BSA | The Software Alliance Comments on the Group of Seven Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems
November 9, 2023

BSA | The Software Alliance appreciates the Group of Seven's (G7) leadership on artificial intelligence (AI). The G7's Hiroshima AI process is an important vehicle for international collaboration on critical AI issues. As part of this effort, the G7 has released the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems (G7 Principles), which serve as a basis for the Hiroshima Process International Code of Conduct. We appreciate the goals of the G7 Principles. We provide below recommendations on enhancing the Principles to help ensure responsible AI development and deployment.

BSA is the leading advocate for the global software industry.[1] BSA members are at the forefront of developing cutting-edge services — including AI — and their products are used by businesses across every sector of the economy.[2] Our member companies provide tools including cloud storage, data processing, customer relationship management software, human resource management programs, identity management services, and collaboration software. BSA has worked on AI policy for more than six years, and our views are informed by our recent experience working with member companies to develop the BSA Framework to Build Trust in AI,[3] a risk management framework for mitigating the potential for unintended bias throughout an AI system's lifecycle. Built on a vast body of research and informed by the experience of leading AI developers, the BSA Framework outlines a lifecycle-based approach for performing impact assessments to identify risks of AI bias and highlights corresponding best practices for mitigating those risks.[4]

We support the objectives of the G7 Principles, which aim to promote safe, secure and trustworthy AI worldwide. We also appreciate that the G7 principles are intended to be a

[1] BSA's members include: Adobe, Alteryx, Asana, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, Databricks, DocuSign, Dropbox, Elastic, Graphisoft, IBM, Informatica, Juniper Networks, Kyndryl, MathWorks, Microsoft, Okta, Oracle, Palo Alto Networks, Prokon, PTC, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, TriNet, Twilio, Unity Technologies, Inc., Workday, Zendesk, and Zoom Video Communications, Inc.

[2] *See* BSA | The Software Alliance, Artificial Intelligence in Every Sector, *available at* https://www.bsa.org/files/policy-filings/06132022bsaaieverysector.pdf.

[3] *See* BSA | The Software Alliance, Confronting Bias: BSA's Framework to Build Trust in AI, *available at* https://www.bsa.org/reports/confronting-bias-bsas-framework-to-build-trust-in-ai.

[4] BSA has testified before the United States Congress and the European Parliament on the Framework. *See, e.g.,* Testimony of Victoria Espinel, Public Hearing on AI & Bias, Special Committee on Artificial Intelligence in a Digital Age, European Parliament, Nov. 30, 2021, *available at* https://www.europarl.europa.eu/cmsdata/244265/AIDA_Verbatim_30_November_2021_EN.pdf; Testimony of Victoria Espinel, The Need for Transparency in Artificial Intelligence, before the Senate Committee on Commerce, Science, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security, *available at* https://www.bsa.org/files/policy-filings/09122023aitestimonyoral.pdf.

"living document" that will be discussed and elaborated on over time. As you review and update the Principles, we recommend:

- Clarifying the scope of the Principles;
- Distinguishing between the different roles in the AI value chain such as developers and deployers in Principles 1 and 2;
- Promoting internal testing of advanced AI systems in Principle 1;
- Ensuring appropriate vulnerability reporting measures in Principles 2 and 4; and
- Removing Principle 11.

A. Clarifying the Scope of the Principles

We recommend clarifying the scope of the Principles. The G7 Principles state that they apply to "advanced AI systems," but do not define that term. We encourage you to update the Principles to explain this term, making clear that "advanced AI systems" only encompass the most capable models that pose a high risk of harm. This approach avoids placing obligations on AI systems that may be used in low-risk scenarios and instead focuses resources on areas that have the most significant impact on individuals. The G7 Principles should also use the term "advanced AI systems" consistently. For example, Principle 2 currently refers to AI systems more broadly, without clearly focusing on advanced AI systems.

B. Principle 1: Measures to Identify, Evaluate, and Mitigate Risks Across the AI Lifecycle

We recommend the G7 update Principle 1 to address two concerns.

First, Principle 1 should promote the use of internal testing and avoid suggesting external tests should always be conducted. As written, Principle 1 addresses the identification and mitigation of risks throughout the AI lifecycle, describing both internal and independent external testing as measures that organizations should perform. We agree that testing is a key part of identifying risks, but we advise against applying this Principle to suggest that organizations should always conduct external testing. There are circumstances where an organization may elect to perform external testing. However, internal testing — which can be performed by a team of employees that is independent from the team tasked with developing an AI system — can identify and mitigate risks without creating concerns about sharing trade secret and other proprietary information that will arise in external testing. As a result, we encourage the G7 to focus Principle 1 on internal testing and remove the reference to independent external testing.

