GENERATIVE A.I. REGULATION AND CYBERSECURITY

A GLOBAL VIEW OF POLICYMAKING

Global Cybersecurity Group | ASPEN DIGITAL

aspen institute

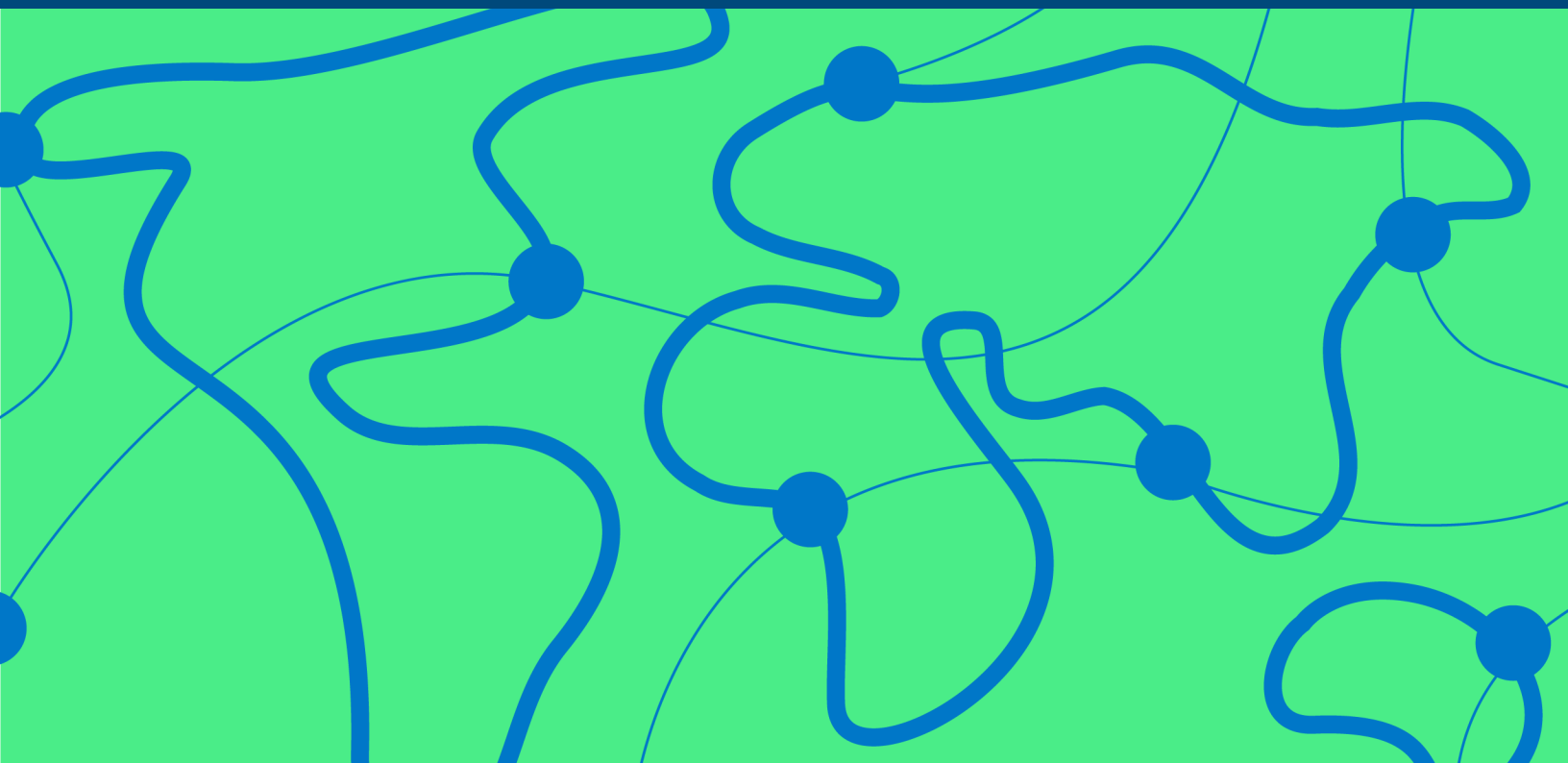


TABLE OF CONTENTS

ABOUT THE GLOBAL CYBERSECURITY GROUP

Founded in 2022, the Aspen Institute's Global Cybersecurity Group is a forum of cross-sector cybersecurity experts who share a commitment to democracy, security, and freedom online. Its agenda is member-driven and informed by international events and their implications in both the digital and physical worlds.

This paper serves as the group's inaugural publication. Chartered in April 2023 based on group consensus, this paper aims to cut through the flurry of AI-focused regulatory activity over the past year and provide a succinct summary of efforts and their focus areas on cybersecurity. It also provides guidance on areas where governments should lean in on regulation and where they should proceed with caution.

INTRODUCTION

In an era marked by unprecedented technological advancements, the explosion of generative artificial intelligence (GenAI) in the public consciousness stands out. It has the possibility to change how people live, learn, and work, and has already shifted the paradigm in cybersecurity. It is therefore not surprising that governments around the world are looking closely at GenAI and how it will impact the lives of their citizens. Some have already begun to regulate or otherwise oversee the usage and development of GenAI tools, while others are moving more cautiously, focusing their efforts on discovery and research. In both cases, GenAI (or foundational) models pose particular regulatory challenges given their adaptability and range of use.

As organizations and individuals entrust an ever-increasing amount of sensitive data to digital systems, the stakes of getting cybersecurity right have never been higher.

As an aid to all such governments and intergovernmental bodies, the Aspen Institute's Global Cybersecurity Group convened a working group to develop guidance on how—and how not—to regulate, oversee, or otherwise address the explosion of GenAI tools when it comes to cybersecurity.

The digital landscape, once dominated by human-operated defenses, now stands at the crossroads of innovation and vulnerability, with AI emerging as both a formidable weapon and a critical shield in the ongoing battle against cyber threats. As organizations and individuals entrust an ever-increasing amount of sensitive data to digital systems, the stakes of getting cybersecurity right have never been higher. Cyberattacks are a daily occurrence and threaten financial stability, national security, and public safety. Against this backdrop, GenAI is a technological marvel capable of learning, adapting, and responding to threats at a pace that surpasses human capacity. The actions that governments, companies, and organizations take today will lay the foundation that determines who benefits more from this emerging capability—attackers or defenders.

In a world where security breaches can have far-reaching consequences, the synergy between AI and cybersecurity is not merely an option–it is an imperative.

A convergence of advances has led to the emergence of GenAI as a disruptive new capability. The availability of large datasets, improvements in deep learning algorithms, increases in computing power, and innovations in training computers have enabled AI systems to create highly realistic synthesized content. Unlike previous AI systems focused on analysis, generative models can unlock creative applications. Progress in the last few years has been remarkably fast. Generative or foundational models can now produce high-fidelity images, human-like text, and natural speech. GenAI promises to revolutionize content creation, art, entertainment, digital marketing, and many other industries. At the same time, it has introduced mounting risks and potential harms, both known and unknown. Not surprisingly, governments are feeling pressure to manage this revolution.

The rapid commercialization of AI tools signals a transition from research concept to real-world deployment. The democratization of GenAI through easy-to-use consumer products, APIs, platforms, and cloud services has enabled widespread adoption. Individuals and startups now have access to capabilities only large tech firms possessed a few years ago.

