SaferAI

G7 Hiroshima AI Process Code of Conduct and EU AI Act GPAI

Commonality Analysis

James Gealy
Daniel Kossack

SaferAI

Introduction

This document contains an analysis of the commonalities and differences between the G7 Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems—herein referred to as the Hiroshima AI Process Code of Conduct (CoC)—and the EU AI Act ("Act" or "AIA") text on general-purpose AI (GPAI) models.

There is substantial commonality between the texts, though each has requirements/recommendations not found in the other. In essence, their commonality can be thought of as fitting a Venn diagram, with approximately 30% high or complete commonality, 50% moderate commonality, and 20% not overlapping where requirements or recommendations from one are not found in the other.

For example, regarding copyright and intellectual property law, the Act has a specific focus on providers complying with EU law on copyright. With regards to public disclosure and reporting to regulators, the CoC intends for public reporting (e.g., Action 3) while the Act intends for documentation to be provided to the AI Office upon request, as well as to organisations further downstream on the value chain. That said, many points are the same or very similar, such as risk assessment, risk mitigation and cybersecurity.

The CoC Actions tend to be more detailed than the requirements in the Act's Articles and give specific examples and expectations. If the Act's Recitals are included though, then the level of detail is more comparable to the CoC. The Act is more detailed in certain ways, such as the documentation and transparency requirements in the Annexes. And many of the CoC requirements that are more detailed can be inferred from the Act's text (e.g., Action 1's secure testing environments requirement can be reasonably inferred from the Act's cybersecurity and evaluations requirements).

Our analysis is based on three assumptions. Firstly, we include the AI Act's Recitals and Article 56 requirements given their additional detail, e.g., 56(2)(d). Secondly, all CoC "shoulds" are considered mandatory in the sense that they are all assumed to be fulfilled. Thirdly, "Advanced AI systems" and GPAI "models with systemic risk" are assumed to be equivalent.

All CoC text is covered in the tables. AIA text relevant to the CoC but not found in an Action is included at the end of each table.

Disclaimer

Note that commonality between these frameworks does not imply mutual compliance. In particular, completing the HAIP Code of Conduct Reporting Framework does not provide presumption of conformity with the requirements in the EU AI Act. Nor does becoming a signatory of the EU GPAI Code of Practice enroll one in the Reporting Framework.

SaferAI

Table of Contents

SaferAI

Summary Table

The following table shows the number of points of comparison between the Code of Conduct and the EU AI Act, per Code of Conduct Action:

- **86 points of comparison in total**
- **31% of the comparisons have high or complete commonality**
- **Just over 80% have at least some commonality**

Subject of the Code of Conduct Action	**Action**	High or complete commonality	Some commonality	Little or no commonality	**Total**
General and Introduction	General and Introduction		1	2	**3**
Risk management and evaluations	Action 1	7	5	5	**17**
Identify and mitigate vulnerabilities	Action 2	2	7	2	**11**
Transparency and documentation	Action 3	1	7		**8**
Incident reporting and information sharing	Action 4	3	7	2	**12**
Risk management framework	Action 5	1	3	2	**6**
Cybersecurity	Action 6	6	4		**10**
Content Authentication and Provenance Mechanisms	Action 7	3	1	1	**5**
Investments in Research and Mitigation Measures	Action 8		3		**3**
Developing AI for the Benefit of the Public	Action 9	1	3	1	**5**
Development and Adoption of Technical Standards	Action 10	2			**2**
Data input measures and protections for personal data and intellectual property	Action 11	1	2	1	**4**
	Total	**27**	**43**	**16**	**86**
		31.4%	**50%**	**18.6%**	

SaferAI

Definitions and clarifications

Acronym/ Initialism	Meaning	AI Act reference example	Refers to
AIA	EU AI Act	(110)	Recital 110
AIO	EU AI Office	55(1)(b)	Article 55, paragraph 1, point (b)
CoC	G7 HAIP Code of Conduct	Annex XI(Sec 2)(1)(a)	Annex XI, Section 2, paragraph 1, point (a)
EC	European Commission	Subparagraphs are not directly referenced	
GPAI	General-purpose AI		
HAIP	G7 Hiroshima AI Process		

