Cambridge Forum on AI: Law and Governance (2025), **1**, e15, 1–26
doi:10.1017/cfl.2024.5

Cambridge Forum

RESEARCH ARTICLE

Enhancements and next steps for the G7 Hiroshima AI Process: Toward a common framework to advance human rights, democracy and rule of law

Hiroki Habuka[1] and David U. Socol de la Osa[2]

[1]Graduate School of Law, Center for Interdisciplinary Studies of Law and Policy, Kyoto University, Kyoto, Japan and
[2]Hitotsubashi Institute for Advanced Study, Graduate School of Law, Hitotsubashi University, Tokyo, Japan
Corresponding author: David U. Socol de la Osa; Email: david.socoldelaosa@r.hit-u.ac.jp

(Received 19 August 2024; accepted 21 October 2024)

Abstract
This article focuses on the G7's Hiroshima AI Process (HAIP) and its flagship document, the Hiroshima Code of Conduct, as pivotal elements in shaping global artificial intelligence (AI) governance. By conducting a comprehensive analysis of AI regulations in G7 member states, the article demonstrates a high degree of interoperability between these national frameworks and the Code of Conduct's principles. The article proposes concrete steps to translate these principles into actionable policies at the G7 level and develops strategic adjustments to incorporate them into national standards. The article then proposes enhancements to the Code of Conduct, including the development of a common AI governance vocabulary, robust risk management frameworks, life cycle standards harmonization, effective stakeholder engagement mechanisms, specific redress mechanisms for AI harms and guidelines for government AI use to ensure democratic principles and human rights are upheld. Ultimately, this research aims to strengthen the G7's role in leading a global AI landscape characterized by the rule of law, democracy, and human rights.

Keywords: Hiroshima AI Process (HAIP); AI governance; G7; international cooperation; regulatory interoperability

1. Introduction

On May 2, 2024, Japanese Prime Minister Kishida Fumio announced the launch of the "Hiroshima AI Process Friends Group" at the Meeting of the Council at Ministerial Level of the Organisation for Economic Co-operation and Development (OECD).[1] This initiative, supported by 49 countries and regions – primarily OECD members – aims to foster international cooperation for ensuring global access to safe, secure, and trustworthy generative artificial intelligence (AI).[2]

The Hiroshima AI Process Friends Group has supported the implementation of international guidelines as stipulated in the Hiroshima AI Process Comprehensive Policy Framework

[1]Hiroshima AI Process, Supporters, Member countries of the Hiroshima AI Process Friends Group (in alphabetical order), Ministry of Internal Affairs and Communications (June 2024), https://www.soumu.go.jp/hiroshimaaiprocess/en/supporters.html; Ministry of Foreign Affairs of Japan, Prime Minister Kishida's attendance at the Side Event on Generative AI at the OECD Ministerial Council Meeting, (May 2, 2024), https://www.mofa.go.jp/ecm/ec/pageite_000001_00332.html; Japan's Kishida unveils a framework for global regulation of generative AI, Associated Press (May 3, 2024) https://apnews.com/article/oecd-ai-japan-kishida-artificial-intelligence-023ac08e04db5a2109cf35f8b8c9b102.

[2]Id.

(Comprehensive Framework).[3] Endorsed by the G7 Digital and Tech Ministers on December 1, 2023, the Comprehensive Framework represents the first policy package agreed upon by the democratic leaders of the G7 to effectively steward the principles of human-centered AI design, safeguard individual rights, and enhance trust-based systems throughout the AI lifecycle. This milestone sends a promising signal of international alignment on the responsible development of AI.[4] Notably, the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems (HCOC),[5] established as an integral part of the Comprehensive Framework, builds upon and aligns closely with existing policies across G7 members.[6]

The G7 has emphasized that the principles are living documents,[7] providing them with significant potential yet to be realized, as well as remarkable questions lying ahead: How does the Hiroshima AI Process (HAIP) contribute to achieving interoperability of international rules on advanced AI models? How can it add value beyond other international collaborations on AI governance?[8] How can the G7, as a democratic referent, leverage its position as a leading advocate for responsible AI to encourage broader adoption of its governance principles, even in regions with differing political or cultural contexts?

To answer these questions, this article (1) provides a brief overview of the history of AI governance and relevant instances of international cooperation; (2) analyzes the structure and content of the HAIP, with specific focus on the HCOC; (3) examines how the HCOC fits into the international tapestry of AI governance, particularly within the context of G7 nations, and how it can foster regulatory interoperability on advanced AI systems; and (4) identifies and discusses prospective areas of focus for the future development of the HCOC.

2. AI governance: A historical overview and international initiatives

2.1 A short history of AI governance

Following the deep-learning breakthroughs of the early 2010s, AI adoption surged across a myriad of industries and sectors (Brynjolfsson & McAfee, 2014; LeCun et al., 2015; Bharadiya et al., 2023). This rapid integration process brought to light a multitude of potential risks associated with deploying AI. From fatal accidents involving autonomous vehicles[9] to discriminatory hiring practices by AI algorithms (Andrews & Bucher, 2022), the real-world consequences of AI development have become

[3]The Comprehensive Framework consists of four elements: (i) The OECD's Report toward a G7 Common Understanding on Generative AI; (ii) the International Guiding Principles for All AI Actors and for Organizations Developing Advanced AI Systems; (iii) the International Code of Conduct for Organizations Developing Advanced AI Systems; and (iv) Project-based cooperation. See generally discussion infra Section 2(i). See also Table 2; The Group of Seven ("G7"), "Hiroshima AI Process G7 Digital & Tech Ministers' Statement (Dec. 1, 2023)," Available at: https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document02_en.pdf ("G7 Ministers' Statement").

[4]Id.

[5]G7, "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" (Oct. 30, 2023). Available at: https://www.mofa.go.jp/files/100573473.pdf.

[6]See discussion infra Section 3(i). See also Annex.

[7]See G7 Ministers' Statement, at II(3), III(5).

[8]AI Safety Summit, "The Bletchley Declaration," (Nov. 1, 2023). Available at: 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 ("The Bletchley Declaration"); "Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity: AI Seoul Summit 2024," (May 22, 2024). Available at: https://www.gov.uk/government/publications/seoul-ministerial-statement-for-advancing-ai-safety-innovation-and-inclusivity-ai-seoul-summit-2024.

[9]See United States Department of Transportation National Highway Traffic Safety Administration, "Summary Report: Standing General Order on Crash Reporting for Automated Driving Systems," (Jun. 2022). Available at: www.nhtsa.gov/sites/nhtsa.gov/files/2022-06/ADS-SGO-Report-June-2022.pdf; F. Siddiqui and J. Merrill, "17 fatalities, 736 crashes: The shocking toll of Tesla's Autopilot," (Jun. 10, 2023) The Washington Post. Available at: https://www.washingtonpost.com/technology/2023/06/10/tesla-autopilot-crashes-elon-musk/; Associated Press, "Cruise recalls all self-driving cars after grisly accident and California ban," (2023) The Guardian. Available at: https://www.theguardian.com/technology/2023/nov/08/cruise-recall-self-driving-cars-gm.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

increasingly evident. Furthermore, the manipulation of financial markets by algorithmic trading and the spread of misinformation on social media platforms (Ferrara, 2024) highlight the broader societal concerns surrounding the technology's integration across sectors.

Fueled by growing awareness of AI risks in the mid-and-late 2010s, national governments (including G7 members), international organizations, tech companies and nonprofits launched a wave of policy and principle publications. Prominent examples include the "Ethics Guidelines for Trustworthy AI" by the European Union (EU) in 2019,[10] the "Recommendation of the Council on Artificial Intelligence" by the OECD in 2019 (updated in 2024),[11] and the "Recommendation on the Ethics of Artificial Intelligence" by the United Nations Educational, Scientific and Cultural Organization (UNESCO) in 2021.[12] These publications emphasized pairing AI development with core values such as human rights, democracy and sustainability as well as key principles including fairness, privacy, safety, security, transparency and accountability.

While fundamental values and AI principles provide a crucial foundation to AI governance, translating them into implementable standards for AI systems remains a challenge, and addressing this challenge requires concrete and material guidance. Various initiatives have been undertaken at different levels to bridge this gap. At the national level, examples include the "AI Risk Management Framework"[13] (RMF) published by the National Institute of Standards and Technology (NIST) in the United States in January 2023, and Japan's "AI Guidelines for Business" published by several ministries in April 2024.[14] On a supranational scale, leading examples include the 2023 AI Safety Summit's "Emerging Processes for Frontier AI Safety"[15] and the G7's HCOC – the latter being the focus of this article. Additionally, nongovernmental organizations such as the International Organization for Standardization (ISO) have contributed by issuing international standards on AI governance. The "AI Management System Standard ISO/IEC 42001"[16] was published in December 2023, specifying AI management system requirements. Another notable contribution to the risk management and stakeholder engagement field is the "Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems" (HUDERAF), proposed by the Alan Turing Institute to the Council of Europe's Ad Hoc Committee on Artificial Intelligence (Leslie et al., 2022). Collectively, these diverse approaches underscore the ongoing efforts to transform abstract AI principles into a practical and implementable reality.

Despite this common direction, many published guidelines and principles for responsible AI development lack legally binding force, making them examples of "soft law." While compliance with these documents helps companies with risk prevention strategies and forward-looking accountability measures, there are no guarantees or enforceability measures to ensure adherence to these standards. Thus, to advance stronger commitment to AI governance – in particular, addressing AI systems

[10]European Commission, "Ethics guidelines for trustworthy AI," (Apr. 8, 2019). Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[11]OECD, "Recommendation of the Council on Artificial Intelligence," (May 3, 2024). Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[12]UNESCO, "Recommendation on the Ethics of Artificial Intelligence," (Nov. 23, 2021). Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[13]United States Department of Commerce (USDoC), National Institute of Standards and Technology (NIST), "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," (Jan. 2023). Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[14]Japan Ministry of Internal Affairs and Communications, Ministry of Economy, Trade and Industry, "AI Guidelines for Business Ver1.0" (Apr. 19, 2024) ('Japan Guidelines for Business' or 'AI Guidelines for Business'). Available at: www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf.