In a world where security breaches can have far-reaching consequences, the synergy between AI and cybersecurity is not merely an option—it is an imperative. Otherwise, the trust we have in widely available authentication measures may erode as GenAI systems broaden training inputs and create increasingly compelling life-like outputs, modeled after and meant to impersonate real individuals. As that trust erodes, we will miss the opportunity to have proactive conversations about the permissible uses of GenAI in threat detection and examine the ethical dilemmas surrounding autonomous cyber defenses as the market charges forward.

With AI as both a potential sword and shield, the future of cybersecurity is as promising as it is uncertain.

As we navigate this uncharted territory, it is crucial to decipher the potential of AI as a guardian of the digital realm while remaining vigilant to the ethical and practical considerations that accompany its deployment. With AI as both a potential sword and shield, the future of cybersecurity is as promising as it is uncertain.

As this technology continues to advance, it is important to analyze its current abilities and limitations in the domain of cybersecurity. Like any transformative technology, GenAI creates new attack vectors even as it improves defenses.

Possible risks include:

- **Creating deepfakes for fraud and scams**
- **Automating phishing and social engineering**
- **Impersonating identities online**
- **Generating malicious code and content**
- **Evading AI-based detection systems**

However, GenAI can also counter these threats by detecting generated content and malicious use. For example, a pattern-matching AI could be used for anomaly detection or classification. Overall vigilance, de-identification, and human oversight are key to maximizing the cybersecurity benefits of GenAI while minimizing harm. As governments dash to install legal and regulatory safeguards, organizations should adopt a multi-faceted approach that includes robust testing, ongoing monitoring, threat modeling, and ethical considerations. Additionally, combining AI with human expertise and maintaining a proactive stance in cybersecurity practices remains crucial to safeguarding digital assets and systems. Finally, these technologies only benefit cyber defenders if they are adopted. It is critical that policymakers consider procurement approaches that enable the adoption of innovative security technologies.

Overly prescriptive policies could stymie progress while permissive frameworks could allow otherwise avoidable risks.

Discussions around GenAI regulation picked up pace throughout 2023, as the technology became increasingly powerful and pervasive. There is a growing consensus that AI needs a governance structure (regulation or otherwise) to ensure that it is developed and used safely and ethically. However, there are wide-ranging debates around the world about how best to do this, given the complexity of the technology and the potential for unintended consequences. Overly prescriptive policies could stymie progress while permissive frameworks could allow otherwise avoidable risks.

In the end, governments will need to find a balanced approach. Mandating dataset openness, human oversight, and transparency in commercial generative models can reduce risks, and industry self-regulation, governance, and codes of ethics are also constructive steps.

AI is a cross-cutting and global issue, and the development of governance principles and frameworks must take into consideration local historical, cultural, and political contexts. These efforts are essential, as most jurisdictions will need time to craft and pass regulation that is both effective and minimizes unintended side effects. Outright bans of the technology or its applications are infeasible given GenAI's myriad uses and ease of access. But targeted legal guardrails guiding both GenAI development and thresholds for undue harm could be effective. Their advancement requires both international collaboration and the right technological expertise with these governing bodies.

AI is a cross-cutting and global issue, and the development of governance principles and frameworks must take into consideration local historical, cultural, and political contexts.

As of the end of 2023, many governments were struggling to identify the role they should play with respect to this rapidly developing technology. This paper delves into the multifaceted ways GenAI is transforming cybersecurity and will shed light on the opportunities and challenges that arise from this convergence.

GLOBAL A.I. GOVERNANCE

FIVE PREDICTIONS BASED ON PREVIOUS CYBERSECURITY POLICYMAKING

There are many themes from previous cybersecurity policymaking and implementation that apply to today's GenAI regulatory processes as governments work to account for the exponential growth in risk and opportunity. For one, past efforts have shown that consistent approaches across like-minded nations provide the foundation to successful governance driving better security. The more baselines and standards to adhere to, the more likely it is for a focus on compliance to overtake a focus on security.

Other lessons from cybersecurity policymaking can inform predictions about the AI-related regulations and cybersecurity outcomes to come:

1

Disclosure: Cyber incident reporting requirements vary widely across the globe, with organizations required to disclose an incident anywhere from 4 to 48+ hours after discovery. With GenAI increasing the speed of both offensive and defense cyber operations, governments may feel pressure to shorten the window for these disclosures moving forward, which may limit an organization's ability to provide human oversight in assessing and remediating the incident in the critical hours after discovery.

2

Attribution: The ability to determine responsibility in cyberspace will be complicated by GenAI technologies, as adversaries have more tools to hide their identities and activities. This will be true for forensics professionals in both government and industry as obfuscation applications of GenAI models mature.

3

Data Jurisdiction: The rise of offshore data centers has raised several questions about data privacy, jurisdiction, collection, and storage mechanisms. Governments are already working to limit the input of individuals' information into GenAI models. We expect similar conversations about data jurisdiction to continue regarding the outputs and outcomes of these models.

4

Leadership Accountability and Liability: Historically, indictments of individual leaders for cybersecurity-related wrongdoing have been relatively rare, however they are becoming more frequent.[1] The emerging Chief AI Officer role may see similar legal or criminal exposure for any incomplete cybersecurity-related or risk-related disclosures.

5

Cyber Assistance: Given the global reach of both cyber- and GenAI-related harms, we anticipate a greater investment in and desire for international capacity building programs, geared at both remediating attacks and proactively hardening cyber defenses.

[1] "SEC Charges SolarWinds and Chief Information Security Officer with Fraud." *U.S. Securities and Exchange Commission*, 31 Oct. 2023, www.sec.gov/news/press-release/2023-227.

A MAP OF CURRENT CONVERSATIONS ON GLOBAL A.I. GOVERNANCE

As GenAI technologies and applications grow, effective regulation will be critical to addressing the challenges, risks, and opportunities. Governments and international organizations have started several governance processes and initiatives, including both national and multistakeholder initiatives, each with varying degrees of cybersecurity focus. These discussions collectively shape the global approach to governing GenAI, making it crucial to assess their interplay and synergies to guide governments and international regulatory bodies in developing informed and effective policies.