SaferAI

High-level analysis

Subject of Code of Conduct Action	Findings
Intro: General and Introduction	• The Act is much more clear about the AI models and systems the CoC and Act apply to, about the values concerned, and how providers may comply with the Act through different conformity mechanisms.
1. Risk management and evaluations	• Both the CoC and AIA are broadly similar regarding the need for a risk management process, model evaluations and documentation in order to mitigate risks throughout the AI lifecycle. • There are important differences, such as the AIA consideration of concepts such as the “state of the art”, and risks along the AI value chain, while the CoC has a unique focus on collaboration and research. • They are not exactly the same in some details, but are mostly common or at least compatible.
2. Identify and mitigate vulnerabilities	• Action 2 focuses on the monitoring of vulnerabilities and incidents and on concrete measures to do so. • Action 2 is relatively explicit, while the AI Act is less so. That said, most of these explicit points can reasonably be inferred from the Act. At the same time, while the CoC is relatively specific, the Act is more broad. Thus, the points in Action 2, in conjunction with other measures, can be seen as specific measures to fulfil the broader requirements of the Act. • There is low commonality regarding the facilitation and incentivisation of finding issues and vulnerabilities, in particular through contests or prizes, though bug bounties may be appropriate. • The least commonality is found where Action 2 states that orgs should use AI systems as intended.
3. Transparency and documentation	• Action 3 requires public reporting while the AI Act Art. 53 (and Annexes XI and XII that specify the content of the reports) require three kinds of reports that are targeted to different audiences: one that can be requested by the AI Office and by national authorities, one that should be provided to downstream providers and one public report about the content used for training. • According to the CoC, reports should be kept up to date, and published for all significant releases, while according to the AIA, they should only be kept up-to-date. • The CoC report and the AIA reports share one topic with high commonality: a technical documentation of the model • The CoC report and the AIA reports share some contents with medium commonality: detail on evaluations red-teaming,

	discussion and assessment of risks to safety or society, instructions of use (only the CoC report and the AIA report to downstream providers share that) and model capacities • There is no topic from the CoC report that is completely ignored in the AI Act reports.
4. Incident reporting and information sharing	• Action 4 requires public reporting while the AI Act Art. 53 (and Annexes XI and XII that specify the content of the reports) require three kinds of report that are targeted to different audiences: one that can be requested by the AI Office and by national authorities, one that should be provided to downstream providers and one public report about the content used for training • Incident reports are required by both CoC and AIA with high commonality. • The CoC report and the AIA reports share some contents with medium commonality: evaluation reports, information on security and safety risks and information on dangerous intended or unintended capabilities • The CoC report contains contents that do not have to be included in the AIA reports: Information on attempts by AI actors to circumvent safeguards
5. Risk management framework	• Action 5 is explicit about providers having “AI governance and risk management policies”. However, the AI Act does not have a binding requirement for such policies. That said, 55(1)(b)’s requirement to “assess and mitigate possible systemic risks” is complemented by Recital 114 stating that providers, “should continuously assess and mitigate systemic risks, including for example by putting in place risk-management policies, such as accountability and governance processes.” • While most of the detailed points in Action 5 are not in the AI Act, they are best practices in risk management and should be followed.
6. Cybersecurity	• Overall, Action 6 and the Act have much in common. The overall concerns are the same, and some of the text is exactly the same. Where one of the two is more specific, these more detailed requirements can be reasonably inferred from the other text.
7. Content authentication and provenance mechanisms	• Both CoC and AIA require content authentication and provenance mechanisms. • CoC prescribes tools or APIs to allow users to determine if particular content was created with their advanced AI system, such as via watermarks; AIA includes watermarks as a possible technique, but additionally mentions metadata identifications and cryptographic methods • CoC prescribes collaboration and investments in research, as appropriate, to advance the state of the field of content authentication and provenance mechanisms; the AIA does not contain the obligation for providers to advance the field, it just prescribes the application of such mechanisms
8. Investments in research and	• Commonality: The objectives of the research investments prescribed by the CoC are in line with the objectives of the