[15]Government of the United Kingdom ("UK") Department for Science, Innovation & Technology, "Emerging Processes for Frontier AI Safety," (Oct. 2023). Available at: https://assets.publishing.service.gov.uk/media/653aabbd80884d000df71bdc/emerging-processes-frontier-ai-safety.pdf.

[16]Joint Technical Committee International Organization for Standardization, International Electrotechnical Commission, "Information technology – Artificial intelligence – Management system: ISO/IEC 42001," (Dec. 2023). Available at: https://www.iso.org/standard/81230.html.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

that pose high risks – there has been active movement to introduce regulations with legally binding force. For instance, the European Commission introduced the draft AI Act in 2021 (subsequently published in the *Official Journal of the European Union* on July 12, 2024),[17] focusing most of its compliance requirements on high-risk systems and even banning certain systems when the risks they present are deemed unacceptable.[18] Similarly, in 2022, Canada presented a legislative proposal, the Artificial Intelligence and Data Act[19] (AIDA), which focuses on establishing compliance requirements for high-impact AI applications. The United States has also seen a surge in legislative activity targeted at AI. As of August, 2024, over 105 draft bills have been introduced addressing AI,[20] over 35 of which specifically target risk mitigation in AI applications.

The 2023 boom in foundation models presents a new layer of complexity to the already challenging landscape of AI governance. While conventional AI has faced issues such as limited explainability, diverse stakeholders and rapid evolution, foundation models expand the scope and reach of these challenges (Bommasani et al., 2021).[21] The application of these systems in countless contexts and their ease of operation create an even more intricate risk environment. As a result, there has been a surge in global efforts to establish rules and foster international cooperation around foundation models. The EU AI Act,[22] for example, has incorporated provisions specifically related to "general-purpose AI" systems.[23] Japan's Liberal Democratic Party proposed the concept note for the Basic Law for the Promotion of Responsible AI in February 2024,[24] which targets advanced foundational AI models with significant societal impact. Similarly, the Chinese government implemented the Interim Measures for the Administration of Generative Artificial Intelligence Services[25] in August 2023, establishing specific requirements for models with "public opinion properties or the capacity for social mobilization."[26] Figure 1 shows the overall structure of AI governance and key documents related to each layer of governance.

The brief history of AI governance is characterized by a complex and multidimensional balancing act between innovation and regulation, rapidly advancing technology, and the integration of multivector interests – encompassing the technology industry, the general public and regulators. While these groups may have differing priorities, there is also growing recognition of the need for collaboration. Responses to AI risks have evolved: Nations and international bodies initially relied on soft-law principles and public–private collaborative efforts, whereas the current momentum is toward binding legislative action, with specific measures addressing advanced AI. Another crucial distinction is the regulatory scope, which can be generally categorized as comprehensive or sectorial. While

[17]See European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) ("EU AI Act" or "AI Act"). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.

[18]See EU AI Act at Chapter II.

[19]Government of Canada, "The Artificial Intelligence and Data Act (AIDA) – Companion document," (2023). Available at: https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document.

[20]American Action Forum, "AI Legislation Tracker," (last accessed Aug. 10, 2024). Available at: https://www.americanactionforum.org/list-of-proposed-ai-bills-table/ ("AAF AI Legislation Tracker").

[21]See generally discussion supra Section 1. See also Schneider et al. (2024) and Myers et al. (2024)

[22]See EU AI Act.

[23]European Commission, "Artificial Intelligence Act," (Mar. 13, 2024). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf. (Regulation – EU – 2024/1689 – EN – EUR-Lex.)

[24]Liberal Democratic Party of Japan, "Basic Law for the Promotion of Responsible AI," (Feb. 2024). Available at: https://note.com/api/v2/attachments/download/006badee3e4d847b3a0c92358b2de63a.

[25]Cyberspace Administration of China, National Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, and State Administration of Radio, "Interim Measures for the Management of Generative Artificial Intelligence Services," (Jul. 2023). Available at: www.cac.gov.cn/2023-07/13/c_1690898327029107.html.

[26]Id.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

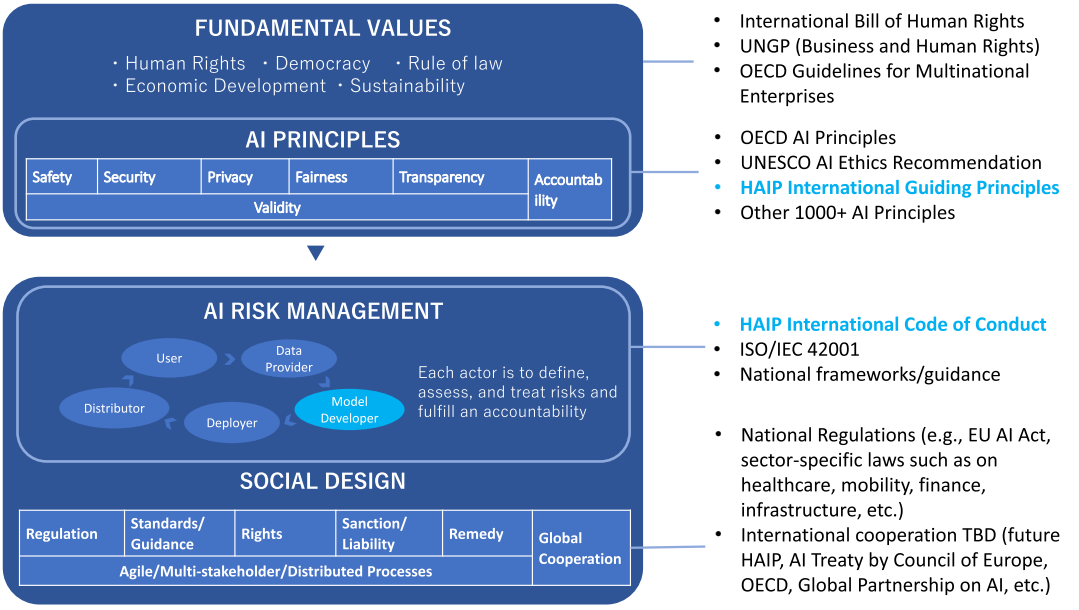

Figure 1. Overall structure of AI governance and key documents related to each layer.

the EU's AI Act and Canada's proposed AIDA encompass regulations that span across industries, Japan, the United Kingdom (UK) and the United States have indicated a policy direction that considers industry-specific AI regulations or focuses on powerful foundational models. Nonetheless, the regulatory emphasis in all of these instances is primarily on high-risk AI, aiming to strike an appropriate stability between fostering technological development and ensuring safety. G7 democracies, in particular, find common ground in core principles such as human rights and democratic values, grounding them in transparency, explainability and bias prevention and forming a common foundation for responsible AI development.[27]

2.2 Advancing international collaboration

This following subsection first (1) provides an overview of key international AI governance initiatives, including significant documents and declarations such as the G7's HAIP Comprehensive Framework, the Bletchley Declaration, the UN's "Governing AI for Humanity" report, and the Council of Europe's AI Treaty. These documents highlight various efforts to establish global standards and frameworks for AI governance. Subsequently, (2) the discussion will examine the pivotal role of the G7's framework in shaping global AI governance. The G7's role in global AI policies is underscored by its active participation in major international initiatives and its significant economic, regulatory, and technological impact.

2.2.1 Key international AI initiatives

As countries make progress with AI rulemaking within their borders, international cooperation is also advancing.[28] The G7 is one of the most impactful forums for such international coordination. During the May 2023 summit, G7 leaders committed to establishing the HAIP by the end of the year

[27] See discussion infra Section 1(ii).
[28] See discussion supra Section 1(i).

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

Table 1. Elements of the Hiroshima Process International Guiding Principles

Risk Management and Governance	Stakeholder Engagement	Ethical and Societal Considerations
1. Risk identification and mitigation 2. Vulnerability and misuse management after deployment 5. Governance and risk management 6. Security investments 7. Content authentication 11. Data quality, personal data and intellectual property protection	3. Transparency and accountability 4. Responsible information sharing 12. Trustworthy and responsible use of advanced AI (not included in the HCOC)	8. Research prioritization for societal safety 9. AI for global challenges 10. Development of international technical standards

Note: The numerals listed for each item correspond to the articles assigned in the HIGP and HCOC. The authors devised the abbreviations for the principles and their categorization.

to foster collaborative policy development on generative AI.[29] Within 6 months, the G7 digital and tech ministers had delivered the Comprehensive Framework.[30] This framework prioritizes proactive risk management and governance, transparency and accountability across the AI life cycle.[31] Additionally, it emphasizes anchoring AI development in human rights and democratic values while fostering the use of advanced AI for tackling global challenges such as climate change, health care, and education.[32]

In November 2023, the **AI Safety Summit** held in the UK produced the "Bletchley Declaration," a significant milestone in international AI collaboration.[33] The declaration addresses crucial aspects of AI governance, such as the protection of human rights, transparency, explainability, fairness, accountability, human oversight, bias mitigation, and privacy and data protection.[34] Additionally, it highlights the risks associated with manipulating or generating deceptive content.[35] Endorsed by 29 countries and regions, the signatories encompass not only G7 and OECD nations but also partners from the Middle East, Africa, South America, Asia, and, notably, China.[36] A second AI Safety Summit was held in Seoul in May 2024,[37] which reiterated the

[29]OECD, "G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI," (Sept. 7, 2023) OECD Publishing, at 6. Available at: https://doi.org/10.1787/bf3c0c60-en; The Government of Japan, "The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI" (Feb. 9, 2024) Kizuna. Available at: https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html.