UNITED STATES
AI Executive Order

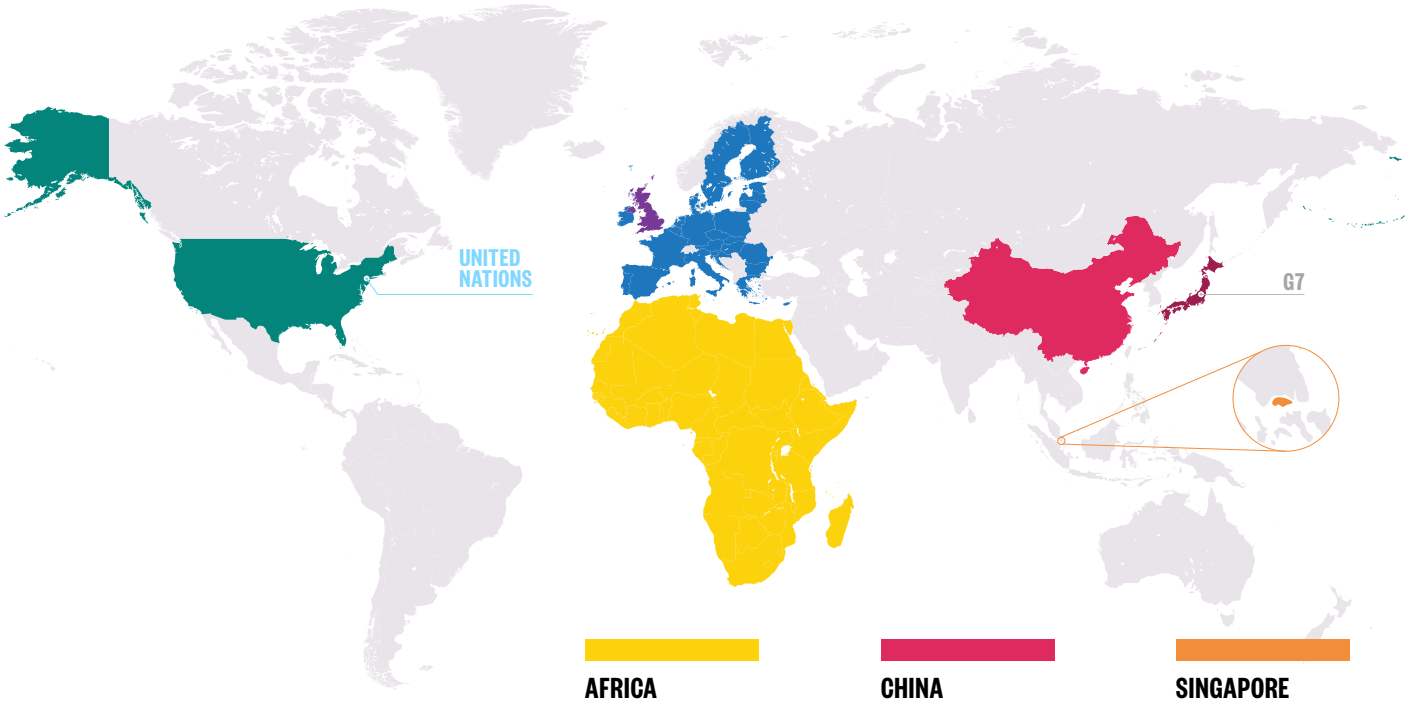
EUROPEAN UNION
AI Act, GDPR

JAPAN

UNITED NATIONS
GDC Process

UNITED KINGDOM
Bletchley Declaration

G7
Hiroshima AI Process



AFRICA
AI Continental Strategy for Africa

CHINA
Draft New/Next Generation Artificial Intelligence Development Plan and Measures for the Management of Generative Artificial Intelligence Services

SINGAPORE
The Model AI Governance Framework

EXAMINING EFFORTS

This list shows a spectrum of the approaches, value sets, and focus of many of the Generative AI governance efforts pertaining to cybersecurity across the globe. The group recognizes it is not inclusive of all such efforts.

AUTHORITY *Relevant Effort(s)*	POLICYMAKING APPROACH	TARGET	PROMINENT CYBERSECURITY PROVISIONS OR THEMES
UNITED NATIONS *Global Digital Compact Process*	Management based	193 UN member states	• Critical Infrastructure Security • Cyber Operations
EUROPEAN UNION *AI Act* *General Data Protection Regulation (GPDR)*	Risk and Rules based	27 EU member states	• Security Risk Assessment • Data Governance
G7 *Hiroshima AI Process*	Management based	7 member states	• Vulnerability Remediation • Incident Reporting
UNITED KINGDOM *Bletchley Declaration*	Risk based	28 signatory countries and the European Union	• Specifies cybersecurity frontier risks as one of two notable domains of concern with "potential for serious, even catastrophic, harm"
UNITED STATES *AI Executive Order*	Risk based	United States	• Vulnerability Remediation • Standards Development • Security Assessment and Submission

CHINA *New/Next Generation Artificial Intelligence Development Plan and Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)*	Rules based	China	• Security Assessment and Submission • Authentication
JAPAN	Goals based	Japan	• Disclosure • Developer Feedback Loops
SINGAPORE *The Model AI Governance Framework*	Management based	Singapore	• Secure Development of AI Tools
AFRICAN UNION *African Union Artificial Intelligence (AU-AI) Continental Strategy for Africa* *Malabo Convention*	Goals based	55 member states	• Privacy and Data Protection • Secure e-Commerce • Cybersecurity and Cybercrime

POLICYMAKING APPROACH

- **Goal based:** An authority sets out an objective, rather than exact rules, specifications, or standards.
- **Management based:** An authority facilitates a process to include research, collaboration, or other exploratory activities to inform action.
- **Risk based:** An authority defines regulations based on its assessment of risks and mitigations.
- **Rule based:** An authority issues a set of rules with fines or punishments expected for transgressions.

As of this paper's publication, many regulatory bodies are continuing to refine the specifics around their intended targets, rules, incentives, and consequences for GenAI use in cybersecurity.

UNITED NATIONS (UN)

The UN's regulatory initiatives on GenAI revolve around three essential themes: ethical development, international cooperation, and the protection of human rights. The overarching objective is to establish responsible governance frameworks that ensure transparency, accountability, and cybersecurity, enabling the harnessing of GenAI's potential while mitigating its societal and ethical challenges on a global scale.

Noteworthy initiatives include UNESCO's groundbreaking 'Recommendation on the Ethics of Artificial Intelligence,' adopted by all 193 Member States in November 2021. This recommendation prioritizes human rights and dignity, emphasizing principles like transparency and fairness, with a focus on human oversight of AI systems. Building on UNESCO's guidance, the UN High-Level Committee on Programmes (HLCP)—Inter-Agency Working Group on Artificial Intelligence released preliminary operational guidance and a series of 10 principles, rooted in the UNESCO Recommendations, for the application of AI by the UN System in September 2022. In addition, the UN Department of Management Strategy, Policy, and Compliance, along with the UN Department of Field Support and the UN Digital and Technology Network (DTN), has issued internal guidance on the use of GenAI for all UN staff members.

Furthermore, within the UN system, the ongoing Global Digital Compact (GDC) process, led by the UN Secretary-General, has introduced a High-Level Advisory Body for AI.[2] This body aims to assemble experts from states, relevant UN entities, industry representatives, academia, and civil society to provide recommendations for international AI governance. Additionally, the proposal includes a digital human rights advisory mechanism facilitated by the Office of the High Commissioner for Human Rights (OHCHR). This mechanism is designed to offer practical guidance on the intersection of human rights and technology issues. These UN-led governance initiatives reflect the organization's proactive response to the increasing public discourse regarding appropriate mechanisms and platforms for global AI oversight.

[2] "New UN Advisory Body Aims to Harness AI for the Common Good." *United Nations*, 26 Oct. 2023, news.un.org/en/story/2023/10/1142867.

In sum, these initiatives are interlinked, united by a commitment to responsible, ethical, and secure AI development. Each initiative builds upon the principles and recommendations of the others, forming a comprehensive approach to address the challenges and opportunities presented by GenAI within the realms of cybersecurity and global governance.

