidents: A Global Perspective

Executive Summary

Many countries have sought to encourage AI development given its potential for technological innovation, advances in healthcare, and military primacy. Even so, AI governance approaches are struggling to keep pace with the speed of development and access to artificial intelligence technology. AI experts and a number of government leaders continue to warn that the structural dynamics regarding the development and use of AI will likely prompt global security challenges states will find increasingly difficult to address.

This brief argues that national security globally may grow increasingly precarious as a result of an increase in the proliferation of serious, large-scale incidents attributable to AI. Following, it outlines key preparedness efforts states globally should prioritize in order to remain resilient against a broad range of national security threats amidst AI's potency increases and integration into critical parts of our society.

Divided into prevention and response, these recommendations include:

Prevention:

1. **Mandate Legal and Incident Reporting Obligations for AI Developers**
   Establish mandates and mature mechanisms for AI developers and operators to report on serious incidents and coordinate with authorities during emergencies, while protecting whistleblowers and proprietary information.
2. **Strengthen Institutional and Technical Preparedness**
   Integrate AI expertise into national incident planning frameworks and cybersecurity exercises; and ensure collaboration and coordination amongst relevant teams across government, academia, and civil society for rapid response and technical advising.
3. **Secure Critical Infrastructure Against AI-Enabled Threats**
   Conduct routine threat assessments, particularly in energy, healthcare, and nuclear sectors; and invest in the physical and digital protections of these systems.

4. **Enhance Oversight of Dual-Use Research and Biosecurity**
   Create national screening and oversight systems for individuals and labs working with hazardous biological materials and the machines that can make them, with proper access controls and incident tracking.
5. **Increase Public Awareness and Global Coordination**
   Promote public literacy on AI risks (e.g., deepfakes, manipulation), and strengthen international cooperation on threat detection, information sharing, and emergency prevention in multilateral forums.

Response:

1. **Develop AI-Specific Emergency Response Protocols**
   Strengthen national incident management frameworks through the inclusion of prevention and response efforts against large-scale AI incidents.
2. **Invest in Threat Detection and Red Teaming Capabilities**
   Develop and deploy monitoring tools and red-team tactics to scan for threats attributable to AI in critical sectors and domains.
3. **Prepare for Rapid Isolation and Containment of Threats**
   Research and develop, as necessary, containment mechanisms to isolate or shut down compromised AI systems.
4. **Ensure Robust, Inclusive Emergency Communication Systems**
   Harden emergency services against AI-enabled interference, and prioritize outreach and preparedness for vulnerable communities, to include international cooperation on incident warning.
5. **Build Capacity for Remediation During and Recovery After Incidents**
   Develop AI-specific forensic and contingency planning capabilities, such as the maintenance computing/data reserves for government continuity.

Introduction

As artificial intelligence accelerates toward transformative but risky capabilities, it is increasingly important to modernize global emergency response systems to prepare for security threats from AI. From AI-enabled cyberattacks on national infrastructure to runaway autonomous systems, the spectrum of plausible AI-related incidents is widening. The scale and speed with which they could unfold could exacerbate AI threats by overwhelming existing emergency response systems. At the same time, geopolitical

dynamics and intense competition between leading AI companies structurally disincentivise regulation and preventive safety measures, further increasing the need for robust incident preparedness and response. Extant national emergency response frameworks, largely designed to address natural disasters, human conflict, nuclear accidents, and public health emergencies, are neither equipped for the technical complexity of AI threats nor calibrated to the tempo at which these crises could escalate.

This brief seeks to provide context surrounding cognizance of AI's risks and capabilities within national security discourse and makes the case for greater emerging threat preparedness from a comparative lens. The recommendations within the brief were developed in cognizance of the many variables states have to consider in their own emergency preparedness, AI governance, and national security prerogatives. They are not one-size-fits-all proposals, they are guiding notions for layered and adaptive resilience to the security threat that AI is and will increasingly present. We must reimagine emergency response for a world where AI not only enables novel forms of disruption but does so with unprecedented scale, subtlety, and speed.