Second, Principle 1 should be updated to reflect the different roles in the AI value chain such as that of developers of AI systems and deployers of AI systems. Developers are organizations that design, code, or produce an AI system, such as a software company that develops an AI system for speech recognition. In contrast, deployers are companies that use an AI system, such as a bank that uses an AI system to make loan determinations. The G7 Principles should recognize these different roles, because developers and deployers will each have access to different types of information and will be able to take different actions to mitigate risks. In its current form, Principle 1 can be read to assume that the developer of an AI system can identify, evaluate, and mitigate risks associated with that AI system — even after the AI system has been acquired and deployed by another organization. That is often not the case. Instead of assuming that all AI actors have access to information created at all stages of the AI lifecycle, the Principles should promote the identification, evaluation, and mitigation of risks by different organizations based on their role in that lifecycle. The relevant responsibility and accountability should be assigned to the most appropriate role based on knowledge, control, and position in the AI value chain that makes it possible to address specific risks. This approach will promote better risk management, focused on the different types of information each organization has access to and its ability to take different risk mitigation measures.

C. Principle 2: Identify and Mitigate Vulnerabilities and Misuse After Deployment

We recommend the G7 address two concerns with Principle 2.

First, Principle 2 should be updated to reflect the different roles of developers of AI systems and deployers of AI systems. Like Principle 1, this Principle appears to assume that all actors in the AI value chain have access to all information about an AI system throughout its lifecycle. Because Principle 2 focuses on vulnerabilities occurring *after* deployment, we strongly recommend that it be revised to apply to the appropriate role – the deployer using the AI system – and not impose such responsibilities on other roles that are not in a position to address the concern, such as the developer of that system.

The importance of distinguishing obligations for different roles such as developers and deployers becomes clear in considering the types of information available to organizations in each role. For example, the developer of an AI system is well positioned to describe features of the data used to train that system, the system's known limitations, and its intended uses, but generally will not have insight into how the system is used after it is acquired by another organization and deployed. In contrast, a deployer is well positioned to understand how the system is actually being used, what type of human oversight is in place, and whether there are complaints about how the system works in practice.[5] Creating role-based obligations is not unique to AI; role-based responsibilities are considered best practice in privacy and security legislation worldwide.

Second, Principle 2 should clarify its approach to vulnerabilities. Principle 2 focuses on identifying and mitigating vulnerabilities and, where appropriate, incidents and patterns of misuse. It also refers to other stakeholders in connection with these efforts. Importantly, vulnerability reporting should be handled confidentially, as it may otherwise interfere with customer contractual agreements or raise concerns about proprietary information. Further, other security implications should be considered when addressing vulnerabilities. It is a prevailing practice in the industry that a vulnerability is not publicly disclosed until a patch or other mitigation measures are in place to limit further harm. Laws and policies related to vulnerability reporting should be risk-based and in line with internationally recognized standards and best practices. We encourage the G7 to recognize the importance of confidentiality in responding to these security incidents.

D. Principle 3: Public Reporting of Advanced AI Systems' Capabilities, Limitations, and Domains of Appropriate and Inappropriate Use

We support Principle 3. We also recommend the G7 further clarify how the transparency responsibilities in Principle 3 should be allocated among the different organizations that play different roles in the AI value chain.

Principle 3 calls for the public reporting of key information about advanced AI systems, including capabilities, limitations, and domains for appropriate or inappropriate uses. We

[5] *See* BSA, AI Developers and Deployers: An Important Distinction, *available at* https://www.bsa.org/files/policy-filings/03162023aidevdep.pdf.

recognize that developers of AI systems are creating a range of new resources to provide transparency to their customers about those AI systems, such as documentation that provides information on responsible AI design choices, as well as best practices for deploying and optimizing the performance of a particular AI service. The G7 should support efforts to provide deployers with this type of information, while avoiding requirements to disclose underlying training data or other information for which disclosure would create trade secret, confidentiality, and privacy concerns.

E. Principle 4: Work Toward Responsible Information Sharing and Reporting of Incidents

We recommend clarifying aspects of the vulnerability reporting referred to in Principle 4. Vulnerability reporting and incident response are key components of an effective security program. The public incident reporting recommended in Principle 4 could interfere with customer contractual agreements and measures to safely address vulnerabilities. As discussed above, a company should generally not report a vulnerability until it has developed a patch or implemented other mitigation measures, and laws and policies related to vulnerability reporting should be risk-based and in line with internationally recognized standards and best practices. We encourage the G7 to recognize the importance of confidentiality in responding to these security incidents.

F. Principle 5: Develop, Implement, and Disclose AI Governance and Risk Management Policies, Grounded in a Risk-Based Approach

We support Principle 5, which recognizes the importance of risk management policies and practices to enhancing organizational accountability and ensuring responsible AI.