EUROPEAN UNION (EU)

Europe is spearheading some of the most advanced initiatives in this field, including the EU's proposed risk-based AI Act, which EU policymakers agreed to on December 8, 2023. While its jurisdiction is limited to the EU, it will have extraterritorial impacts across the globe given its applicability to all entities with operations in the EU. Policymakers will soon finalize details in the law, which is expected to take effect in 2025 at the earliest. Until the rules are fully applicable, the EU is asking companies to voluntarily commit to implementing key parts of the Act by signing an AI Pact.[3] [4]

The EU dimension of regulatory efforts, particularly the EU AI Act, holds significant international relevance in addressing GenAI's challenges and opportunities. Among its key elements are stringent rules governing high-risk AI systems, transparency requirements, and comprehensive data governance measures. The new rules describe high-risk AI systems as posing a significant risk to critical infrastructure, medical systems, education, law enforcement, and democratic processes. While minimal risk AI systems may only be subject to transparency rules, high-risk systems will be required to comply with requirements including detailed documentation, risk mitigation, activity logging, and human oversight.

However, its application to GenAI, which includes deep learning models like GPT-3, presents unique challenges. GenAI systems produce creative and potentially unpredictable outputs, making risk assessment complex. Striking a balance between innovation and accountability is crucial, requiring tailored guidelines for

[3] Von der Leyen, Ursula. "Statement by President Von Der Leyen on the Political Agreement on the EU AI Act." *European Commission*, 9 Dec. 2023, ec.europa.eu/commission/presscorner/detail/en/statement_23_6474.

[4] "AI Pact." *European Commission*, 15 Nov. 2023, digital-strategy.ec.europa.eu/en/policies/ai-pact.

GenAI. Moreover, addressing issues of bias, intellectual property, and content generation ethics will be essential within the framework of the EU AI Act.

Contained in the latest public draft are new transparency requirements for the foundation models underpinning GenAI, including publishing summaries of algorithm training content in compliance with EU copyright laws. Foundation models posing a "systemic risk" are subject to further requirements, including model evaluations, risk assessments, incident reports, and energy efficiency standards.[5]

The latest draft of AI Act institutes bans on several AI uses, including the bulk scraping of facial images to build databases, social scoring, and emotion recognition in the workplace. Live facial recognition is also restricted, with some exceptions for national security and law enforcement purposes.

Once the law takes effect, companies that are not in compliance with the Act will be fined in the range of 1.5% to 7% of global sales or up to 35 million euro, whichever is greater. The newly-formed European AI Office within the European Commission will oversee coordination among European authorities, as well as implementation and enforcement of the rules on general purpose AI.

In addition to EU-specific policies, there are global-level efforts underway, facilitated by the Council of Europe's Committee on Artificial Intelligence (CAI), which is presently in the process of developing the world's inaugural AI treaty. Despite its origins within a European entity, this instrument, projected to come into existence in 2024, has the potential to establish itself as a global benchmark that can be embraced by nations beyond the Council of Europe.

[5] Foon Yun Chee. "What Are Europe's Landmark AI Regulations?" *Reuters*, 9 Dec. 2023, https://www.reuters.com/technology/what-are-europes-landmark-ai-regulations-2023-12-09/.

G7

In May 2023, G7 Leaders met in Hiroshima, Japan and published a communiqué based on discussions on responsible AI and AI governance.[6] The communiqué highlighted a need for interoperable AI governance frameworks, but noted that approaches may vary among member countries. Leaders encouraged the development of international technical standards and tools for trustworthy AI through multistakeholder approaches. In acknowledgement of the rise of GenAI, the communiqué also established the "Hiroshima AI process."

The Hiroshima AI process is a working group in cooperation with the OECD and the Global Partnership on AI (GPAI) conducting discussions on the opportunities, risks, and policies associated with GenAI. Topics include "governance, safeguard of intellectual property rights including copyrights, promotion of transparency, response to foreign information manipulation, including disinformation, and responsible utilization of these technologies."

In September 2023, the OECD Directorate for Science Technology and Innovation (STI) published a report[7] to inform the ongoing Hiroshima AI process. The report detailed results from a questionnaire circulated to G7 members in June, including:

- All seven G7 members identified "disinformation/ manipulation" as the top risk presented by GenAI in achieving national and regional goals.
- Several G7 members stressed challenges that require international cooperation, including "preventing the use of GenAI to create chemical or biological threats (e.g. viruses), or massive disinformation/misinformation (including from foreign actors)" and "addressing international AI cyber security risks on a global level."
- G7 members indicated the most urgent and important action they can recommend is "providing effective tools for safety, quality control, and capacity / trust building, and voluntary codes of conduct."

[6] "G7 Hiroshima Leaders' Communiqué." *G7 Hiroshima*, 20 May 2023, www.g7hiroshima.go.jp/documents/pdf/Leaders_Communique_01_en.pdf.

[7] "G7 Hiroshima Process on Generative Artificial Intelligence (AI)." *OECD*, 7 Sept. 2023, www.oecd.org/publications/g7-hiroshima-process-on-generative-artificial-intelligence-ai-bf3c0c60-en.htm.

On September 7, 2023, the G7 Digital and Tech Ministers released a statement[8] following a virtual meeting building on the Hiroshima AI process. The statement endorsed the development of a policy framework and international guiding principles for AI actors, a code of conduct for organizations developing advanced AI systems, and "project-based cooperation in support of the development of responsible AI tools and best practices."

As an outcome of the Hiroshima process, the G7 leaders agreed on a list of international guiding principles[9] and a voluntary Code of Conduct for organizations developing AI systems in October 2023. Both the guiding principles and Code of Conduct[10] are living documents building on the OECD AI Principles.

The risk-based list of guiding principles includes risk mitigation across the AI lifecycle; vulnerability monitoring; transparency reports; information-sharing and incident reporting among AI developers; risk-based AI governance and risk management policies; security controls; content authentication mechanisms, i.e., watermarking; research on societal risks; prioritizing AI systems that address global challenges, in support of the United Nations Sustainable Development Goals (SDGs); advancing international technical standards; and implementing appropriate data protection measures.

The corresponding Code of Conduct calls on organizations in academia, civil society, the private sector, and the public sector to abide by actions based on the 11-point list of guiding principles.

The G7 leaders also asked ministers to develop the Hiroshima AI Process Comprehensive Policy Framework and work plan for advancing the Hiroshima AI Process by the end of 2023. [11]

[8] G7 Hiroshima AI Process: G7 Digital & Tech Ministers' Statement." *University of Toronto*, 7 Sept. 2023, www.g8.utoronto.ca/ict/2023-statement.html.

[9] "Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System." *G7 2023 Hiroshima Summit*, www.mofa.go.jp/files/100573471.pdf.

[10] "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems." *G7 2023 Hiroshima Summit*, www.mofa.go.jp/files/100573473.pdf.

[11] "G7 Leaders' Statement on the Hiroshima AI Process." *G7 2023 Hiroshima Summit*, 30 Oct. 2023, www.mofa.go.jp/files/100573466.pdf.

THE UNITED STATES (US)

In October 2023, the White House released an Executive Order (EO) on AI, which includes new standards for AI safety and security, privacy and consumer protections, equity and civil right considerations, workforce development, innovation and competition drivers, and responsible government use of AI.