mitigation measures	AIA. ● Discrepancy: The CoC requires research investments and collaboration to promote those objectives. The AIA does not require such research investments and collaboration and does not require providers to share research and best practices on risk mitigation. However, there are some ways for GPAI model providers to share their research with institutions, namely through the advisory forum and the drawing-up of the codes of practice.
9. Developing AI for the benefit of the public	● CoC and AIA have similar (or at least compatible) end-goals regarding developing AI for the benefit of the public, but the details of their scope and who is responsible differ
10. Development and adoption of technical standards	● Both the CoC and the Act encourage organisations to participate in the development and use of content provenance methods, along with other methodologies and measures more broadly.
11. Data input measures and protections for personal data and intellectual property	● The CoC prescribes measures to manage data quality in order to mitigate against harmful biases; in the AIA, such measures are implied to be part of the mitigation of systemic risks and their sources. ● With respect to privacy, personal data, copyright, and intellectual property, the AIA (especially when combined with Union law) is much more detailed, explicit and comprehensive in its requirements. ● Both require assurance of privacy and compliance with other legal frameworks; the AIA explicitly mentions the need to comply with *Directive (EU) 2019/790.*

SaferAI

Cross-reference Table

G7 Hiroshima AI Process Code of Conduct	EU AI Act (GPAI focus)
Intro: General and Introduction	Protected values, conformity mechanisms, and definition of advanced AI / general-purpose AI (with systemic risk). • (112): Notifying AIO of models w/ systemic risk • (113): If EC becomes aware of a model with systemic risk, EC can designate it as such • 1(1): Subject matter • 3(63): 'general-purpose AI model' • 3(64): 'high-impact capabilities' • 3(65): 'systemic risk' • 40: Harmonised standards and standardisation deliverables • 41: Common specifications • 51: Classification rules for general-purpose AI model with systemic risk • 52: GPAI model with systemic risk classification procedure • 56: Codes of practice • XIII: Criteria for the designation of general-purpose AI models with systemic risk referred to in Article 51
1. Risk management and evaluations	• (110): List of systemic risks • (114): Details on Article 55, evaluations and risk management • (116): Details on Codes of Practice drafting, risk taxonomy, and risk assessment and mitigation measures • 3(2): 'risk' • 3(65): 'systemic risk' • 53(1)(a): Draw up technical documentation • 53(1)(b): Transparency information for downstream • 53(3): Providers of GPAI models shall cooperate as necessary • 55(1)(a): Perform model evaluation • 55(1)(b): Assess and mitigate systemic risks • 55(1)(d): Cybersecurity protection • 56(2)(c): Systemic risk identification • 56(2)(d): Measures, procedures and modalities for the assessment and management of systemic risks • 92(1): AIO may conduct model evaluations

	• 92(1)(b): AIO may investigate systemic risks of GPAI models with systemic risk • 92(2): AIO may appoint experts to conduct evaluations on its behalf • XI(Section 1)(2)(a-e): Relevant information of the process for the development • XI(Section 2)(1): Description of the evaluation strategies • XI(Section 2)(2): Description of adversarial testing and model adaptations • XIII(b): the quality or size of the data set, for example measured through tokens
2. Identify and mitigate vulnerabilities	• (110): List of systemic risks • 53(1)(b):Transparency information for downstream • 53(1)(b)(i): Enable providers of AI systems to understand the capabilities and limitations of GPAI models • 53(1)(b)(ii): Link to Annex XII • 55(1)(b): Assess and mitigate systemic risks • 55(1)(c): Track, document, and report serious incidents and corrective measures • 55(1)(d): Cybersecurity protection • 56(2)(d): Measures, procedures and modalities for the assessment and management of systemic risks • XII(1)(b): Acceptable use policies
3. Transparency and documentation	• 51(1): GPAI model with systemic risk classification conditions • 51(1)(a): Classification based on high impact capabilities • 51(1)(b): Classification based on a decision of the Commission, *ex officio* • 51(2): 10^25 FLOP • 51(3): Amend the thresholds and supplement benchmarks and indicators • 52(1): provider shall notify the Commission of a model with systemic risk • 53(1)(a): Draw up technical documentation • 53(1)(b): Provide information and documentation to downstream providers • 53(1)(b)(i): Enable providers of AI systems to understand the capabilities and limitations of GPAI models • 53(1)(d): Make publicly available a summary of training content • 55(1)(a): Perform model evaluation • 55(1)(b): Assess and mitigate systemic risks • 56(2)(d): Measures, procedures and modalities for the assessment and management of systemic risks • 91: Power to request documentation and information • XI(Section 1)(1)(a): tasks the model is intended to perform and info on AI systems it can be integrated in • XI(Section 1)(2)(a): the technical means required for the GPAI model to be integrated in AI systems • XI(Section 2)(1): Description of the evaluation strategies