[30]The Comprehensive Framework consists of four elements: (i) The OECD's Report toward a G7 Common Understanding on Generative AI; (ii) the International Guiding Principles for All AI Actors and for Organizations Developing Advanced AI Systems; (iii) the International Code of Conduct for Organizations Developing Advanced AI Systems; and (iv) Project-based cooperation. See generally discussion infra Section 2(i). See also Table 1; The Group of Seven ("G7"), "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems," (Oct. 30, 2023). Available at: https://www.mofa.go.jp/files/100573473.pdf.

[31]Id. See also G7 (Oct. 30, 2023), "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" at arts. 1-7, 11.

[32]See sources cited supra note 30. See also G7 (Oct. 30, 2023), "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" at arts. 8–10.

[33]See AI Safety Summit (2023), "The Bletchley Declaration."

[34]Id.

[35]Id.

[36]Id.

[37]See UK, "About the AI Seoul Summit 2024," (last accessed Aug. 9, 2024). Available at: https://www.gov.uk/government/topical-events/ai-seoul-summit-2024/about; Department for Science, Innovation and Technology, "Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity: AI Seoul Summit 2024," (May 22, 2024). Available at: https://www.gov.uk/government/publications/seoul-ministerial-statement-for-advancing-ai-safety-innovation-and-inclusivity-ai-seoul-summit-2024. See also Gregory C. Allen and Georgia Adamson, "The AI Seoul Summit," Center for Strategic & International Studies (May 23, 2024). Available at: https://www.csis.org/analysis/ai-seoul-summit; Jessica Birch, Öykü Özfırat, "Key Takeaways from the AI Seoul Summit 2024," Access Partnership (May 22, 2024). Available at:

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

anchoring point of safety, and highlighted inclusion and innovation as critical priorities for global convergence.[38]

The **United Nations** is also active in forming international AI governance principles. In December 2023, the UN AI Advisory Body issued the interim report "Governing AI for Humanity."[39] The report outlines a set of guiding principles and institutional roles designed to create a global AI governance framework, proposing essential considerations and actions to ensure that AI development and deployment serve the broader interests of humanity.[40] These include principles such as inclusivity,[41] public interest[42] and the importance of aligning AI governance with data governance and promoting a data commons.[43] Institutional functions highlighted in the report include assessing the future directions and implications of AI[44]; developing and harmonizing standards,[45] safety and RMFs[46]; and facilitating the development, deployment, and use of AI for economic and societal benefit through international multi-stakeholder cooperation.[47]

In March 2024, the **Council of Europe's Ad Hoc Committee on Artificial Intelligence** introduced the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (AI Treaty), a groundbreaking treaty on AI governance that sets a high bar on responsible AI development.[48] The AI Treaty, adopted in May 2024,[49] emphasizes the obligation of signatory nations (parties to the convention) to proactively ensure AI activities are aligned with human rights, democratic integrity and the rule of law. The treaty calls for comprehensive safeguards throughout the AI life cycle – including mechanisms for accountability[50] and transparency[51] – and introduces comprehensive RMFs.[52] Furthermore, it calls for robust remedies and procedural protective measures against rights violations,[53] promotes rigorous risk and impact assessments,[54] and delineates duties for international cooperation and implementation, focusing on nondiscrimination and rights protection.[55]

Nations participating in these initiatives vary. Figure 2 maps the structural involvement of various jurisdictions in the abovementioned international processes.

---

https://accesspartnership.com/key-takeaways-from-the-ai-seoul-summit-2024/; Ministry of Foreign Affairs of Japan, "Seoul Declaration for Safe, Innovative and Inclusive AI by Participants Attending the Leaders Session of the AI Seoul Summit," (May 21, 2024). Available at: https://www.mofa.go.jp/files/100672534.pdf. The event maintained the same attendee list as the previous summit, with China invited to the ministerial meetings. Although present for the discussions, China declined to become signatory to the final document, the "Seoul Ministerial Statement."

[38] Id.

[39] United Nations AI Advisory Body, "Governing AI for Humanity," (Dec. 2023). Available at: https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf.

[40] See UN Secretary-General's AI Advisory Body (2023), "Interim Report on Governing AI for Humanity."

[41] Id. at pp. 13.

[42] Id.

[43] Id. at pp. 14.

[44] Id. at pp. 15.

[45] Id. at pp. 16.

[46] Id.

[47] Id. at pp. 17.

[48] Council of Europe Ad Hoc Committee on Artificial Intelligence, "Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law," (Mar. 2024). Available at: https://rm.coe.int/1680afae3c.

[49] Council of Europe, "Committee on Artificial Intelligence (CAI)," (last accessed on Aug. 8, 2024). Available at: https://www.coe.int/en/web/artificial-intelligence/cai.

[50] Id. Article 8.

[51] Id. Article 7.

[52] Id. Article 16.

[53] Id. Article 15.

[54] Id. Article 16.

[55] Id. Article 25.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

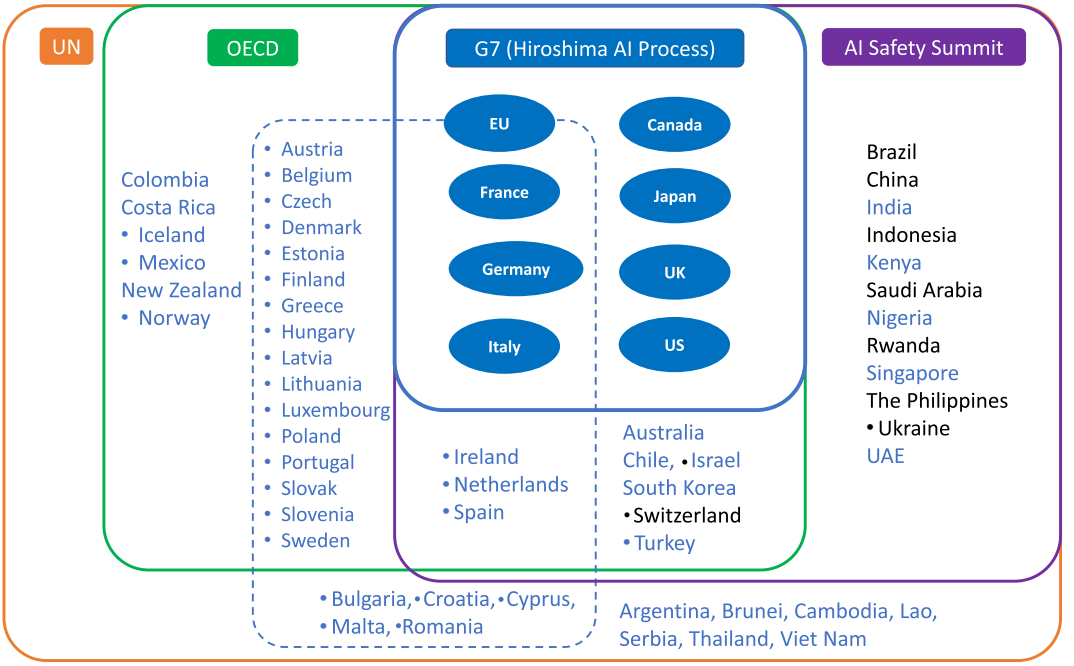

Figure 2. The global AI governance landscape.
Note: The European Union (EU) is considered a non-enumerated member of the G7. Both the European Council and European Commission presidents participate in the leaders' summit. The countries shown in blue indicate those participating in the Hiroshima AI Process Friends Group (as of December 31, 2024). The nations with dots (•), plus the G7 and EU members, are members or observers of the Council of Europe, the host organization of the AI Treaty.

2.2.2 The importance of the G7's AI governance framework

Figure 2 shows why and how the G7 HAIP has significance in global rulemaking on advanced AI systems. First, the G7 nations participate in all significant initiatives mentioned previously – namely the AI Safety Summit, the UN AI Advisory Body and the AI Treaty. Second, the G7 represents a group of nations with significant economic, regulatory and technological impact and leadership. In 2023, the GDP of the G7 countries (excluding the EU, which is a nonenumerated member) accounts for approximately 26.4 percent[56] of the global total.[57] Moreover, most global companies developing advanced AI systems are based in one of the G7 member countries.[58] Establishing interoperable rules for advanced AI systems in these countries is crucial to avoid duplicate compliance costs and to facilitate innovation on a global scale. Third, the G7 is a group of democratic nations, which sets it apart from institutions that include nondemocratic states as members, such as the United Nations and the AI Safety Summit.[59] The HAIP will likely serve as a key foundation, not just for safety but also for realizing fundamental values such as human rights, democracy, and the rule of law in the development and implementation of advanced AI systems.

[56]World Economics, "G7," (2024). Available at: www.worldeconomics.com/Regions/G7/.

[57]Id. See also A. Murphy, H. Tucker, "The Global 2000" (2023) Forbes. Available at https://www.forbes.com/lists/global2000/?sh=763a56c55ac0 for a more fulsome comparative state of corporate presence throughout the G7 and internationally.

[58]T. Keary, "Top 10 Countries Leading in AI Research & Technology in 2024" (2024) Techopedia. Available at: www.techopedia.com/top-10-countries-leading-in-ai-research-technology.