It includes many cybersecurity-specific references throughout, including "enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks" and "creat[ing] guidance and benchmarks for evaluating and auditing AI capabilities," with a focus on cybersecurity capabilities through which AI could cause harm.[12]

In November 2023, the US Department of Homeland Security's (DHS) Cybersecurity and Infrastructure Security Agency (CISA) and the United Kingdom's National Cyber Security Centre (NCSC) released nonbinding guidelines[13] for providers to deploy 'secure by design' AI systems. Sixteen countries signed on to the guidelines along with the US and UK. The cybersecurity guidelines apply to four stages of the AI development lifecycle—design, development, deployment, and operation and maintenance—and are intended to produce AI that is functional, available, and protects sensitive data. Secure design covers threat modeling and understanding risks and trade-offs. Secure development and deployment involves securing supply chains and infrastructure, protecting assets and AI models, and developing procedures for technical debt and incident management. Secure operation and maintenance applies to logging and monitoring, update management, and information sharing.

The AI EO was preceded by the Blueprint for AI Rights, a non-binding guide to guide policy and practice on the responsible use of AI published by the White House Office of Science

[12] "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." *The White House*, 30 Oct. 2023, www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[13] "DHS CISA and UK NCSC Release Joint Guidelines for Secure AI System Development." *U.S. Cybersecurity and Infrastructure Security Agency*, 26 Nov. 2023, www.cisa.gov/news-events/news/dhs-cisa-and-uk-ncsc-release-joint-guidelines-secure-ai-system-development.

and Technology Policy in October 2022 (prior to ChatGPT's launch).[14] The Blueprint set forth five guiding principles: safe and effective systems; algorithmic discrimination protections; data privacy; notice and explanation; and human alternatives, consideration, and fallback.

Additionally, the US Department of Commerce's National Institute of Standards and Technology (NIST) released an AI Risk Management Framework (AI RMF) for organizations designing or deploying AI systems, intended to promote the development of "trustworthy and responsible" AI.[15] NIST launched a Public Working Group on GenAI in July 2023 to build on the success of the AI RMF and develop guidance on the specific risks of GenAI. The working group will release a cross-sector risk management profile on GenAI for public review in early 2024.[16]

THE UNITED KINGDOM (UK)

The UK approach to regulation is motivated by the country's ambition to "become a science and technology superpower by 2030," as set forth in a white paper published in March 2023 titled "A pro-innovation approach to AI regulation."[17] Rather than introducing "rigid and onerous" legislation, the UK government opted to create a principles-based regulatory framework with "proportionate" rules for different sectors' use of AI.

Of note, the UK has not created a new AI regulatory body, but instead diffused responsibility across the existing regulators, including the UK Information Commissioner's Office (ICO) and the Medicines and Healthcare products Regulatory Agency (MHRA).

[14] "Blueprint for an AI Bill of Rights." *The White House*, Oct. 2022, www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[15] "AI Risk Management Framework." *U.S. National Institute of Standards and Technology*, 26 Jan. 2023, airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF.

[16] "NIST AI Public Working Groups." *U.S. National Institute of Standards and Technology*, airc.nist.gov/generative_ai_wg.

[17] "A Pro-innovation Approach to AI Regulation." *Gov.UK*, 3 Aug. 2023, www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.

The framework aims to:

- Create a unified definition of AI to support regulation
- Adopt a "context-specific approach"
- Provide a set of five cross-sector principles (with application at the discretion of regulators), including: safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress; and
- Design new central government functions to support regulators over the following 12 months, i.e., via investing in the AI Standards Hub,[18] an initiative created in 2022 to improve AI standards adoption and development in the UK; and by developing an AI regulatory sandbox[19]

Building on this strategy, the UK government hosted a Global Summit on Artificial Intelligence Safety in November 2023. 150 leading experts contributed to conversations on two categories: risks and potential actions. On the risk side, conversations focused on global safety, misuse, unpredictable advances, loss of control, and the integration of AI into society. The other conversations explored what AI developers, national policymakers, and the international and scientific community should do about these risks and opportunities. This convening culminated in 28 countries signing on to the Bletchley Declaration[20], which encourages "context-appropriate transparency and accountability on their plans to measure, monitor and mitigate potentially harmful capabilities and the associated effects that may emerge, in particular to prevent misuse and issues of control, and the amplification of other risks." The Declaration sets forth a two-fold agenda, first to develop a shared understanding of AI risks and second to build risk-based policies including transparency requirements, evaluation metrics, safety testing tools, and public sector scientific research. Signatories will meet again in furtherance of this agenda in 2024.

[18] "About the AI Standards Hub." *AI Standard Hub*, aistandardshub.org/the-ai-standards-hub/.

[19] "A Pro-innovation Approach to AI Regulation, Section 334." *Gov.UK*, 3 Aug. 2023, www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#section334.

[20] "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023." Gov.UK, 1 Nov. 2023, www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

CHINA

China published its long-term plan on AI in 2017, entitled the "New/Next Generation Artificial Intelligence Development Plan" (2017–2030).[21] The first half of the document lays out China's plans and ambitions. The second lays out what it has done and aims to achieve in technology, binding AI to political ambition in increasing China's economic growth and setting out three main objectives:

1. To make China a global leader in AI research and development by 2030.
2. To promote the application of AI in key economic and social sectors, such as manufacturing, healthcare, and transportation.
3. To develop a robust and ethical AI governance framework.

Of note, the plan lists "*Enhance AI civil-military integration*" before "*Build safe and efficient AI infrastructure system.*" The approach does recognize the need to set up AI safety regulation or assessment systems.

In 2023, China pursued rules-based measures for regulating GenAI[22], which deal with both legal transgressions of accuracy and copyright and pertain to nearly every part of the Generative AI lifecycle, including permissible inputs, algorithm transparency, licensing, and more. Of note, the regulations state "[c]ontent generated through the use of GenAI shall reflect the Socialist Core Values, and may not contain: subversion of state power; overturning of the socialist system; incitement of separatism; harm to national unity; propagation of terrorism or extremism; propagation of ethnic hatred or ethnic discrimination; violent, obscene, or sexual information; false information; as well as content that may upset economic order or social order." This lays bare the political priorities (and perhaps fears) in the development of this technology.

[21] Full Translation: China's 'New Generation Artificial Intelligence Development Plan' (2017)." *DigiChina, Stanford University*, 1 Aug. 2017, digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.

[22] "Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment) – April 2023." *DigiChina, Standard University*, 12 Apr. 2023, digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/.

China has a forward-leaning approach to develop AI and integrate it into all areas of civil society, with the clear goal of making China the leader in the field. Recent developments from Baidu, a leading Chinese AI company, indicate that Chinese AI developments are indeed progressing at pace.[23]

SINGAPORE

Singapore has taken a measured approach to the global race to regulation, with government officials confirming "we are currently not looking at regulating AI" as of July 2023.[24] In lieu of formal regulation, Singapore is advocating for responsible AI measures and testing and guidance for individuals and enterprises, building on strategy planning conducted prior to widespread AI access in late 2022, which include the following pillars:

- The Model AI Governance Framework (Model Framework).
- The Advisory Council on the Ethical Use of AI and Data (Advisory Council).
- The Research Programme on the Governance of AI and Data Use (Research Program).[25]

The government has focused on researching and understanding AI development and applications through cross-sector partnerships, including launching AI Verify, a governance and testing toolkit.