Context

AI's ability to reason, plan, and take actions, with access to vast amounts of information, allow for greater automation and faster procession of critical tasks with significantly minimal need for human control.[1] AI's computing capacity and versatility enables a substantially wider and deeper range of technological integration into societal systems, physical and nonphysical, than traditional software and Internet of Things devices.[2] However, through the failure of the systems' performance, misuse, or autonomous malevolence, these factors also create a greater risk of injury and/or damage from incidents attributable to AI.[3] The interconnectedness of today's world also means that any sufficiently cognizant AI can affect an extremely wide range of commons and third-party systems. An *AI incident* refers to the circumstances, events, or series of events that, in the use, development, or malfunction of AI, directly or indirectly lead to harm, injury or disruption to people, systems, property, and the environment.[4] In the absence of sufficient risk-based safety and security measures in place throughout AI's development and use, AI can act as a target for attacks, as a vector to weaponize technology, an aide to create other novel technology, or as a threat actor itself.[5] AI experts, industry leaders, and governments have increasingly voiced concern that those threats on a larger and integrated scale could fuel

[1] European Union, European Parliament "How Artificial Intelligence Works ". March, 2019
[2] UN AI Advisory Group, "Governing AI for Humanity". Sept. 2024
[3] Harvard University, "Into the Abyss: Examining AI Failures and Lessons Learned". https://www.ethics.harvard.edu/blog/post-8-abyss-examining-ai-failures-and-lessons-learned
[4] OECD, "Defining AI Incidents and Related Terms". 2024. www.oecd.org/en/publications/defining-ai-incidents-and-related-terms_d1a8d965-en.html
[5] Crowdstrike, "AI-Powered Cyber attacks". Jan. 2025.

many serious and more visceral threats: its capacity to increase the severity and speed of cyber and physical threats to critical infrastructures in healthcare and energy sectors, economic disruptions, or lowering the barrier for the production and use of bioweapons and autonomous weapons used by state and nonstate actors.[6] This brief is particularly concerned with emergency response capabilities to prevent and respond to the types of AI incidents that could cause serious injury and/or harm, up to and including *AI disasters*-occurrences so disruptive that they may exceed a community or State's ability to respond and mitigate.[7]

The Case for Disaster Preparedness in the Age of Artificial Intelligence

Many countries and AI developers have begun to recognize some of the risks from advanced AI, though current regulatory, diplomatic and security approaches to address both recognized and unrecognized threats remain underdeveloped.

States have sparsely sought to reduce AI risks through regulatory frameworks, governance and safety research initiatives, strategic dual use investments, and international cooperation. The US, in its since-rescinded 2023 Executive Order, planned to start addressing the risks of AI capabilities within biosecurity, cybersecurity, and critical infrastructure through risk assessment and standardization lines of efforts among federal and industry stakeholders.[8] In the European Union, the EU AI Act establishes risk management, incident reporting and certain prohibitions for high-risk AI systems.[9] Africa has established an AI Council and is working toward safety implementation alongside the African Union's AI Strategy.[10] Singapore specifically noted AI's threat to national cybersecurity. Multilateral institutions such as the UN, the G7, OECD are currently engaged in managing systems and systemic level ai risks at a global scale, including on the ethical and security considerations surrounding the use of Lethal Autonomous Weapons Systems (LAWS).[11] Other countries such as China has incorporated an algorithmic registry into governance frameworks. Moreover, intelligence agencies and countries with Cyber Emergency Response Teams are engaged with AI labs, industry stakeholders and AI

[6] www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat; International AI Safety Report, January 2025. International_AI_Safety_Report_2025_accessible_f.pdf; statement-on-ai-risk
[7] OECD, "Defining AI Incidents and Related Terms". 2024.
[8] Federal Register : Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence
[9] European Union, EU AI Act, July 2024.
[10] Brookings Institution, "Building regional capacity for AI safety and security in Africa". May 2025.
[11] OECD "Assessing potential future artificial intelligence risks, benefits and policy imperatives", OECD Artificial Intelligence Papers". 2024; UN AI Advisory Group, "Governing AI for Humanity". Sept. 2024; G7 Leaders Statement on AI for Prosperity" https://g7.canada.ca/en/news-and-media/news/g7-leaders-statement-on-ai-for-prosperity/. June, 2025.

experts to assess immediate and medium-term threats. National security agencies are engaged in information sharing internationally to combat malicious cyber actors and emerging technology threats, while there are early efforts to share incident and safety reports from AI developers in multilateral fora[12].