[59]See Figure 2.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

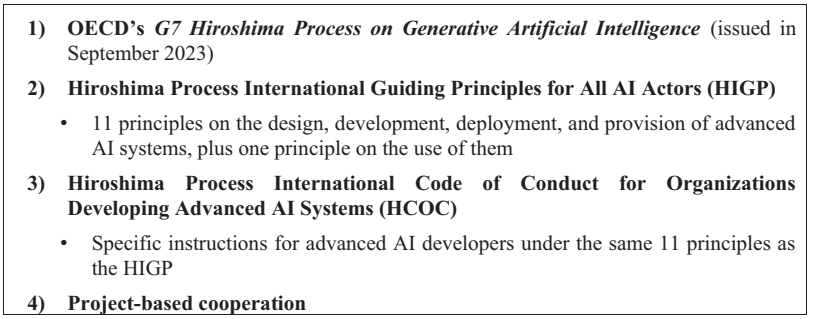

Figure 3. Four elements of the HAIP Comprehensive Framework.

3. Analyzing the Hiroshima AI Process Comprehensive Framework

3.1 Structure of the Comprehensive Framework

In response to the rapid development and global spread of advanced AI, the G7 nations launched the HAIP in May 2023 under Japan's presidency.[60] This international forum aims to establish common ground for responsible AI development and use. It focuses on fostering safe, secure and trustworthy AI by addressing key ethical issues, promoting collaboration on research and development, and encouraging international standards for a future where humanity benefits from AI advancements. Although the HAIP focuses on governance of advanced AI systems, the Comprehensive Framework avoids a rigid definition of this technology by providing the tentative definition of "the most advanced AI systems, including the most advanced foundation models and generative AI systems."[61] This flexibility likely reflects a desire to adapt to future advancements in AI performance, functionalities, and deployment landscapes.

The Comprehensive Framework consists of four elements (see Figure 3). First, the OECD's "G7 Hiroshima Process on Generative Artificial Intelligence"[62] serves as a background analysis of the opportunities and risks of advanced AI systems. Second, the "Hiroshima Process International Guiding Principles for All AI Actors"[63] (HIGP) provides 12 general principles for designing, developing, deploying, providing and using advanced AI systems without providing detailed guidance. Third, the HCOC[64] consists of a set of detailed instructions for the developers of advanced AI systems under the general principles the HIGP provides. Finally, the "project-based cooperation" on AI includes international collaborations in areas such as content authentication and the labeling of AI-generated content.

The following section summarizes the contents of the HIGP and HCOC.

3.2 Contents of the HIGP

The HIGP is a comprehensive set of values and best practices promoting responsible development and use of advanced AI on a global scale. It consists of 12 core principles that serve as a foundation for responsible AI governance. These principles closely mirror the values and approaches that G7 nations

[60] See discussion supra Section 1(ii).

[61] European Commission "Hiroshima Process International Guiding Principles for Advanced AI system" (2023). Available at: https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system.

[62] See OECD (2023), "G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI."

[63] G7, "Hiroshima Process International Guiding Principles for All AI Actors," (2023). Available at: https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document03_en.pdf.

[64] G7, "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems," (2023). Available at: https://www.soumu.go.jp/main_content/000912748.pdf.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

are already exploring for their individual AI governance frameworks.[65] The analysis here suggests that the 12 principles may be divided into the following three groups (see Table 1):

1. **Risk management and governance**: This group recommended actions to assess and mitigate risks associated with AI systems, ensuring they are reduced to a level that relevant stakeholders deem acceptable.
2. **Stakeholder engagement**: This group recommended actions to ensure clear communication with and accountability to all relevant stakeholders.
3. **Ethical and societal considerations**: This group recommended actions to ensure the development, deployment and usage of AI are in alignment with ethical standards and societal values.

3.3 Overview of the Code of Conduct

Building on 11 of the HIGP's 12 core principles (excluding the trustworthy and responsible use of advanced AI),[66] the HCOC translates these principles and materializes them into a more specific code of practice for organizations developing and deploying advanced AI systems. The HCOC provides a comprehensive road map for AI processes and risk mitigation, outlining general actionable items on the matters of risk management and governance, stakeholder engagement, and ethical considerations.[67]

3.3.1 Risk management and governance

The HCOC emphasizes in items 1, 2, 5, 6, 7 and 11 the importance of comprehensive risk management for organizations developing advanced AI across the life cycle of development and implementation. These practices include the following:

- **Risk identification and mitigation**: implementing rigorous testing throughout the AI life cycle, such as red-teaming, to identify and address potential safety, security, and trustworthiness issues
- **Vulnerability and misuse management after deployment**: post-deployment monitoring for vulnerabilities and misuse, with an emphasis on enabling third-party and user vulnerability reporting, possibly via bounty systems
- **Governance and risk management**: creating transparency about organizations' governance and risk management policies and regularly updating users on privacy and mitigation measures
- **Security investments**: implementing robust security measures throughout the AI life cycle to protect critical system components against threats
- **Content authentication**: developing content authentication methods (e.g., watermarking) to help users identify AI-generated content
- **Data quality, personal data and intellectual property protection**: prioritizing data integrity, addressing bias in AI, upholding privacy and respecting intellectual property, and encouraging alignment with relevant legal standards

3.3.2 Stakeholder engagement

The HCOC highlights in items 3 and 4 the critical role of transparency and multistakeholder engagement:

- **Transparency and accountability**: emphasizing public transparency for organizations developing advanced AI, including reporting on both the capabilities of AI systems and their limitations

[65] See discussion infra Section 3(i). See also Annex.

[66] In this case, with the exclusion of "Trustworthy and Responsible Use of Advanced AI."

[67] G7, "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" (2023).

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

- **Responsible information sharing**: encouraging organizations to share information on potential risks, incidents, and best practices with each other, including industry, governments, academia and the public

3.3.3 Ethical and societal considerations

The HCOC establishes in items 8–10 a series of parameters to ensure AI is developed and deployed within the boundaries of human rights and democracy to address global challenges:

- **Research prioritization for societal safety**: emphasizing collaborative research to advance AI safety, security and trustworthiness, focusing on key risks such as upholding democratic values, respecting human rights and protecting vulnerable groups
- **AI for global challenges**: prioritizing development of advanced AI systems to address global challenges such as climate change, health and education, aligning with the UN Sustainable Development Goals
- **International technical standards**: encouraging contribution to the development and use of international technical standards, including practices to promote transparency by allowing users to identify AI-generated content (e.g., watermarking), testing methodologies and cybersecurity policies

A detailed summary of the HCOC is presented in Table 2.

4. The potential of the Hiroshima Code of Conduct: Toward interoperable frameworks for advanced AI systems

The HCOC, as articulated in the Comprehensive Framework, serves as a pivotal instrument to enhance interoperability between various AI governance frameworks.[68] But how compatible is the HCOC with the regulatory frameworks of G7 members? What are the mechanisms or functionalities that make this interoperability possible? Firstly, the HCOC (and similar voluntary codes of conduct) can operate as a potent, nonbinding 'common guidance.' Although not legally enforceable, the gravitas and direction of these documents can wield significant practical influence as soft law (Guzman & Meyer, 2010; Schwarcz, 2020; Wallach et al., 2022; Guruparan & Zerk, 2021). Soft law documents like the HCOC can shape compliance behaviors either as the foundations for good corporate governance or in anticipation of further regulation; they can serve as a reference in private contracts; and can even factor into civil or tort liability decisions.[69]Moreover, such frameworks can provide stability and certainty in an evolving regulatory landscape, enabling organizations to navigate complex AI governance requirements effectively. Second, the HCOC may be integrated into each jurisdiction's regulatory framework in a direct manner.[70] G7 nations are generally poised to either introduce new regulations or revise existing structures on AI governance.[71] If these regulations draw upon the HCOC – whether by reference, content consistency, or formal incorporation – this will increase and facilitate regulatory interoperability as well as international cohesion, integrating an AI governance framework that promotes human rights, democracy and the rule of law.

This section explores the space the HCOC holds within the G7 regulatory context and how it can foster interoperability between the legislative frameworks of different G7 jurisdictions on advanced AI systems. For this, the section first (1) examines the current state of AI regulation within each G7

[68]G7, "Hiroshima AI Process G7 Digital & Tech Ministers' Statement" (2023) at 1. Available at: https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document02_en.pdf.

[69]Id.

[70]See discussion infra Section 3(ii).

[71]See discussion infra Section 3(i).

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

Table 2. Summary of the Hiroshima Process International Code of Conduct

No.	Principle	Summary
Risk Management and Governance		
1	Risk identification and mitigation	Organizations should take appropriate measures to identify, evaluate and mitigate risks throughout the development and deployment of advanced AI systems. This includes diverse testing methods such as red-teaming and ensuring that systems are trustworthy, safe and secure throughout their life cycle.
2	Vulnerability and misuse management after deployment	After system deployment, organizations should monitor for vulnerabilities, incidents, and misuse, adapting their responses based on the level of risk. They are also encouraged to facilitate third-party and user reporting of vulnerabilities through mechanisms such as bounty systems.
5	Governance and risk management policies	Organizations should develop and disclose governance and risk management policies regarding advanced AI systems, including privacy policies and mitigation measures, using a risk-based approach. These policies should be updated regularly and include organizational mechanisms for their implementation.
6	Security investments	Organizations should invest in and implement robust security controls across the life cycle of advanced AI systems, including physical security, cybersecurity and safeguards against insider threats. This aims to secure model weights, algorithms, servers and data sets appropriately.
7	Content authentication	Organizations should endeavor to develop reliable content authentication and provenance mechanisms, such as watermarking, to enable users to identify AI-generated content. This includes implementing tools or application programming interfaces (APIs) that allow users to verify the origin of content created with the organization's advanced AI systems.
11	Data quality, personal data and intellectual property protection	Organizations should take measures to ensure data quality and mitigate harmful biases through transparency, privacy-preserving training techniques, or testing and fine-tuning. They are also encouraged to implement measures to respect privacy and intellectual property rights and compliance with applicable legal frameworks.
Stakeholder Engagement		
3	Transparency and accountability	Organizations should ensure transparency by reporting on the capabilities and limitations of their advanced AI systems, aiming to clarify their safe and appropriate use. This involves sharing meaningful information through transparency reports about the safety, security and societal impacts of systems and any limitations that might affect their use.
4	Responsible information sharing	Organizations should share information on safety and security risks and report incidents in a responsible manner. This collaboration extends across the development community, including industry and academia, to foster the adoption of best practices and standards for security and trustworthiness of advanced AI systems.
Ethical and Societal Considerations		
8	Research prioritization for societal safety	Organizations should invest in research to mitigate societal, safety and security risks, such as prioritizing research on upholding democratic values, respecting human rights, protecting children and vulnerable groups, safeguarding intellectual property rights and privacy, and avoiding harmful bias, mis- and disinformation and information manipulation.
9	AI for global challenges	Organizations should aim their AI development efforts at addressing major global challenges, such as climate change, health and education. This aligns with supporting progress on the UN Sustainable Development Goals and encourages developing AI for the benefit of all.
10	Development of international technical standards	Organizations should contribute to the development and adoption of international technical standards for AI, including best practices for security, content authentication and public reporting. This effort seeks to promote interoperability and helps distinguish AI-generated content from human-created content.