[23] Marr, Bernard. "China's AI Landscape: Baidu's Generative AI Innovations In Art And Search." *Forbes*, 27 Sept. 2023, www.forbes.com/sites/bernardmarr/2023/09/27/chinas-ai-landscape-baidus-generative-ai-innovations-in-art-and-search/?sh=58991e4e419a.

[24] Chiang, Sheila. "Singapore Is Not Looking to Regulate A.I. Just yet, Says the City-state'S Authority." *CNBC*, 19 Jun. 2023, www.cnbc.com/2023/06/19/singapore-is-not-looking-to-regulate-ai-just-yet-says-the-city-state.html.

[25] Thong, Josh Lee Kok. "AI Verify: Singapore's AI Governance Testing Initiative Explained." *Future of Privacy Forum*, 6 Jun. 2023, fpf.org/blog/ai-verify-singapores-ai-governance-testing-initiative-explained/.

JAPAN

As of December 2023, Japan is still in the process of developing its own domestic policy on GenAI informed by the G7 guiding principles and code of conduct. Still, while other models are focusing on risk-based or rules-based elements of AI regulation, Japan has focused its regulatory efforts on harnessing Generative AI's positive applications, including increasing efficiency and innovation. In addition to hosting the G7's Hiroshima AI process and sharing support for those goals, Japan has been working on a goals-based approach to advance economic development. From a cybersecurity lens, Japan has been focused on creating disclosure channels and developer feedback loops for any improper or incorrect uses or outputs.

AFRICAN UNION (AU)

Building on engagement from the AU's Information, Communications, and Technology ministers, the AU launched a working group chaired by Egypt with the objective of determining an AU AI strategy. From a cybersecurity perspective, the working group determined the need for AI governance as well as the protection and availability of data.

These goals are mirrored in principles set forth by the AU's Convention on Cyber Security and Personal Data Protection (the Malabo Convention), which came into effect in 2023 after 9 years and ratification by 15 countries. It set plans to examine the "use of artificial intelligence, measures to ensure proper resourcing for domestic data protection frameworks, and the establishment of regional bodies to monitor implementation."[26] More broadly it focuses on security principles for e-commerce, personal data protection, and cybersecurity and cybercrimes.[27]

[26] "Africa: AU'S Malabo Convention Set to Enter Force After Nine Years." *Data Protection Africa*, 19 May 2023, dataprotection.africa/malabo-convention-set-to-enter-force/.

[27] "African Union Convention on Cyber Security and Personal Data Protection." *African Union*, 27 Jun. 2014, au.int/sites/default/files/treaties/29560-treaty-0048_-_african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf.

GENERATING EFFECTIVE GOVERNMENT ACTION

SECTION SUMMARY

1. Start with the End User in Mind
2. Assess Criminal and Civil Liability
3. Consider Technology Safeguards and Feasibility
4. Establish Standards

START WITH THE END USER IN MIND

Before governments act, they need to have a clear endstate and objective, beyond mitigating risks or minimizing harms. If they do not know what conduct, outcomes, or values they are advancing for their citizens, their efforts are unlikely to be successful. As the adoption of GenAI tools by the general population moved AI out of technology boardrooms, computer labs, universities, and government halls, it began to resonate deeply in our daily lives as consumers, workers, and citizens. On one side, it empowers individuals with the ability to create, whether it's crafting code, composing music, generating videos, or weaving intricate text, all with significantly lower skill prerequisites. Yet, on the other side is the specter of AI voice cloning,[28] misinformation, AI "hallucinations," and increased cybersecurity risks and criminal adoption.

Before governments act, they need to have a clear endstate and objective, beyond mitigating risks or minimizing harms.

[28] "Klobuchar Fighting AI Voice Cloning." *Amy Klobuchar, United States Senator*, 7 Nov. 2023, www.klobuchar.senate.gov/public/index.cfm/2023/11/klobuchar-fighting-ai-voice-cloning.

Government policies should focus on the areas where self-regulation by industry is likely to inflict harm on the individual. The prevailing narrative surrounding AI often centers on industry and governmental concerns, such as assessing its macroeconomic impact on the labor market or the necessity of regulations, licensing, and risk frameworks. These discussions are undoubtedly important, but they can overlook the immediate needs of and risks to individuals. The reality is that most individuals lack the means to adequately protect themselves. Individual education, both for users of GenAI and the general public, operates with a lag and struggles to keep pace with the breakneck speed of technological advancements.

Government policies should focus on the areas where self-regulation by industry is likely to inflict harm on the individual.

Therefore, as governments consider how to oversee and regulate generative AI, they are best suited by following the old adage "start with the end in mind." What outcome do they want for individuals? What actions do they want to influence? What conduct do they want to encourage or dissuade? This approach to AI safety will emphasize design that sees end-users as not just sophisticated enterprises, but as everyday citizens, consumers, and individuals who employ technology for creative purposes in their personal and professional lives. This individual is the place where governance will have lasting impact, and regulators should ask whether the policies they are considering are constructed in a way that they will make a difference. Simply banning specific actions, products, or outcomes may look and feel like action but it is unlikely to have the desired positive impact.

This approach necessitates consultation, testing, and evaluation, which likely includes red teaming of AI models before they are released in the hands of the general public. In the absence of new laws, this places the responsibility on industry and governments equally. And while this process is underway (i.e., before formal regulations and directives are in place), governments and regulatory agencies must use what authorities they do have to

ensure that AI models do not cause immediate harm, for example by generating compelling impersonations for authentication purposes or expediting the production of malicious code.

The ease and efficiency that make GenAI popular with the general public applies equally to those who would use it for nefarious purposes.

ASSESS CRIMINAL AND CIVIL LIABILITY

Though only recently available to the public, bad actors are already working to exploit GenAI systems and pose dangers to public safety. AI-enhanced threats vary from data breaches to sexual abuse to terrorism. With such expansive potential for misuse of AI, there is no one-size-fits-all approach to protecting the public. Even now, decades into the professionalization of cybersecurity as a field, many countries do not have comprehensive cybersecurity laws and regulations. Thus, in addition to relying on existing civil and criminal codes or even updating those laws, regulators should raise public awareness to the threat as well as engage with developers to help shape their tools and define their responsibilities.

The ease and efficiency that makes GenAI popular with the general public applies equally to those who would use it for nefarious purposes. Attackers don't need to know how to code to use AI to generate ransomware and dangerous hacking tools. Schemers who might otherwise struggle with language barriers can now generate phishing text that can convincingly impersonate the people most trusted by their targets. Grooming, trafficking, and sexual abuse can all be facilitated by AI-generated fake profiles and believable chats. And terrorists and extremists can generate effective propaganda and incite misinformation for rapid, targeted recruiting.

In the civil context, AI-generated tools, text, and products could give rise to myriad claims, from copyright infringement to civil tort claims. Current laws were written without consideration of GenAI, and in many cases before it was even imagined. At minimum, governments should review current statutes to see if they need revision to account for these developments and the legal disputes that could come with them.

Current laws were written without consideration of GenAI, and in many cases before it was even imagined.