However, the patchwork of frameworks for risk management, the current state of defensive AI capabilities, and expert consensus on loss of control risks all indicate broad challenges in security planning to combat large AI incidents. States' reliance on AI developers to develop and incorporate security and safety features for AI systems, without mandated standards, transparency, and accountability measures, indicate a fundamental failure within a first line of defense against potential AI misconduct. States have implicitly and explicitly communicated a desire not to stifle innovation from the AI industry.[13] AI labs, whose financial interests and public safety responsibilities are inherently divergent from the interests of safety and security stakeholders, do not have strong incentives to maintain sufficient safety and security measures. Further, there is inscience amongst policy makers and the public of how AI systems work at a baseline technical level, and amidst AI developers themselves at deeper compute levels.[14] Subsequently, AI developers establish risk management into their product design that are insufficient to combat the growing risks to individuals, systems, and property. As a result, extant risk management frameworks remain largely voluntary, and insufficiently scoped requirements and liability regimes for "high-risk" systems fail to specifically target AI hazards from general-purpose systems, or their downstream effects, that could lead to serious large-scale incidents.

Moreover, the lack of reporting requirements for AI incidents, and the nascence of AI risk research, create additional obstacles for risk-based planning. Industry fears of legal exposure, fragmented governance frameworks, and the nascency of risk assessment for the rapid evolving of AI capabilities and its comprehensive integration all contribute to relevant stakeholders' distorted path for understanding and building consensus on AI-related security policy. With the knowledge that there is a lack of full consensus on the timeline for development and capability of highly advanced AI, as well as the actual threat of loss of control risks, the described confounding considerations create a cyclical feedback loop for the current immaturity of AI governance.[15]

Further, governments have begun to invest in offensive and defensive AI capabilities to augment their own national security strategies, which include attempts to minimize the

[12] OECD G7 Hiroshima AI Process Voluntary Reporting Framework
[13] Columbia University, "The False Choice Between Digital Regulation and Innovation". 2024; "Request for Information on the Development of an Artificial Intelligence (AI) Action Plan" Feb. 2024. www.federalregister.gov/documents/2025/02/06/2025-02305/request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan
[14] AI Action Summit, International AI Safety Report, 2025. Assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International_AI_Safety_Report_2025_accessible_f.pdf
[15] Id.

adversarial threat of AI itself. Key players within AI development, including the US, UK, Russia, and China, as well as multilateral institutions such as the EU and NATO, are investing in defensive AI capabilities to enhance detection and reaction to cyber threats in energy infrastructure, or surveillance.[16] Similarly, countries are investing in AI offensive products to, among others, enhance surveillance, intrusion, and imaging for autonomous devices. Offensive AI weapons and AI-enabled cyber offense technology are outpacing their defensive counterparts and have considerable ethical and escalatory risks.[17] For policy makers, the challenge of regulating dual use technology while working toward national security and economic interests prompt regulatory hesitancy, increasing the imbalance between safety and development. Additionally, due to resource constraints, many smaller businesses and units of government are less able to invest in defensive AI applications, increasing their vulnerability to attacks from state and nonstate actors.[18]

Historically, States have belatedly worked toward national and international response regimes to combat nationally significant security threats. Whether it be coordination and intelligence authorities to combat terroristic threats, such as the profound reorientation of the U.S. homeland security apparatus in the aftermath of 9/11, or the nonproliferatory treaties concerning chemical, biological, and nuclear weapons. Related to emergency preparedness specifically, treaties such as the emergency communications treaty, or the convention to marshall assistance in the case of a nuclear accident – and the recently signed treaty on pandemics – were created and signed after failure, realized yet foreseen threats, and a need for coordination. The pace of AI's evolving capabilities, the uncertainty of its mechanics, and the potential for large-scale harm from AI systems, across sectors and societies, and historical belated prevention and response efforts – illustrate less risk tolerance for post-factum security measures.