Note: The numerals listed for each item correspond to those assigned in the HIGP and HCOC. The authors devised the abbreviations for the principles and their categorization.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

member state. This analysis assesses the compatibility between the HCOC principles and existing G7 member frameworks. Notably, a significant overlap already exists between the core elements of the G7 nations' regulatory documents and the HCOC.[72] Second, (2) building on this compatibility, the section explores various avenues for integrating the HCOC into the regulatory frameworks of G7 member states. By exploring these options, the section identifies the most effective means of leveraging the HCOC to achieve interoperability in G7 AI governance.

4.1 Status of AI governance in the G7 and HCOC as common guidance

The HCOC serves as a central reference point in the evolving global landscape of AI governance. This section provides insight into how HCOC aligns with the existing frameworks in G7 jurisdictions, including Canada, the EU, Japan, the UK and the United States. Next, the section contains an AI-focused overview of each jurisdiction's regulatory status, identifies the documents that closely align with the HCOC's structure and functionality, and evaluates their compatibility with the HCOC's content. The summary of the analysis is shown in the "Annex."

1. **Canada**: Canada is in the process of formulating a comprehensive regulatory framework for AI under Bill C-27, known as AIDA.[73] This legislation prioritizes risk mitigation for "high-impact" AI systems.[74] Additionally, Canada has published a Voluntary Code of Conduct for Responsible Development and Management of Advanced Generative AI Systems,[75] offering nonbinding guidelines for AI industry stakeholders.
2. **European Union**: The EU has positioned itself at the forefront of AI regulation with the AI Act, published in July 2024.[76] This legislation sets a robust and comprehensive framework for trustworthy AI development and implementation, emphasizing a risk-based regulatory approach.[77] The AI Act mandates the development of codes of practice to guide its implementation, ensuring alignment with international standards as well as evolving technology and market trends.[78]
3. **Japan**: Japan's approach to AI governance emphasizes maximizing the positive societal impacts of AI and capitalizing on a risk-based and agile governance model.[79] Taking a sector-specific approach, Japan seeks to promote AI implementation through regulatory reforms tailored to specific industries and markets, such as transportation, finance, and medical devices.[80] This strategy includes updating more than 10,000 regulations or ordinances that require "analog" compliance methods, including requirements for paper documents, on-site periodic inspections and dedicated in-person staffing.[81] In addition, Japan launched the AI Guidelines for

[72]See Annex.

[73]See Government of Canada (2023), "The Artificial Intelligence and Data Act (AIDA) – Companion document."

[74]Government of Canada, "The Artificial Intelligence and Data Act (AIDA) – Companion document" (2023). Available at: The Artificial Intelligence and Data Act (AIDA) – Companion document (canada.ca).

[75]Government of Canada, "Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems," (Sep. 2023). Available at: https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems.

[76]See European Parliament (2024), "EU AI Act."

[77]See EU AI Act, Chapter V.

[78]See EU AI Act, Art. 56. As of the date of this publication, these codes are in development.

[79]Hiroki Habuka, Center for Strategic & International Studies, "Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency," (Feb. 14, 2023). Available at: https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency.

[80]デジタル庁, アナログ規制見直しの取組, デジタル臨時行政調査会での決定事項等 (2023年9月11日更新); デジタル庁, アナログ規制見直しの取組, 構造改革のためのデジタル原則の全体像 ((令和3年6月); デジタル庁, デジタル原則に照らした: 規制の一括見直しプラン, デジタル臨時行政調査会 (令和4年6月3日); Jiji Press "Japan Govt to Review Nearly 10,000 Items of Analog Regulations" (Dec. 21, 2022), https://sp.m.jiji.com/english/show/23777.

[81]Id.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

Business[82] as a voluntary AI risk management tool. The principles for advanced AI systems established in the HIGP are directly integrated into these guidelines, following Japan's presidency of the G7 during the HAIP Comprehensive Framework drafting process.

4. **United Kingdom**: The UK is developing a decentralized regulatory approach focusing on sector-specific guidelines, a pro-innovation stance and public–private collaboration through specialized AI institutions.[83] While the UK is not currently enforcing a comprehensive AI law or drafting a central code of conduct, it emphasizes traditional AI governance principles such as safety, security, transparency, and fairness to inform its sector-driven regulations.[84] The UK Department for Science, Innovation and Technology also published a practical guidance code in the form of the Emerging Processes for Frontier AI Safety[85] ahead of the UK AI Safety Summit. The summit culminated in the Bletchley Declaration, a shared commitment to safe and responsible AI development signed by 28 nations and the EU.[86]
5. **United States**: The United States has adopted a decentralized, multitiered regulatory strategy for AI governance, with agencies overseeing sector-specific regulations.[87] Key initiatives include the "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,"[88] which directs sector-specific agencies to formulate regulations; the Risk Management Frameworks,[89] developed by NIST to provide guidelines for risk assessment and management; the "White House's Blueprint for an AI Bill of Rights,"[90] outlining foundational principles for AI development; and nonbinding voluntary commitments for ensuring safe, secure and trustworthy AI[91] endorsed by companies such as Amazon, Anthropic, Google, Inflection, Meta, Microsoft, Nvidia and OpenAI, among others.

4.2 Achieving and enhancing regulatory interoperability: The HCOC as a reference point for AI governance development

The AI governance landscape across the G7 is complex and multifaceted. The EU has instituted robust and comprehensive regulations through its AI Act, and Canada is in the process of developing similar

[82] See Japan Guidelines for Business.

[83] See discussion infra Section 3(ii) (addressing the UK's approach to AI governance). See also UK Department for Science, Innovation & Technology, "A pro-innovation approach to AI regulation: government response" (2024) Art. 5. CP 1019. E03019481 02/24. Available at: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

[84] See "A pro-innovation approach to AI regulation: government response." generally and at Art. 5.

[85] See United Kingdom Department for Science, Innovation & Technology (Oct. 2023), "Emerging Processes for Frontier AI Safety."

[86] See AI Safety Summit (2023), "The Bletchley Declaration."

[87] See generally The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" (2023). Available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ ("The White House Executive Order on the Use of AI"); OECD: AI Policy Observatory Policy initiatives of United States. In: OECD AI Policy Obs. https://oecd.ai/en/dashboards/policy-initiatives?conceptUris=http:%2F%2Fkim.oecd.org%2FTaxonomy%2FGeographicalAreas%23UnitedStates (last accessed Aug. 10, 2024); R.B.L. Dixon (2023); Plotinsky and Cinelli (2024).

[88] See The White House Executive Order on the Use of Artificial Intelligence.

[89] USDoC NIST AI RMF 1.0; USDoC NIST, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," NIST AI 600-1, (July 2024). Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

[90] The White House, "Blueprint for an AI Bill of Rights," (Oct. 2022) Available at: https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[91] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI," (Jul. 21, 2023). Available at: https://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/; The White House, "Ensuring Safe, Secure, and Trustworthy AI." Available at: https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf (last visited Aug. 10, 2024).

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

hard-law frameworks.[92] Conversely, the United States, Japan and the UK lean toward sector-specific and lighter-touch regulatory approaches.[93] This diverse regulatory environment, marked by varying levels of stringency, scope and focus, poses challenges for global operations, requiring businesses to navigate a complex regulatory patchwork, as well as differing rights and obligations across G7 nations. The HCOC holds promise as a unifying mechanism, to bridge these regulatory disparities and promote interoperability.

The HCOC may be integrated into national regulations across G7 countries through various means, such as direct formal legal referencing or recognition, material content integration, and leveraging or harmonizing specific aspects of regulatory developments. Pathways for integration into the regulatory frameworks of the G7 jurisdictions include the following:

- **Canada:** Overall, Canada's Voluntary Code of Conduct specifically, and its regulatory trajectory generally, demonstrate alignment with the international conversation on ethical AI development and the HCOC's principles.[94] As AIDA evolves, it presents the potential to translate these principles into enforceable regulations, further solidifying Canada's commitment to responsible AI advancement. Given that AIDA could likely address advanced AI systems specifically within its regulatory scope, this upcoming law opens a clear possibility to find common ground with the HCOC's principles and functionality.
- **European Union**: The EU AI Act mandates the development of codes of practice that complement its implementation.[95] These codes of practice align with the HCOC's focus, addressing practical aspects of responsible AI development. Furthermore, the EU acknowledges in the EU AI act that international standards should play a role in shaping these codes of practice,[96] presenting an opportunity to materially integrate or formally reference the HCOC in the EU AI governance framework.
- **Japan**: In February 2024, the Liberal Democratic Party proposed the concept note for the Basic Law for the Promotion of Responsible AI.[97] The proposed legislation specifically targets advanced foundational AI models with significant societal impact. It requires model developers to adhere to seven key measures,[98] including third-party vulnerability checks and the disclosure of model specifications. The requirements align with the voluntary commitments the White House has requested from U.S. companies.[99] The HCOC could serve as a valuable reference point for implementation of these key measures, especially considering that the HCOC principles are already integrated into Japan's AI Guidelines for Business.[100]
- **United Kingdom**: Besides leading international discussions on AI governance through initiatives such as the Bletchley Declaration,[101] the UK is proactively formulating its own AI governance framework. According to "A Pro-innovation Approach to AI Regulation,"[102] the UK government is undergoing technical policy analysis on regulation and life-cycle accountability of capable general-purpose systems. It also commits to updating the Emerging Processes for

[92] See generally discussion supra Section 3(i).