Though these threats cut across regulator industries and regulatory silos, law enforcement, litigators, and regulators will each face the same pressing questions. Foremost, if GenAI is used to facilitate or commit a cybercrime, breach a contract, or harm someone online, to what extent might the developers of the AI system be liable? At what point do the benefits of GenAI as a creative tool become overshadowed by the technology's ability to aid, abet, and conspire? In the criminal context, how should AI developers cooperate with law enforcement to gather evidence of the inputs, outputs, and algorithms used to generate content associated with this criminal activity?

At the core of these questions lies the decision between regulating the creation versus the use of GenAI. Regulators focusing on the former might ban particular AI capabilities, whereas regulators focusing on the latter might criminalize particular user inputs or outputs. The most successful approach will strike a balance

Part of the solution should be working directly with AI developers to strike a balance that promotes innovation without compromising global safety.

between the two. Every country has a different approach to law enforcement, both civil and criminal. But all should at a minimum take a hard look at what gaps GenAI is revealing and assess how, and even whether, to address them legislatively. Part of the solution should be working directly with AI developers to strike a balance that promotes innovation without compromising global safety, as has been the case with the cooperative efforts in the fight against ransomware.

Any regulatory and legal safeguards proposed must be flexible to keep up with technological advances.

Above all, it is essential to educate the general public on how GenAI can be misused or weaponized against them. GenAI can be a fun tool in the mainstream. But regulators must raise public awareness about the darker capabilities of AI or risk leaving their citizens vulnerable and exposed. Anti-phishing education campaigns have proven successful, if only by forcing fraudsters to develop new tactics, and may be illustrative in educating the public about GenAI-driven crime.

CONSIDER TECHNOLOGY SAFEGUARDS AND FEASIBILITY

The full uses and applications of AI will never be easily defined, since the possibilities for utilization increase as the technology develops. Therefore, any regulatory and legal safeguards proposed must be flexible to keep up with technological advances.

Since AI is already being utilized by individuals, companies, and governments around the globe, creating and enforcing regulatory and legal safeguards will inevitably lack complete uniformity. There is also danger of a bias towards more lenient, permissive, light-touch norms and regulation as countries that impose more restrictive safeguards will fear falling behind internationally. Other

challenges include the delays in creating and enforcing regulatory and legal safeguards. These are unavoidable, leaving only national and international voluntary compliance as an interim solution, posing further challenges for policymakers.

Many initiatives stipulate users should have the ability to opt out of engagement with AI systems and have access to a human alternative, where appropriate. The need for reasonable and appropriate AI safeguards is clear but determining, implementing, and enforcing those safeguards poses significant challenges with this rapidly-developing technology.

A few key concepts should be at the core of all safeguards being considered over and above data security:

1. The end-user should know when they are engaging with an AI system
2. Discrimination and bias must be minimized and eliminated if possible
3. Transparency and information-sharing concerning vulnerabilities, potential dangers, and inappropriate uses are critical
4. Human-controlled break points must be in place when AI is utilized in critical systems, such as the health and safety of humans, national security, or other critical matters.

ESTABLISH STANDARDS

Standards serve as the operational bedrock for AI, intricately weaving systems, processes, and tools into a cohesive fabric. Much like the ubiquitous Wi-Fi standard that effortlessly unites diverse devices worldwide, AI standards are poised to define the future landscape of innovation.

Given their emergence in late 2023, the US AI Executive Order and the EU AI Act are poised to shape the next era of standards. However, who truly holds the reins in this complex domain? In the quest for global coherence and cooperation, history reveals that industry stalwarts have often spearheaded this process.

Technical expertise, the lifeblood of specific standards, predominantly resides in the corridors of industry and academia. Yet, the narrative shifts when it comes to the socio-ethical dimensions of AI. Here, the stage broadens, and governments and civil society take center stage. Governments possess the authority to set standards and delineate rules, yet their jurisdictional boundaries threaten to fracture the international AI ecosystem. The remedy? Harmonization. The imperative lies in navigating complex global organizations like the International Standards Organization/ International Electrotechnical Commission (ISO/IEC) and the Institute of Electrical and Electronics Engineers (IEEE), which transcend geopolitical lines and could forge a unified AI framework. Industry organizations have historically led the charge, armed with talent, budgets, and resources to mold technical standards.

However, as the geopolitical landscape exerts its influence on AI standards, bridging the chasm between social and technical considerations becomes paramount. Dialogue, convenings, and harmonization emerge as the conduits that can ensure standards encapsulate the best interests of all stakeholders. As the curtain rises on the next act of AI standards, the critical actors are not just the architects of algorithms or policymakers, but the convergence of government, industry, and civil society. Only through their collaboration can the intricate choreography of AI standards unfold, navigating the delicate balance between technological prowess and societal well-being.

CAUTIONS AND CAVEATS FOR GOVERNMENT ACTION

SECTION SUMMARY

1. Creating Consent Fatigue
2. Mistaking Actions for Results
3. Ignoring the Openness of GenAI Tools

CREATING CONSENT FATIGUE

Transparency is a key component of trust. So while labeling norms and regulations are typically only followed by honest parties, they remain essential to trust and ideally would follow a standard format across the globe. A unified labeling scheme indicating the presence of AI-generated content will make misrepresentation easier to police. Lack of proper signposting can also serve as an element of civil or criminal penalty, especially in cases where the GenAI content is used for impersonating or authenticating as others. In this case, regulators might consider if fraud that utilizes an undeclared AI should be subject to enhanced sentencing, both for the crime itself and for the lack of disclosure.

Governments, companies, and organizations should not just consider the "how" of labeling, but also the "why."

With that said, governments, companies, and organizations should not just consider the "how" of labeling, but also the "why." Effective AI transparency will inform the end user when the use of AI is relevant to her needs, while ubiquitous labeling of every instance that potentially involves an AI could reduce even the best label to easy-to-ignore background noise. Take, for example, the numerous consent processes introduced to satisfy GDPR. While useful to some, to many they have become just one more meaningless click before reaching a desired page or app to many others. This is the now well-understood concept of "Consent Fatigue."[29]

In the case of AI and in the interests of transparency, users should always know if the 'entity' they are communicating with, whether a chat-bot, text, email, or other pathway, is a human-to-human or a human-to-AI generated exchange. This could be achieved by the designation of a suitable, universally accepted icon or emoji to represent AI to be made visible by browsers, email clients, and other applications whenever the exchange involves an AI. Users could also be presented with an explanation and other details needed in the event of complaint.

Whatever solution is settled upon, it is important to understand that end users are already presented with an avalanche of information beyond what they may actually be seeking. Governments and the private sector should develop a common approach to labeling AI content and a framework for determining when transparency is necessary and beneficial to an end user.

MISTAKING ACTIONS FOR RESULTS

Our collective focus should not be the speed at which governments regulate AI but the consideration they put into crafting regulation that is flexible enough to adapt with such a dynamic technology and grounded in what is already known about securing software.

Many overlook the fact that AI is software. It's the culmination of decades of best practices, frameworks, and international standards that help keep everyone more secure. Governments

[29] Borner, Peter. "Consent Fatigue." *The Data Privacy Group*, 13 Jun. 2022, thedataprivacygroup.com/blog/consent-fatigue/.

should leverage those existing tools to help us address the most pressing security concerns and give ourselves the appropriate time to think through the emerging risks, bringing the right mix of stakeholders—including government, industry, civil society, and independent security researchers—into these conversations to ensure we're thinking broadly enough about the new challenges we'll face.