Building Capacity for AI-Related Emergency Response

In the face of significant natural disasters, CBRN incidents, terrorist attacks, and, of increasing importance, significant cybersecurity threats, many states have developed a

[16] NATO, "Emerging and Disruptive Technologies". 2025. www.nato.int/cps/en/natohq/topics_184303.htm; US Cyber Command "USCYBERCOM Unveils AI Roadmap for Cyber" Operations. www.cybercom.mil/Media/News/Article/3905064/uscybercom-unveils-ai-roadmap-for-cyber-operations/; Center for New American Security "The Role of AI in Russia's Confrontation with the West" www.cnas.org/publications/reports/the-role-of-ai-in-russias-confrontation-with-the-west

[17] International AI Safety Report, January 2025. International_AI_Safety_Report_2025_accessible_f.pdf

[18] Id.

layered approach toward prevention and response.[19] Amidst the potential for large and evasive AI hazards and threats, countries should review their preparedness systems to account for AI's ubiquity and potency.

Many countries and intergovernmental organizations have begun to incorporate the risks from emerging technology into their preparedness models and designations of technology in sector risks frameworks, though these efforts are incipient. Research indicating AI's potential to create new and exponential harm indicates the potential for preparedness postures to become overwhelmed.[20] Though full-scale AI incidents are nascent, the evidence for serious harm and rapidly evolving AI capabilities should prompt greater global AI risk preparedness. In order to maintain a robust security posture that prevents, deters, and mitigates AI-enabled threats, countries will need to take a national approach toward national preparedness that orients a state's security apparatus to prioritize AI threats, while enhancing the legal and public awareness environment.

Countries with relatively mature preparedness systems have emergency response apparatus that consist of regulatory requirements and national plans that prioritize, authorize, and resource response mechanisms. More effective protocols are tailored to specific realized and incoming threats and follow a common emergency response framework; prevention and awareness, monitoring and detection, public warning, response and mitigation, and remediation and development.[21] Cognizant of the difference in resources and capacity for response for a given incident, national and subnational units of emergency response stakeholders typically have a clear understanding of the delineation of responsibility for large-scale incidents of familiar types. Additional indicators of mature emergency preparedness include widescale and efficient communication mechanisms between government and the public, sufficient legal authority to compel private action as appropriate, and international cooperation. It is largely difficult to assess the efficacy of national and international approaches for combatting complex, and often chronic, large-scale incidents. However, many states have moved to establishing capability metrics, routine review, and stress test exercises. Effective strategies will require a multi-pronged, whole-of-nation approach toward resilience, including routine novel technology scans integrated into their preparedness assessments, a strong legal culture on AI risks addressing users, affected non-users, developers, deployers, and hosts, strong public stakeholder knowledge and participation in security incidents and exercises, global cooperation, and sufficient resources for response and remediation.

---

[19] The International Journal of Disaster Risk Reduction, " Top hazards approach – Rethinking the appropriateness of the All-Hazards approach in disaster risk management", 2020. doi.org/10.1016/j.ijdrr.2020.101559

[20] www.science.org/doi/10.1126/science.adn0117; RAND, " The Operational Risks of AI in Large-Scale Biological Attacks. www.rand.org/pubs/research_reports/RRA2977-1.html; OpenAI, " Preparing for future AI capabilities in Biology", preparing-for-future-ai-capabilities-in-biology

[21] "Top Hazards Approach"; FEMA, National Preparedness Goal, www.fema.gov/emergency-managers/national-preparedness/goal

Recommendations

Prevention

Prevention efforts against large-scale AI incidents are a crucial component to national preparedness because they reduce risk by attempting to minimize vulnerabilities and mitigate potential and current harms. The risk of serious harm among AI's rapidly evolving capabilities across layers of societal exposure and integration demand a robust proactive framework that emphasizes cognizance of interconnectedness.[22] In proactively addressing the hazardous aspects of AI development, deployment, and use, States can reduce the often larger response and recovery costs, resulting in less harm and injury to society and the public. In order to facilitate better preparedness, below we provide key legal, institutional and emergency preparedness objectives that states should endeavor toward in the near term.