	• XI(Section 2)(2): Description of adversarial testing and model adaptations • XII(1)(a): tasks the model is intended to perform and info on AI systems it can be integrated in • XIII(e): benchmarks and evaluations of capabilities of the model
4. Incident reporting and information sharing	• (114): Details on Article 55, evaluations and risk management • (139): Objectives of the AI regulatory sandboxes • (167): Confidentiality of information and data obtained in carrying out tasks • (172): Whistleblower protection • 3(49): 'serious incident' • 40(3): Harmonised standards and standardisation deliverables • 52(1): provider shall notify the Commission of a model with systemic risk • 52(6): GPAI model with systemic risk classification procedure • 53(1)(a): Draw up technical documentation • 53(1)(b)(i): Enable providers of AI systems to understand the capabilities and limitations of GPAI models • 53(1)(d): Make publicly available a summary of training content • 53(3): Cooperate with authorities • 55: Obligations of providers of general-purpose AI models with systemic risk • 56(3): Participation in the drawing-up of codes of practice • 62(1)(d): facilitate participation in the standardisation development process • 78(1): Confidentiality of information and data obtained in carrying out tasks • 78(1)(a): intellectual property rights and confidential business information or trade secrets • 91: Power to request documentation and information • XI(Section 1)(1)(a): tasks the model is intended to perform and info on AI systems it can be integrated in • XI(Section 1)(2): detailed description of the elements of the model and information on development process • XIII(e): benchmarks and evaluations of capabilities of the model
5. Risk management framework	• (28): AI misuse can contradict Union values • (110): List of systemic risks • (114): Details on Article 55, evaluations and risk management • 3(65): 'systemic risk' • 4: AI literacy • 9(5)(c): Training to deployers • 14(5): Human oversight • 26(2): Deployers - training and authority of human oversight

	• 55(1)(b): Assess and mitigate systemic risks • 56(2)(d): Measures, procedures and modalities for the assessment and management of systemic risks • 60(4)(j): [Testing of high-risk outside of sandboxes - training and authority] • Annex XI(Section 2)(1): Description of the evaluation strategies • Annex XI(Section 2)(2): Description of adversarial testing and model adaptations
6. Cybersecurity	• (115): Details on Article 55: Track and report serious incidents; Cybersecurity • 55(1)(b): Assess and mitigate systemic risks • 55(1)(d): Ensure adequate cybersecurity protection • 56(2)(d): Measures, procedures and modalities for the assessment and management of systemic risks
7. Content authentication and provenance mechanisms	• (133): Watermarking and other techniques • 50(1): Natural persons are informed that they are interacting with an AI system • 50(2): Mark outputs as artificially generated (for synthetic content)
8. Investments in research and mitigation measures	• (20): AI literacy • (48): High-risk AI and fundamental rights • (110): List of systemic risks • (113): If EC becomes aware of a model with systemic risk, EC can designate it as such • (115): Details on Article 55: Track and report serious incidents; Cybersecurity • (116): Details on Codes of Practice drafting, risk taxonomy, and risk assessment and mitigation measures • (133): Watermarking and other techniques • (148): Governance framework of AI Office, Board, scientific panel, and advisory forum • (149): Details about the AI Board • (150): Details about the advisory forum • (164): AIO to monitor compliance • (174): EC report every 4yrs • (176): Objective of AI Act better achieved at Union level • 1(1): Subject matter • 38(3): EC shall provide for the exchange of knowledge and best practices between notifying authorities • 53(1)(b): Provide information and documentation to downstream providers • 55(1)(b): Assess and mitigate systemic risks • 56(2)(d): Measures, procedures and modalities for the assessment and management of systemic risks • 56(3): Participation in the drawing-up of codes of practice • 65(2): AI Board and its meetings