The new challenges are complex, and the technology is evolving quickly. It can be tempting for organizations (government and private sector) to focus on what appear to be "easy wins" and avoid the tough elements of securing AI altogether. However, this strategy is only beneficial in the short-term. Simply "banning" malicious activity rarely works; theft is illegal around the world yet it persists. So while "banning" a particular misuse of AI may make for good headlines, it is unlikely to change conduct significantly or improve the lives of individuals. So how do we avoid this with AI? It will require time and a collaborative, thoughtful multistakeholder process assessing both what the future should be and how to shape digital security to get there.

Finally, governments should consider how bad actors will respond to regulatory action. The extent to which bad actors will care or even monitor regulation depends on their motivations, risk tolerance, technical capabilities, and the overall regulatory environment in which they are operating. Although regulation may deter some harmful behavior, many bad actors will find ways to circumvent regulation or develop new techniques to avoid detection. Countering them will require a multifaceted approach that includes legal measures, cybersecurity, international cooperation, and public education.

IGNORING THE OPENNESS OF GENERATIVE A.I. TOOL ACCESS

Powerful artificial intelligence technologies have a huge potential for upside as well as downside. Governments need to carefully consider how open these technologies can or should be to the general public.

Traditional wisdom in computer security is that security by obscurity does not work, and that systems built on open source are more secure than closed source systems. If anyone can read the code, bugs and vulnerabilities are found faster and anybody can propose a potential fix. However, in the regulatory realm, it is critical to consider thresholds for societal harms derived from open access to GenAI tools.

OpenAI—one of the best-known examples of AI companies—named their company after the core idea that they want to build their models out in the open and provide everyone access to the source code of their systems. When GPT turned out to be more powerful than they expected, OpenAI reversed course, and went from open source to fully closed. Today, almost all the frontier GenAI systems are closed, hosted in the cloud and unavailable for analysis by outsiders. You cannot download ChatGPT—you can only use it via the web. If OpenAI doesn't like what you're doing with ChatGPT, they can close your account and kick you out. The same thing applies to other GenAI systems such as Google Bard, Anthropic Claude, Inflection Pi, OpenAI Dall-E, Midjourney, and MusicLM.

There are notable exceptions like Stability (with the Stable Diffusion image-generation framework and Stable LM large language model) and GPT-J from EleutherAII. They were instantly abused; there are ports of Stable Diffusion which are solely used to create realistic AI pornography, often using the likeness of real people, and there are ports of GPT-J and LLAMA where restrictions and safety precautions have been removed, allowing the language model to freely write malware, craft phishing emails, or outline election influencing messaging.

Malware can be generated with any large language model that's capable of programming; the first piece of malware that used GPT to rewrite its code every time it replicated was found in April 2023. However, this malware (known as LLMorpher) requires access to OpenAI servers and needs a unique API key. OpenAI can easily blacklist such usage and there's nothing the attackers can do about it, since GPT is closed source.

The argument for more open AI systems can be made as well: the more people can study and interact with open-source foundation models, the more we learn about them, and the better we can improve them. Thus, the use of open source could be good for security.

Finally, as is the case with any powerful technology breakthrough, we would hope it's achieved by a responsible party, hopefully from a democratic nation. The more we publish powerful open-source models, the more we hand them over to governmental actors in non-democratic countries. Once we invent something, we can't uninvent it. We can only try to limit access to it or establish guidelines for its use. This applies to foundational models as well.

CONCLUSION

Governments and regulatory bodies will continue to consider GenAI's implications throughout 2024 and beyond. Effective governance and regulation will require finding a balance between hope and fear. While that balance is found, these technologies will continue to change our lives, both for better and for worse.

Meanwhile, industry leaders are grappling with how best to govern their own technologies, as they await government action. While effective and enforceable regulation of these technologies is underway, industry is working to adopt commitments and codes of conduct, to only varying degrees of success and enforcement. For example, the White House convened[30] seven leading AI companies and secured voluntary commitments[31] on principles of "safety, security, and trust." In September 2023,

[30] "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI." *The White House*, 21 Jul. 2023, www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[31] "Ensuring Safe, Secure, and Trustworthy AI." *The White House*, 21 Jul. 2023, www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

the White House secured a second round of commitments[32] from eight others. Some commitments include "internal and external security testing," "information sharing" on managing AI risks, and "invest[ing] in cybersecurity and insider threat safeguards." However, all commitments are voluntary and not legally binding.

The flurry of regulatory efforts in 2023 seeded decades of further conversation on potential and yet unfathomed cybersecurity harms, risks, and costs. However, it also presented a rare opportunity where GenAI had urgent attention from both countries and industries across the globe, allowing for a global discourse on what GenAI means for the future.

While policymakers face immediate challenges, average people will be the ones to pay the price most immediately when the governments and industry do not keep pace with AI's applications in cybersecurity. Voice and video generated by AI technologies will shake the assumptions of trust many of us enjoyed in the first three decades of increasingly digital connection on the internet. Fortunately, with the right principles and safeguards in place, these risks can be minimized.

[32] "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI." *The White House*, 12 Sept. 2023, www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

APPROACH AND ACKNOWLEDGMENTS

TIMELINE

This paper concept was developed in virtual meetings of the Generative AI working group from May to November 2023.

OBJECTIVES AND AUDIENCE

This paper sets out to provide guidance for governments as they consider whether and how to regulate, oversee, or otherwise address the explosion of generative AI tools in cybersecurity contexts. It focuses on the intersection of cybersecurity and generative AI, and is intended to assist international regulatory bodies and governments around the world.

CONTRIBUTORS

Jane Horvath (Co-chair)

Govind Shivkumar (Co-chair)

Francesca Bosco
John Carlin
Tod Eberle
Katherine Fang
Stew Garrick
Mikko Hypponen
Boon Hui Koo
Tzipi Livni
Chelsea Magnant
Vanessa Moody
Shinichi Yokohama

ASPEN DIGITAL TEAM

Jeff Greene
Katie Brooks
Devon Regal

ASPEN GLOBAL CYBERSECURITY GROUP MEMBERS January 2024

David Koh
Corey Thomas
Marina Kaljurand
Jorge Guajardo
Dmitri Alperovitch
Paul Ash
Carl Bildt
Christophe Blassiau
Cecilia Bonefeld-Dahl
Yasmin Brooks
Inge Bryan
Paolo Dal Cin
Oleh Derevianko
Anriette Esterhuysen
Tobias Feakin
Camille Francois
Stew Garrick
Katherine Getao
Dario Gil
Ron Green
Jane Horvath
Mikko Hypponen
Chris Inglis
Boon Hui Khoo
Tzipi Livni
Ciaran Martin
Chris Painter
Guillaume Poupard
Michelle Price
Greg Rattray
Latha Reddy
Runa Sandvik
Rob Strayer
Eli Sugarman
Yigal Unna
Phil Venables
Grant Verstandig
Alicia Wanless
Alberto Yepez
Shinichi Yokohama

Global Cybersecurity Group | ASPEN DIGITAL

aspen institute