Legal and Oversight

- ➢ As seen during large-scale cybersecurity incidents, the owners and operators of AI systems will sometimes be the first to discover failures, misuse or malfeasance, and other hazards that could lead to significant actual harm. As such, States should establish legal requirements to ensure AI developers, deployers, and hosts have public safety obligations, including incident reporting and coordination measures during emergency circumstances.
- ➢ States should provide legal protections for employees who whistleblow on imminent or other potential large scale AI threats.
- ➢ To encourage early notification of potential large-scale AI-based concerns from industry and the public, States, through legislation, should consider liability protections to encourage disclosure and mechanisms to handle proprietary commercial information for use in emergency circumstances.

Institutional Infrastructure

- ➢ Governments should establish regular coordination mechanisms, industry, and civil society to support research, governance, and monitoring objectives.
- ➢ Emergency management agencies should incorporate the breadth of large-scale AI incidents into their national readiness mechanisms to ensure relevant stakeholders

[22] UN AI Advisory Group, "Governing AI for Humanity". Sept. 2024 governing_ai_for_humanity_final_report_en.pdf

have the knowledge, training, and resources necessary for appropriate preparedness. Extending current cybersecurity exercises to address current and potential AI incidents will help to orient readiness mechanisms more robustly.

- ➢ Governments should establish cadres of technical experts with national security and law enforcement agencies and a mechanism for coordinating swift collaboration with outside experts. These experts would either operate in their own expert capacities at universities, research centers, and other civil society organizations and be able to quickly assess and advise relevant security stakeholders during emergency circumstances, or they can be used in a reserve internal consultant capacity. States should work with their relevant intelligence and defense apparatus to undertake necessary precautions to allow for the immediate collaboration with relevant external stakeholders, particularly in dealing with sensitive information. This is especially relevant for threats from the novel technologies and unknown unknowns we can expect from transformative AI.
- ➢ Technology-enabled threats and threat actors increasingly blur the lines between foreign and domestic sources of threats, which have impacts on civil-military capacity to respond. National security stakeholders are already sounding the alarm on hybrid threats or campaigns, or the results of cyber attacks having digital and physical consequences. States should ensure that there are clear lines of responsibility between civil and military defense agencies, appropriate mechanisms in place for interagency coordination, and appropriate collaboration when necessary.

Critical Systems Protection

- ➢ National and subnational units should collaborate to maintain awareness of AI risks toward cyber and physical critical infrastructure systems, particularly in the energy and water sectors. This effort should include inventory of high-risk systems integration into sensitive government and critical infrastructure systems. It should also consider environmental resources such as air and soil to be critical infrastructure.
- ➢ Governments should harden IT systems, using verified software where possible, and also ensure that rapid response security protocols are recognized in sector-based emergency response plans.
- ➢ States should adopt screening standards for individuals and labs engaged in research of highly infectious microorganisms and hazardous bio materials. A national system to evaluate the use of CBRN-relevant materials, heighten the control of material access, and inventory/incident reporting is an important strategy to counter the threat that AI will by default lower the barriers for CBRN threats.
- ➢ Operators and owners of important cyber-physical systems (CPS) and emergency managers should conduct routine threat assessments to ascertain potential threats to these systems. AI can be used to cause technology-enabled attacks on industrial

control systems, medical equipment, energy grids, and nuclear systems, among others. Threat scanning and red-teaming can assess the threat environment at multi-scale systems and geographic levels, and AI/ML can help detect anomalies, false and/or malicious data, and human manipulation techniques in a manner that strives to close the gaps with the speed and invasiveness of AI-cyber tactics.

Public Awareness

- National and subnational governments, together with civil society stakeholders, should build greater stakeholder awareness about AI-related threats and emergency preparedness procedures.
- Governments should work toward greater awareness of AI-enabled misinformation and deepfakes and build resilience against circumstances in which human manipulation or employment by AIs can lead to significant emergencies or the instrumentalized exacerbation of existing threats.

International Engagement

- States should cooperate on a variety of emergency prevention efforts in the AI risks context. This should include monitoring, detection, threat model and broader information sharing mechanisms, resource sharing, identification and prioritization of communities and states particularly vulnerable to emergencies.
- AI governance is on the agenda of many regional and multilateral forums, though emergency response within regulatory discourse is in its latency. States should increase their engagement in awareness and prevention efforts of potential AI incidents, emergencies, and crises to shore up resilient defenses, proactively

Response

Effective response capabilities and plans that are adaptive to serious AI incidents are of vital importance to reduce the impact of disasters while maintaining public safety and critical functions that communities rely on. Strengthening existing emergency response protocols by emphasizing rapid detection and containment, coordinated and collaborative action among government, industry and civil society, and resilient communication systems are essential to adept national preparedness regimes.