[93] Id.

[94] See Annex.

[95] See EU AI Act, Art. 53.4.

[96] See EU AI Act, Art. 56.1.

[97] See Liberal Democratic Party of Japan (2024), "Basic Law for the Promotion of Responsible AI."

[98] Id.

[99] See The White House (2023), "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI."

[100] See Annex; Japan's AI Guidelines for Business.

[101] See discussion supra Section 1(i). See also See AI Safety Summit (2023), "The Bletchley Declaration."

[102] See UK Department for Science, Innovation & Technology (2024) "A pro-innovation approach to AI regulation: government response."

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

Frontier AI, Safety,[103] which is highly compatible with the HCOC, by the end of 2024. For these purposes, the UK is opting for collaborative private-public development through institutions such as the Digital Regulation Cooperation Forum[104] and the AI Safety Institute.[105] Considering the current institutional inertia and the stalled progress regarding its draft intellectual property code,[106] the UK could leverage the HCOC and its international scope to inform these regulatory initiatives.

- **United States**: The United States is in active development of its AI governance frameworks. The AI executive order has directed multiple agencies to deliver sector-specific guidance publications, and as of August 2024 there are more than 105 draft bills addressing AI, with over 35 focused on risk mitigation.[107] Notably, after releasing RMF 1.0 in January 2023, NIST established the Generative AI Public Work Group[108] to spearhead development of a cross-sectoral AI RMF profile for managing the risks of generative AI models or systems.[109] The HCOC's emphasis on responsible risk management and governance aligns seamlessly with the United States' principles-based trajectory and could fit into proposed risk mitigation legislation, positioning the HCOC as a crucial reference in shaping AI regulatory policy in the United States.

5. HCOC 2.0: Next steps toward a more harmonized and impactful AI governance framework

The current AI governance landscape is characterized by jurisdictional fragmentation, with disparate national regulations imposing varying obligations on developers and offering inconsistent protections to users. While the HCOC holds promise for harmonizing G7 approaches and inspiring broader international cooperation, its lack of specificity currently limits its practical utility. The following section posits that, to realize the HCOC's full potential, future G7 discussions should prioritize development in key areas such as (1) terminology and definitional interoperability, (2) risk management, (3) stakeholder engagement, (4) ethical considerations and (5) further areas for exploration not currently contained in the HCOC. By establishing a robust and adaptable framework, the G7 can position the HCOC as a global benchmark for responsible AI development, anchored in shared values of human rights, democracy and the rule of law.

5.1 Terminology and definitions: Indexing a common vocabulary

The HCOC can serve as a foundation for a consistent definition or methodology for identification of terms for advanced AI systems governance, facilitating smoother regulatory implementation across jurisdictions. Future terminology consensus includes the following:

- **Bridge the terminology gap**: The HCOC can endorse consistent definitions for streamlined regulatory implementation across jurisdictions, fostering a common understanding of critical concepts. This could be achieved by including a glossary of key terms with clear, agreed-upon definitions or by establishing methodologies for identifying and classifying AI systems based on the factors relevant to risk assessment. By establishing a common language, the HCOC can ease communication, regulatory certainty and business-sector collaboration across borders. Underscoring the importance of shared language around AI, the EU and the United States are currently in the development of 65 key terms "essential to understanding risk-based approaches

[103] See UK Department for Science, Innovation & Technology (Oct. 2023), "Emerging Processes for Frontier AI Safety."
[104] Digital Regulation Cooperation Forum, "About the DRCF." Available at: https://www.drcf.org.uk/about-us.
[105] UK Department for Science, Innovation & Technology, "Introducing the AI Safety Institute," (Jan. 17, 2024). Available at: https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute.
[106] See "A pro-innovation approach to AI regulation: government response," at Art. 29; Joseph and Berry (2024).
[107] See AAF AI Legislation Tracker; discussion supra Section 3(i).
[108] See USDoC NIST. Available, "NIST AI Public Working Groups." Available at: https://airc.nist.gov/generative_ai_wg.
[109] Id.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

to AI."[110] Notably, even when common terminology has been developed (e.g., through the U.S.-EU Trade and Technology Council, OECD or the ISO), the definition of advanced AI systems is unclear, leaving the question of which criteria (e.g., floating point operations, quality and size of data set, or input and output modalities)[111] should be used to determine advanced AI systems.

5.2 Risk management and governance: Building a common and robust framework

Effective risk management stands as a cornerstone of responsible development of advanced AI systems. The HCOC can significantly contribute to this endeavor by advocating for shared principles and best practices. Risk management cohesion across jurisdictions includes the following:

- **Identify and share security risks, particularly systemic risks**: The HCOC can enhance its interoperability contribution by explicitly listing and addressing security risks, particularly those with systemic consequences. This can be achieved through a two-pronged approach. First, the HCOC can integrate a comprehensive list of typical AI risks common to advanced AI systems, such as AI hallucinations (generating inaccurate outputs), fake content generation (deepfakes), intellectual property infringement (copyrighted content integration in data sets), job market transformations due to automation, the environmental impact of AI systems, bias amplification based on training data and privacy concerns, among others. Case studies can be implemented through "project-based cooperation," which constitutes the fourth element of the Comprehensive Framework. Second, the HCOC can establish a risk assessment framework to categorize AI systems based on their potential for harm. This framework could leverage existing models such as the EU AI Act's categorization of general-purpose AI models with systemic risk and its classification rules for high-risk AI systems.[112] By prioritizing systems with the greatest potential for systemic or high-impact issues, the HCOC can provide a clearer road map for identifying, understanding, and mitigating various risks.
- **Enhance clarity in the risk management process**: The HCOC can encourage the development of standardized risk management policies tailored to specific AI applications. Future drafting can reference or draw insights from established RMFs, such as ISO/IEC 42001:2023 or NIST's RMF – especially the RMF developed by the Generative AI Public Working Group[113] in July 2024. Additionally, policies can incorporate learnings from other reputable sources to enhance clarity and comprehensiveness.
- **Develop standard data governance, risk management and information security policies**: Establishing robust data protection protocols is essential for building trust and mitigating risks associated with AI development. The development of standardized policies can leverage established frameworks such as ISO/IEC 27001 and ISO/IEC 27002 or NIST's Cybersecurity Framework, which provide a structured foundation adaptable to the unique risk landscape of the development of advanced AI systems.
- **Implement content authentication mechanisms**: The HCOC can list reliable content authentication and provenance mechanisms to enable users to identify the originators of content or establish common labeling mechanisms to help users understand that AI has generated the content. These contributions could be based on input from the HAIP's project-based cooperation. Authentication mechanisms can safeguard against misinformation and uphold democratic values and human rights by verifying data sources and outputs. However, it is imperative to balance

[110] European Commission, "EU-U.S. TTC: Call for input on first edition of WG1 Terminology and Taxonomy for Artificial Intelligence." (last update: Nov. 3, 2023). Available at: https://digital-strategy.ec.europa.eu/en/news/eu-us-ttc-call-input-first-edition-wg1-terminology-and-taxonomy-artificial-intelligence.

[111] See EU AI Act, Annex XIII; Art. 51.2.

[112] See EU AI Act, Chap. III, Sec. 1; Art. 6; Annex III.

[113] USDoC NIST, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," NIST AI 600-1, (July 2024). Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

these efforts with the protection of individual privacy, ensuring authentication processes do not compromise personal data. This balance is key to maintaining public trust and promoting the responsible and user-centric deployment of AI technologies.

5.3 Stakeholder engagement: Fostering transparency and accountability

Building trust in AI necessitates robust stakeholder engagement. A transparent and accountable AI development process fosters public confidence and encourages information sharing. Future pathways for stakeholder engagement include the following:

- **Establish standardized formats for transparency reports**: The HCOC can promote the adoption of standardized formats for transparency reports. By consolidating best practices and identifying common risks, the HCOC can offer a template for companies to self-assess and disclose relevant information consistently across jurisdictions. A potential model for standardized formatting pursuant to transparency reports is the UK Algorithmic Transparency Recording Standard.[114] Standardization would enable companies to have uniform international disclosure criteria, enhancing cross-border cohesive reporting and auditing consistency as well as allowing the public to better understand the development and operation of AI systems.
- **Define clear formats for incident sharing**: Encouraging adoption of clear incident-sharing formats can facilitate the exchange of information about security breaches, biases or unintended consequences observed in deployed AI systems. This collaborative approach to sharing and learning from incidents enables stakeholders to develop effective mitigation strategies, ultimately enhancing the safety and reliability of AI technologies.