	• 66(b): AI Board to collect and share expertise and best practices among Member States • 68(2)(b): Scientific panel experts to have independence from any provider of AI systems or GPAI models • 68(3): The scientific panel shall advise and support the AI Office • 112(6): EC report on development of standards on energy-efficient development of GPAI models • XI(Section 1)(2)(e): known or estimated energy consumption of the model
9. Developing AI for the benefit of the public	• (8): A Union legal framework laying down harmonised rules on AI is needed • (165): Voluntary codes of conduct for non-high-risk AI systems • 1(1): Subject matter • 4: AI literacy • 53(3): Cooperate with authorities • 57(5): AI regulatory sandboxes • 66(f): AI Board to support the Commission in promoting AI literacy • 67(1): Advisory forum to AI Board and EC • 67(2): Advisory forum to AI Board and EC • 95(2)(c): Codes of conduct for voluntary application of specific requirements
10. Development and adoption of technical standards	• 40(3): Harmonised standards and standardisation deliverables • 50(1): Natural persons are informed that they are interacting with an AI system • 50(2): Mark outputs as artificially generated (for synthetic content) • 56(3): Participation in the drawing-up of codes of practice • 62(1)(d): facilitate participation in the standardisation development process
11. Data input measures and protections for personal data and intellectual property	• (28): AI misuse can contradict Union values • (105): Copyright applies, providers need permission from copyright holders • (106): Providers need to put in place a policy to comply with copyright law • (107): Detailed summary of training content • (108): AIO to verify copyright compliance, but not work-by-work • (110): Systemic risk list • 2(7): Union law on protection of personal data, privacy, etc. applies to personal data processed • 3(65): ‘systemic risk’ • 53(1)(c): Policy to comply with Union law on copyright • 53(1)(d): Provide summary of training content • 55(1)(b): Assess and mitigate systemic risks • XI(Section 1)(2)(c): information on the data used for training

SaferAI

Detailed Analysis

Legend:
High or complete commonality
Some commonality
Little or no commonality

General and Introduction

High-level findings:

- Because it is EU law, the Act is much more clear about the AI models and systems that the Act applies to, about the values concerned, and how providers may comply with the Act through different conformity mechanisms.

#	Point of Comparison	Hiroshima Process Code of Conduct (Introduction/Preamble)	EU AI Act (GPAI focus)	Comments
1	**Values**	Intro: "the [HAIP CoC] for Organizations Developing Advanced AI Systems aims to promote safe, secure, and trustworthy AI worldwide" Intro: "While harnessing the opportunities of innovation, organizations should respect the rule of law, human rights, due process, diversity, fairness and non-discrimination, democracy, and human- centricity, in the design, development and deployment of advanced AI systems." Intro: [international human rights law]	1(1): "The purpose of this Regulation is to…promote the uptake of…trustworthy [AI], while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union"	*AIA is more stringent.* *Generally compatible, but not all fundamental rights enshrined in the Charter are covered by international human rights law (at least to our knowledge)*

		"seek to ensure the trustworthiness, safety and security of systems throughout their entire lifecycle"		
2	**Conformity mechanisms**	"Different jurisdictions may take their own unique approaches to implementing these actions in different ways." *OECD Informal Task Force Reporting Framework*	40: *Harmonised standards* 41: *Common specifications* 56: *Codes of practice*	*AIA is more specific.* *As a conformity mechanism for the AIA, the Codes of Practice (and standards) have specific expectations set out in the legal text.* *The CoC does not have such expectations.*
3	**Definition of advanced AI / general-purpose AI**	"the most advanced AI systems, including the most advanced foundation models and generative AI systems (henceforth "advanced AI systems")."	3(63): [definition of] 'general-purpose AI model' 3(64): [definition of] 'high-impact capabilities' 3(65): [definition of] 'systemic risk' 51: "Classification of general-purpose AI models as general-purpose AI models with systemic risk" 52: "Procedure [for the classification of general-purpose AI models as general-purpose AI models with systemic risk]" (112): "It is also necessary to clarify a procedure for the classification of a	*The CoC doesn't clearly define the category "most advanced AI systems", while the Act has an explicit focus on defining "general-purpose AI models" and "general-purpose AI models with systemic risk" with multiple definitions, articles, recitals, and an annex.*

			general-purpose AI model with systemic risks.” (113): “If the Commission becomes aware of the fact that a general-purpose AI model meets the requirements to classify as a general-purpose AI model with systemic risk…” Annex XIII: “Criteria for the designation of general-purpose AI models with systemic risk referred to in Article 51”	

SaferAI

Action 1

Risk management and evaluations

High-level findings:

- Both the CoC and AIA are broadly similar regarding the need for a risk management process, model evaluations and documentation in order to mitigate risks throughout the AI lifecycle.
- There are important differences, such as the AIA consideration of concepts such as the “state of the art”, and risks along the AI value chain, while the CoC has a unique focus on collaboration and research.
- They are not exactly the same in some details, but are mostly common or at least compatible.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 1)	EU AI Act (GPAI focus)	Comments
1	**Level of effort, level of mitigation**	Intro: “Organizations should follow these actions in line with a risk-based approach.” “Take appropriate measures” “implementing appropriate mitigation to address identified risks and vulnerabilities.” “seek to ensure the trustworthiness, safety and security of systems throughout their entire lifecycle so that they do not pose unreasonable risks.”	56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): “**the measures, procedures and modalities for the** assessment and **management of the systemic risks** at Union level, including the documentation thereof, which **shall be proportionate to the risks**, take into consideration their severity and probability and take into account the specific challenges of tackling those risks in light of the possible ways in which such risks may emerge and materialise along the AI value chain.”	*Need to consider difference in meaning between CoC use of "appropriate" and "unreasonable" vs AIA use of "proportionate"*
2	**Life cycle**	“Take appropriate measures throughout the development of	55(1)(b): “assess and mitigate possible systemic risks at Union level, including	*Both cover the entire life cycle (development, placing on*

		advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle" "Testing and mitigation measures, should, for example, seek to ensure the trustworthiness, safety and security of systems throughout their entire lifecycle so that they do not pose unreasonable risks." "[Testing should] be performed at several checkpoints throughout the AI lifecycle in particular before deployment and placement on the market"	their sources, that may stem from the **development, the placing on the market, or the use** of general-purpose AI models with systemic risk;" (114): "perform the necessary model evaluations **in particular prior to its first placing on the market**"	*market, use)*
3	**Risk management process**	"Identify, evaluate, and mitigate risks"	55(1)(b): "assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;"	*Same, assuming that "assess" consists of "identify and evaluate", which it typically does in ISO risk management standards.*
4	**Evaluation methodologies**	"combination of methods for evaluations"	Annex XI(Sec 2)(1): "A detailed description of the **evaluation strategies**, including evaluation results, on the basis of **available public evaluation protocols and tools or otherwise of other evaluation methodologies**. Evaluation strategies	*Broadly similar*

			shall include evaluation criteria, metrics and the methodology on the identification of limitations."	
5	**Internal vs external testing**	"Diverse internal and independent external testing"	(114): "Internal or independent external testing" 92(1): "The **AI Office**, after consulting the Board, **may conduct evaluations** of the general-purpose AI model concerned: [...] (b): **to investigate systemic risks** at Union level of general-purpose AI models with systemic risk, in particular following a qualified alert from the scientific panel in accordance with Article 90(1), point (a)." 92(2): "**The Commission may decide to appoint independent experts to carry out evaluations on its behalf**, including from the scientific panel established pursuant to Article 68. Independent experts appointed for this task shall meet the criteria outlined in Article 68(2)." Annex XI(Sec 2)(2): "Where applicable, a detailed description of the measures put in place for the purpose of conducting **internal and/or external adversarial testing** (e.g. red teaming), model adaptations, including alignment and fine-tuning."	*CoC uses "and" and uses the term "diverse", while AIA uses "or".* *AIA empowers AIO to conduct evaluations, and to appoint independent experts to do so on its behalf.*

6	**Red-teaming**	"This includes employing diverse internal and independent external testing measures, through a combination of methods for evaluations, such as **red-teaming**"	55(1)(a): "perform model evaluation in accordance with standardised protocols and tools reflecting the state of the art, including **conducting** and documenting **adversarial testing of the model** with a view to identifying and mitigating systemic risks;" Annex XI(Sec 2)(2): "Where applicable, a detailed description of the measures put in place for the purpose of conducting internal and/or external **adversarial testing (e.g. red teaming)**, model adaptations, including alignment and fine-tuning."	*Red-teaming is an example of testing in both the CoC and AIA.*
7	**State of the art**	"**Take appropriate measures** throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle." "This includes employing **diverse** internal and independent external **testing measures**"	55(1)(a): "perform model evaluation in accordance with **standardised protocols and tools reflecting the state of the art**, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating systemic risks;"	*"State of the art" not mentioned in the CoC.*
8	**Traceability**	"developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development."	XI(Sec 1)(2): "A detailed description of the elements of the model referred to in point 1, and relevant information of the process for the development, including the following elements:	*Both require traceability, but the AIA is more explicit and detailed.* *The AIA Annexes have explicit*