As Principle 5 recognizes, organizations should develop and implement risk management programs to help them evaluate and mitigate risks throughout the AI lifecycle. We encourage the G7 to recognize that a key part of an effective risk management program is conducting impact assessments. Impact assessments enable organizations to identify and mitigate risks and should be conducted by developers and deployers for high-risk uses of AI systems. By allowing personnel across the organization to examine the objectives, data preparation, design choices, and testing results, these assessments help refine AI products and services and drive internal changes to an organization's risk management program. Implementing these changes enables organizations to better address existing concerns and adapt to new risks as they emerge.

Principle 5 also refers to disclosing AI governance policies and organizational mechanisms to implement policies. As you further consider this principle, we encourage the G7 to recognize that impact assessments should be treated as confidential to preserve the incentives for organizations to implement them through rigorous processes that identify and mitigate a wide range of potential risks. The fact that assessments are being performed for high-risk uses of AI systems promotes trust for external stakeholders because they will know that an organization is conducting a thorough examination of AI systems; those assessments should also be available to regulators in the course of an investigation, under existing domestic laws. We support Principle 5's aim of ensuring the implementation of risk management policies, and we encourage the G7 to refer to impact assessments as an important accountability tool that can help achieve this goal.

G. Principle 6: Invest in and Implement Robust Security Controls

We support Principle 6. The responsibility in this Principle to provide robust security controls, including physical security measures, is critical. Security controls, including limiting the employees who have access to data, implementing zero trust architecture where appropriate, and monitoring networks for malicious activity, are essential. Providing robust cybersecurity protections is a priority for BSA members. An AI governance program should effectively address security risks. Notably, organizations are using AI to improve cybersecurity, including developing more secure code, detecting and responding to malicious threats, protecting against malware, and improving identity management.[6]

H. Principle 7: Develop and Deploy Reliable Content Authentication and Provenance Mechanisms

We support Principle 7. The development and deployment of reliable content authentication and provenance mechanisms (e.g., watermarking) that can help users identify AI-generated content is an important focus of AI policies. Any content provenance requirements for AI-generated content should focus on images, audio, and video content, since it is unlikely that tools developed for labeling image and audio-visual content would be effective for text. To ensure provenance of text-based AI generated content, we recommend focusing on other transparency mechanisms that can help individuals know when they interact with an AI system.

We encourage the G7 to build on work by organizations including the Content Authentication Initiative and Coalition for Content Provenance and Authenticity, which promote the adoption of an open industry standard for content authenticity and provenance. This work can enable viewers to identify the origins of an image or video, such as the photographer, the location where the image was generated, and if it was edited using software, assisting viewers in determining the content's authenticity.

I. Principle 8: Prioritize Research to Mitigate Societal, Safety, and Security Risks and Prioritize Investment in Effective Mitigation Measures

We support Principle 8's focus on research and investment in risk mitigation measures. Investment in research is critical to continued development of AI. It can help inform many areas, including risk measurement and mechanisms for addressing socio-technical concerns. Notably, governments play an important role in research and development, including through providing investments, open government datasets, and shared computing resources. We encourage the G7's support of public-private research partnerships, as well as international collaboration on research initiatives.

J. Principle 9: Prioritize the Development of Advanced AI Systems to Address the World's Greatest Challenges

We support Principle 9. AI is helping solve the world's most complex challenges. For example, AI is being used to help detect medical diseases and transform patient care. It is also being used in education, helping teachers to access lessons and personalize learning. Addressing global challenges should be a priority for both the public and private sector.

[6] *See* AI for Cybersecurity Ensuring Cyber Defenders Can Leverage AI to Protect Customers and Citizens, *available at* https://www.bsa.org/files/policy-filings/20231004aiforcybersecurity.pdf.

Government policies on workforce development, research, and innovation are also important in tackling society's most pressing problems.

K. Principle 10: Advance the Development of International Technical Standards

We strongly support Principle 10's focus on international standards development. The G7 rightfully acknowledges in Principle 10 the importance of advancing the development and adoption of international technical standards on AI. It is important that we build on existing secure software development standards. Benchmarking should be performed to measure model performance against internationally recognized standards. The ongoing work of international standards organizations, such as the International Organization for Standardization (ISO), will help policymakers to create harmonized AI standards across jurisdictions. Aligning AI policies with internationally recognized standards will improve global interoperability of AI policies and promote the ability of organizations that operate across G7 member countries and beyond to benefit from the most advanced resources, concepts, and options available.

L. Principle 11: Implement Appropriate Data Input Measures and Protections for Personal Data and Intellectual Property

This Principle is unnecessary. Unlike the first 10 G7 Principles, which address system-level risks that are not covered in existing regulatory frameworks, Principle 11 implicates issues for which existing regulations are already in force. We agree about the importance of implementing appropriate safeguards regarding input data, but including a new AI principle implies organizations are not already required to maintain proper data governance. In addition, the description of Principle 11 refers to transparency of datasets, without recognizing that those datasets may be confidential and contain a range of proprietary information and, therefore, should not be disclosed.

* * *

We appreciate the G7's leadership on AI and look forward to a continuing dialogue about these important issues.