5.4 Ethical and societal considerations: Upholding the rule of law, human rights and core democratic values

The G7, a group of leading democracies, has a unique opportunity to shape the global conversation around responsible AI development. The HCOC, as an initiative stemming from this group, can play a crucial role in ensuring AI development aligns with the ethical and societal considerations that underpin democratic values and secure human rights in AI development and implementation. Potential pathways to prioritizing these principles include the following:

- **Reinforce the primacy of rule of law, human rights and democratic principles:** The HCOC already champions these core values and emphasizes human-centric design. However, there is room for further enhancement and substantiation for practical application. For instance, the HCOC could enhance its guidance on how organizations should foster research and AI development that prioritizes the protection of fairness, privacy and intellectual property rights while also tackling global challenges such as climate change, health and education. Rather than providing detailed descriptions itself, the HCOC could reference other international agreements or widely recognized standards. Furthermore, the HCOC could strengthen democratic principles and the rule of law by highlighting due safeguards for freedom of expression, ensuring AI does not minimize dissent or impose undue restrictions on information access, guaranteeing a right to remedy for individuals adversely affected by AI and promoting transparency and accountability in AI decision-making processes. Enhancing human-centricity could involve advocating

[114]United Kingdom Department for Science, Innovation & Technology, "Algorithmic Transparency Recording Standard – Guidance for Public Sector Bodies," (Jan. 5, 2023). Available at: https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

for effective oversight in high-risk applications, providing individuals with explanations regarding AI-driven decisions affecting them, and promoting inclusive design that caters to the diverse needs and perspectives of various populations to ensure equitable AI benefits.

5.5 Further areas for exploration

The HCOC can play a key role in exploring several critical areas for further development in responsible AI:

- **Acknowledge special considerations for government use of AI:** The HCOC can play a pivotal role in delineating special considerations for government use of AI, ensuring governmental powers in AI deployment are appropriately circumscribed and limited. Drawing inspiration from the AI Treaty[115] and leveraging principles from the OECD Declaration on Government Access to Personal Data Held by Private Sector Entities,[116] the HCOC can establish clear guidelines that emphasize due process in developing and deploying advanced AI systems by the public sector, such as legal basis, legitimate aims, oversight, and redress, in addition to shared principles such as privacy, transparency and accountability. By aligning with these principles, the HCOC can become a democratic referent, and governments can leverage the power of AI responsibly while mitigating potential risks and fostering public trust.
- **Harmonize full life cycle regulatory approaches:** The HCOC can explore the potential for incorporating best practices from various jurisdictions' regulations. This could involve elements such as certification mechanisms, robust oversight mechanisms and iterative audit controls.
  - **Certification mechanisms:** The HCOC can establish a framework for certification and registration mechanisms for high-risk advanced AI systems. This system could ensure rigorous evaluation throughout the life cycle of high-risk and advanced AI systems, from pre-market integrative assessments to ongoing post-market analyses and compliance reviews. The HCOC could define risk categories and establish criteria for the need for certification.
  - **Oversight methodologies:** The HCOC can emphasize the importance of effective oversight in AI systems to mitigate potential harm and address incidents effectively. In some cases, human involvement in critical AI processes is necessary, while in other cases machines can detect risks much faster and more precisely than humans. The HCOC could propose guidelines about whether to prioritize human judgment and intervention, especially in high-risk AI applications, ensuring a balance between automation and human control.
  - **Audit mechanisms:** The HCOC can extend procedural cohesion beyond AI implementation by establishing common processes for iterative audits, ensuring continuous monitoring and evaluation of AI systems' compliance with established principles and guidelines. By considering and potentially adapting existing frameworks, such as the UK Guidance on the AI Auditing Framework,[117] the HCOC can equip organizations with practical tools for ongoing evaluations. These iterative audits would allow for continuous improvement and ensure AI systems remain aligned with responsible development principles throughout their life cycle.

[115]See generally Council of Europe, Committee on Artificial Intelligence, Draft Framework Convention on artificial intelligence, human rights, democracy and the rule of law (2024). Available at: https://search.coe.int/cm/pages/result_details.aspx?objectid=0900001680aee411.

[116]Organisation for Economic Co-operation and Development, Declaration on Government Access to Personal Data Held by Private Sector Entities (2023) OECD/LEGAL/0487. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0487.

[117]Information Commissioner's Office, Guidance on the AI auditing framework: Draft guidance for consultation (2023). Available at: https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

- **Establish means for redress:** The HCOC could expand discussions about redress for harms caused by advanced AI systems. This could involve exploring access to remedies and explanations for individuals affected by AI decisions in areas as diverse as copyright and intellectual property to judicial processing. As AI plays a growing role in judicial decision-making, for example, developing specific appeal mechanisms for harms caused by AI-based judicial decision making may become also crucial. The HCOC could encourage developers and deployers of advanced AI systems to provide appropriate dispute resolution mechanisms to users and harmed parties. Furthermore, to make victim relief more effective, G7 members could discuss shifting the burden of proof of damages or causal links and establishing accessible, fast and low-cost dispute resolution mechanisms for damages caused by advanced AI systems.
- **Foster shared responsibility in the AI ecosystem:** The HCOC addresses developers of advanced AI systems only.[118] However, its scope could expand in the future to other actors within the AI value chain, such as deployers and users of advanced AI systems. In addition, it is important to examine how to distribute responsibility and liability among stakeholders, ensuring all parties are accountable for their respective roles in potential harms.

By focusing on these key areas, the HCOC can evolve into a powerful tool for facilitating a more cohesive and effective approach to AI governance on a global scale. The HCOC's dynamic nature positions it to bridge the gap between diverse national frameworks, fostering a future of responsible AI development for the G7 nations and beyond.

6. Conclusion

The G7 nations' endorsement of the HIGP and the HCOC, supported by more than 40 countries through the Hiroshima AI Process Friends Group, represents a significant milestone in global AI governance.[119] This unified stance by the world's leading democratic economies underscores a robust international commitment to advancing human-centered AI development, safeguarding individual rights, and strengthening trust in AI systems. The collective weight and global influence of the nations lending their support to this process amplifies the significance of its agreements, marking them as pivotal steps in shaping the future of AI governance.

However, for the promise of the Comprehensive Framework to be fully realized, its key practical instrument, the HCOC, requires further development. While the HCOC, as this article reveals, significantly aligns with the trajectory of existing G7 policies, it currently lacks the material specificity to provide truly effective guidance for practical implementation. Moving forward, it is crucial to engage in substantive discussions on enhancing the HCOC in several key areas. These areas include the following:

- **Coordinating a common vocabulary:** A unified understanding of key terms and definitions is essential for ensuring consistent interpretation of AI terms across borders.
- **Developing robust RMFs and risk-based categorization:** The HCOC should provide clear guidance on assessing and mitigating risks associated with advanced AI systems throughout the entire AI life cycle, from pre-market duties to post-market updates.
- **Promoting harmonized stakeholder engagement:** The HCOC can play a valuable role in encouraging cohesive approaches to stakeholder engagement and developing consistent transparency standards.
- **Strengthening democratic and human rights principles**: The HCOC should provide more concrete and actionable steps for upholding democratic values and safeguarding human rights in the context of AI development and deployment.

[118] See generally HCoC.
[119] See generally discussion supra Section 1.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

- **Pursuing further areas for discussion:** The HCOC's potential extends beyond its current scope. The G7 can leverage this collaborative document to explore critical areas such as developing special considerations for government AI use, harmonizing life cycle regulatory practices (e.g., certification mechanisms, oversight methodologies and audit mechanisms), fostering shared responsibility within the AI ecosystem, and establishing efficient redress for AI harms.

By addressing these crucial areas, the HCOC has the potential to evolve into a truly robust and impactful instrument for global AI governance. A strengthened HCOC can serve as a valuable reference point not only for G7 nations and friends, but also for a broader international audience seeking to navigate the complexities of responsible AI development and deployment. This international alignment can help ensure the power of AI is harnessed for the benefit of all while mitigating potential risks and upholding core human values.

Funding statement. The authors declare no funding. This article is an updated version of a report published by the Center for Strategic & International Studies.

Competing interests. The authors declare no competing interests to disclose.

References.

Andrews, L., & Bucher, H. (2022). Automating discrimination: AI hiring practices and gender inequality. *Cardozo Law Review*, *44*, 145.

Bharadiya, J. P., Thomas, R. K., & Ahmed, F. (2023). Rise of artificial intelligence in business and industry. *Journal of Engineering Research and Reports*, *25*(3), 85–103. doi:10.9734/jerr/2023/v25i3893

Bommasani, R., Liang, P., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R. E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S. M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L. & Zhou, K. (2021). On the opportunities and risks of foundation models. In *Center for Research on Foundation Models (CRFM)*, Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Co.

Dixon, R. B. L. (2023). A principled governance for emerging AI regimes: Lessons from China, the European Union, and the United States. *AI and Ethics*, *3*(3), 793–810. doi:10.1007/s43681-022-00205-0

Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Sciences*, *7*(1), 549–569. doi:10.1007/s42001-024-00250-1

Guruparan, K., & Zerk, J. (2021). Influence of Soft Law Grows in International Governance. *Chatham House*. https://www.chathamhouse.org/2021/06/influence-soft-law-grows-international-governance.

Guzman, A., & Meyer, T. (2010). International soft law. *Journal of Legal Analysis*, *2*(1), 171–225. doi:10.1093/jla/2.1.171

Joseph, P., & Berry, K. (2024). UK Fails to Agree AI/Copyright code of practice. *Linklaters*. https://techinsights.linklaters.com/post/102j0q6/uk-fails-to-agree-ai-copyright-code-of-practice.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi:10.1038/nature14539

Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., & Briggs, M. (2022). Human rights, democracy, and the rule of law assurance framework for AI systems: A Proposal Prepared for the Council of Europe's Ad Hoc Committee on Artificial Intelligence. *The Council of Europe*. https://rm.coe.int/huderaf-coe-final-1-2752-6741-5300-v-1/1680a3f688.

Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y. H., Henshaw, H., Alhawawreh, M., Berdik, D., & Jararweh, Y. (2024). Foundation and large language models: Fundamentals, challenges, opportunities, and social impacts. *Cluster Computing*, *27*(1), 1–26. doi:10.1007/s10586-023-04203-7

Plotinsky, D., & Cinelli, G. (2024). Existing and proposed federal AI regulation in the United States. *Morgan Lewis*. Available at: https://www.morganlewis.com/pubs/2024/04/existing-and-proposed-federal-ai-regulation-in-the-united-states.