Threat Detection and Assessment

- States should work toward establishing mechanisms that monitor and detect emerging technology- and cyber-derived threats through intelligence and civil defense apparatuses. Many countries do this through civil, military, and intelligence information sharing agreements at the national and international level. More countries should adopt this approach, especially in high-consequence domains not typically subject to conventional cyber risk, like biotechnology.
- As the nature and methods of AI-enabled CBRN- and cyber-relevant threats become clearer, states should implement adapted CBRN & cyber detection technologies at strategically prioritized facilities and sensor locales.
- In States' preparedness systems (such as FEMA's National Incident Management System), there should include further integration of assessments that evaluate the potential impacts of technology-enabled threats, at regular intervals.
- States should establish a mechanism for determining conditions that would activate emergency response protocols given each type of threat that may be attributable to AI. These mechanisms should tailor to each State's specific security environment, yet given the cross-border and digital connectedness of AI and critical societal systems, they should also consider international engagement on alerts and definitional harms where possible.
- Operators and owners of important or dangerously exploitable cyber-physical systems (CPS) and emergency managers should conduct routine threat assessments to ascertain potential threats to these systems. AI can be used to cause technology enabled attacks on industrial control systems, medical equipment, energy grids, and nuclear systems, among others. An attack might also take control of robotics to perform physical actions in an otherwise protected environment. Threat scanning and a red team can assess the threat environment at systems and geographic levels, and AI/ML can help detect anomalies, false and/or malicious data, and human manipulation techniques, in a manner that strives to close the gaps with the offensive AI tactics.

Threat Mitigation

- States should review their emergency procedures routinely to account for threats based on varying timeline estimates and severity of potential hazards and threats.
- States should perform table top exercises across units of government and across critical sectors, that test for both capacity for response and resourcing to AI threats as enablers and amplifiers. Metrics should be established to determine capability targets based on the probability that AI causes given degrees of damage.
- States should explore strategies to isolate or shut down offending, compromised, or vulnerable high-risk systems, including limiting their access to external networks

and, where appropriate, restricting power or connectivity to prevent further risk propagation or uncontrolled escalation.
- States should develop protocols for ensuring that misinformation does not exacerbate emergency procedures.

Emergency Communications

- Architect domestic emergency services to be resistant to emerging technology hacking, loss of carrier services, loss of grid power, and radiation spikes.
- Low income, ability-impaired, poorly educated, and rural communities are commonly more vulnerable to a significant range of hazards. As AI threats come to be realized, Countries should ensure preparedness awareness, physical resources, and protection accessible to these groups.
- States should ensure their laws and national communications procedures are in-line with or exceed the provisions of the International Emergency Communications Treaty.

International Engagement

- Threat intelligence sharing mechanisms at regional and broader multilateral levels should include protocols for sharing AI-based threat information, and coordinate with industry, at the international level, to ensure incident reporting before, during, and after an emergency, where appropriate.
- States working through multilateral institutions should develop frameworks or build on existing frameworks to determine where emergency response and remediation resources specific to AI-driven threats can be developed and shared.

Remediation

- In both live situations and in the aftermath of large-scale AI incidents or disasters, States should incorporate AI forensic examination expertise into their active and post-mortem investigations.
- Relevant authorities should work collaboratively with industry and media to ensure that mis- and dis-information on response efforts are minimized.
- Relevant authorities should ensure computing and data reserves for restoring capacity after large-scale events affecting government devices and systems.
- In preparation as AI capabilities evolve and incidents increasingly realize, States should establish and update contingency planning for critical services and systems as forecasted necessary per plausible threats.

- ➢ Develop domain-specific rapid response remediation capabilities in areas like CBRN and public health. To prepare for threats from novel unseen technologies that advanced AI systems can design, prepare ready access to broad and high competency in the sciences and engineering so as to counter new technologies in near-real-time.