Schneider, J., Meske, C., & Kuss, P. (2024). Foundation models: A new paradigm for artificial intelligence. *Business and Information Systems Engineering*, *66*(3), 221–231. doi:10.1007/s12599-024-00851-0

Schwarcz, S. (2020). Soft law as governing law. *Minnesota Law Review 104*(3265), 2471.

Steenkiste, S., Chang, M., Greff, K., & Schmidhuber, J. (2018). Relational neural expectation maximization: Unsupervised discovery of objects and their interactions. In *International Conference on Learning Representations*.

Wallach, W., Reuel, A., & Kaspersen, A. (2022). Soft law functions in the international governance of AI. *Carnegie Council for Ethics in International Affairs*. https://cdn.carnegiecouncil.org/media/cceia/Soft-Law-in-International-AI-Governance.pdf?v=1695911519.

https://doi.org/10.1017/cfl.2024.5 Published online by Cambridge University Press

Annex. Mapping Jurisdictional Coverage of Key Principles in the Hiroshima Process International Code of Conduct: Alignment with National AI Regulations and Guidance

Item[a]	Hiroshima AI Process	Canada[b]	European Union	Japan	United Kingdom	United States
General						
Document	Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems	Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems	AI Act	AI Guidelines for Business	Emerging Processes for Frontier AI Safety	Voluntary Commitments for Ensuring Safe, Secure, and Trustworthy AI
Target	Advanced AI systems	Advanced generative AI systems	General-purpose AI	Advanced AI systems	Frontier AI	Generative AI (foundation) model technology
Definition	The most advanced AI systems, including the most advanced foundation models and generative AI systems	Generative AI systems that have advanced capabilities enabling them to be adapted for a wide variety of uses in different contexts, including uses for which they were not specifically trained	An AI model – including one trained with a large amount of data using self-supervision at scale – that (1) displays significant generality and the capability to competently perform a wide range of distinct tasks regardless of the way the model is placed on the market and (2) can be integrated into a variety of downstream systems or applications	The most advanced AI systems, including the most advanced foundation models and generative AI systems (same as HCOC)	Highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models.	–
Risk Management and Governance						
Risk Identification and Mitigation	1 (Risk Identification and Mitigation)	2 (Safety), 3 (Fairness and Equity), 6 (Validity and Robustness)	Art. 6 (high-risk classification); Annex III (categorical high-risk systems); Art. 51 (classification of general-purpose AI models with systemic risk); Arts. 10–13 (risk mitigation); Art. 55(a) (model evaluation)	1 (Risk Identification and Mitigation)	1 (Responsible Capability Scaling), 2 (Model Evaluations and Red Teaming)	1 (Commit to internal and external red-teaming of models or systems in areas including misuse, societal risks, and national security concerns, such as bio, cyber, and other safety areas)

(*Continued*)

(*Continued.*)

Item[a]	Hiroshima AI Process	Canada[b]	European Union	Japan	United Kingdom	United States
Vulnerability and Misuse Management after Deployment	2 (Vulnerability and Misuse Management after Deployment)	5 (Human Oversight and Monitoring)	Chapter IX generally Art. 72 (post-market monitoring); Art. 55(1)(b), (c) (obligations for providers); Art. 85 (right to lodge a complaint); Art. 87 (reporting of infringements and protection of reporting persons)	2 (Vulnerability and Misuse Management After Deployment)	1 (Responsible Capability Scaling), 8 (Preventing and Monitoring Model Misuse)	4 (Incent third-party discovery and reporting of issues and vulnerabilities)
Governance and Risk Management Policies	5 (Governance and Risk Management Policies)	1 (Accountability)	Art. 9 (risk management system for high-risk systems) generally; Art. 13 (policy transparency); Art. 14 (oversight measures commensurate to risks for high-risk AI systems); Art. 53 (obligations for providers of general-purpose AI models); Art. 55 (obligations for providers of general-purpose AI models with systemic risk)	5 (Governance and Risk Management Policies)	All	2 (Work toward information sharing among companies and governments regarding trust and safety risks, dangerous or emergent capabilities, and attempts to circumvent safeguards)
Cybersecurity Investments	6 (Cybersecurity Investments)	6 (Validity and Robustness)	Art. 15 generally (high-risk systems); Art 55 (1)(c); Art. 14 (human oversight requirements for high-risk systems); Art. 26 (obligations for deployers of high-risk AI systems); Art. 70 (designation of national authorities)	6 (Cybersecurity Investments)	4 (Security Controls Including Securing Model Weights)	3 (Invest in cybersecurity and insider threat safeguards to protect proprietary and unreleased model weights)
Content Authentication	7 (Content Authentication)	4 (Transparency)	Art. 55 (obligations for providers) generally; Art. 10(2)(c) (data and data governance); Art. 13. (3)(b)(vi),(vii) (transparency and provision of information to deployers), Art. 48 (CE marking for high-risk systems); Art. 50(1a) (marking of outputs)	7 (Content Authentication)	6 (Identifiers of AI-generated Material)	5 (Develop and deploy mechanisms that enable users to understand if audio or visual content is AI-generated, including robust provenance, watermarking, or both, for AI-generated audio or visual content)

(*Continued*)

(*Continued.*)

Item[a]	Hiroshima AI Process	Canada[b]	European Union	Japan	United Kingdom	United States
Data Quality, Personal Data, and Intellectual Property Protection	11 (Data Quality, Personal Data, and Intellectual Property Protection)	3 (Fairness and Equity)	Art. 53 (1)(b), (c) (copyright and intellectual property); Art. 10 generally (bias); Art. 15 (accuracy and robustness)	11 (Data Quality, Personal Data, and Intellectual Property Protection)	9 (Data Input Controls and Audits)	–
Stakeholder Engagement						
Transparency and Accountability	3 (Transparency and Accountability)	2 (Safety), 3 (Fairness and Equity)	Art. 11 (technical documentation), Art. 12 (record keeping) generally; Art. 13 generally (transparency and provision of information to deployers); Art. 53(1)(a) (technical documentation), (b) and (d) (availability of documentation); Art. 50 (transparency to user); Art. 17(1)(m) (accountability requirement for high-risk AI systems)	3 (Transparency and Accountability)	3 (Model Reporting and Information Sharing)	6 (Publicly report model or system capabilities, limitations, and domains of appropriate and inappropriate use, including discussion of societal risks, such as effects on fairness and bias)
Responsible Information Sharing	4 (Responsible Information Sharing)	1 (Accountability), 5 (Human Oversight and Monitoring)	Title III Chapter 4 (notifying authorities and notified bodies) generally; Arts. 11 (technical documentation), 12 (record keeping) generally; Art. 55 (1)(c) (obligations for providers); Art. 73 (reporting of serious incidents)	4 (Responsible Information Sharing)	3 (Model Reporting and Information Sharing), 5 (Reporting Structure for Vulnerabilities)	2 (Work toward information sharing among companies and governments regarding trust and safety risks, dangerous or emergent capabilities, and attempts to circumvent safeguards)
Ethical and Societal Considerations						
Research Prioritization for Societal Safety	8 (Research Prioritization for Societal Safety)	–	Art. 1(1) (purpose of EUAI Act); Art. 60(4)(g); Art. 14 (preventing risks to health, safety, or fundamental rights); Art. 5 (prohibited systems)	8 (Research Prioritization for Societal Safety)	7 (Prioritising Research on Risks Posed by AI)	7 (Prioritize research on societal risks posed by AI systems, including on avoiding harmful bias and discrimination, and protecting privacy)

(*Continued*)

(*Continued.*)

Item[a]	Hiroshima AI Process	Canada[b]	European Union	Japan	United Kingdom	United States
AI for Global Challenges	9 (AI for Global Challenges)	–	Art. 1(1) (environmental protection); Art. 59 (sandbox conditions); Art. 95 (2)(b) (code of conduct); Art. 112 (evaluation)	9 (AI for Global Challenges)	–	8 (Develop and deploy frontier AI systems to help address society's greatest challenges)
Development of International Technical Standards	10 (Development of International Technical Standards)	Premise	Art. 40 (harmonized standards); Art. 56 (codes of practice)	10 (Development of International Technical Standards)	Model Reporting and Information Sharing	2 (Work toward information sharing among companies and governments regarding trust and safety risks, dangerous or emergent capabilities, and attempts to circumvent safeguards)
Additional Elements						
Additional Elements	N/A	• Human oversight • Description of the types of training data	• Sufficiently detailed summary about the content used for training • Pre-market conformity assessments and registration • Increased compliance requirements for high–risk AI • Ban of AI systems with unacceptable risk (e.g., biometric categorization based on sensitive characteristics, certain predictive policing algorithms, and social scoring systems) • Development of codes of practice for implementation specifications	–	• Data input controls and audits	• Specific note to work on a strong international code of conduct • References to AI within the context of national security

[a]Based on the structure of the HCOC.
[b]Given the structure of Canada's Voluntary Code of Conduct, the numbers in this column correspond as follows: 1 = Accountability, 2 = Safety, 3 = Fairness and Equity, 4 = Transparency, 5 = Human Oversight and Monitoring, 6 = Validity and Robustness.
[c]The numbers listed in each cell indicate the section or article numbers of the corresponding documents in each country.

Cite this article: Habuka H and Socol de la Osa D.U. (2025). Enhancements and next steps for the G7 Hiroshima AI Process: Toward a common framework to advance human rights, democracy and rule of law. *Cambridge Forum on AI: Law and Governance* **1**, e15, 1–26. https://doi.org/10.1017/cfl.2024.5