			(a) the technical means (e.g. instructions of use, infrastructure, tools) required for the general-purpose AI model to be integrated in AI systems; (b) the design specifications of the model and training process, including training methodologies and techniques, the **key design choices including the rationale and assumptions made**; what the model is designed to optimise for and the relevance of the different parameters, as applicable; (c) **information on the data used for training, testing and validation, where applicable, including the type and provenance of data and curation methodologies** (e.g. cleaning, filtering, etc.), the number of data points, their scope and main characteristics; **how the data was obtained and selected as well as all other measures to detect the unsuitability of data sources and methods to detect identifiable biases, where applicable**; (d) the computational resources used to train the model (e.g. number of floating point operations), training time, and **other relevant details related to the training**; (e) known or estimated energy consumption of the model." XI(Sec 2)(1): "A detailed description of the evaluation strategies, including	*traceability requirements regarding design choices, processes, and data used.*

			evaluation results, on the basis of available public evaluation protocols and tools or otherwise of other evaluation methodologies. Evaluation strategies shall include evaluation criteria, metrics and the methodology on the identification of limitations." XIII: "Criteria for the designation of general-purpose AI models with systemic risk referred to in Article 51" (b): "**the quality or size of the data set**, for example measured through tokens;"	
9	**Documentation**	"[Testing and mitigation] measures should be documented and supported by regularly updated technical documentation"	53(1)(a): "draw up and **keep up-to-date the technical documentation of the model**, including its training and **testing** process and the results of its evaluation…" 53(1)(b): "draw up, **keep up-to-date** and make available information and **documentation to providers of AI systems** who intend to integrate the general-purpose AI model into their AI systems…" 55(1)(a): "...conducting and **documenting** adversarial testing of the model…" 55(1)(b): "...assess and mitigate	*Both require testing and mitigation documentation, AIA is more explicit and detailed.* *Art. 53 requires testing documentation to be kept up to date. Mitigation documentation is not explicitly stated, but is generally required for risk management and can be seen as "proportionate to the risks" (Art. 56).*

			possible systemic risks…" 56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): "the measures…for the assessment and management of the systemic risks…**including the documentation thereof**, which shall be proportionate to the risks…"	
10	**Secure environments**	"testing should take place in secure environments"	55(1)(a): "perform model evaluation in accordance **with standardised protocols and tools reflecting the state of the art**, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating systemic risks;" 55(1)(d): "ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model."	*Equivalent, assuming that the AIA considers the testing environment to be part of the physical infrastructure of the model*
11	**Testing supports risk identification and mitigation**	"[Testing should take place] to identify risks and vulnerabilities, and to inform action to address the identified AI risks to security, safety and societal and other risks, whether accidental or intentional"	55(1)(a): "**perform model evaluation** in accordance with standardised protocols and tools reflecting the state of the art, including conducting and documenting adversarial testing of the model **with a view to identifying and mitigating systemic risks**;"	*Equivalent* *CoC explicitly adds "whether accidental or intentional"*

12	**Risk taxonomy**	*Seven top-level risks listed in Action 1*	Recital 110: [List of risks, along with hazards and hazardous situations] Recital 116: "codes of practice should help to establish a risk taxonomy of the type and nature of the systemic risks…including their sources" 3(65): "'systemic risk' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable **negative effects on public health, safety, public security, fundamental rights, or the society as a whole**, that can be propagated at scale across the value chain;"	*All risks in Action 1 are in Recital 110. However, Recital 110 lists more risks, as well as their nature, contributing factors, ,* *"Useful defensive [cyber] applications" is noted in the CoC but not in the AIA. Same for "non-state actors" under CBRNE risks.* *CoC does not require further risk taxonomy.*
13	**Collaboration**	"Organizations commit to work in collaboration with relevant actors across sectors, to assess and adopt mitigation measures to address these risks, in particular systemic risks."	53(3): "Providers of general-purpose AI models shall cooperate as necessary with the Commission and the national competent authorities in the exercise of their competences and powers pursuant to this Regulation" 92(1): "The AI Office, after consulting the Board, may conduct evaluations of the general-purpose AI model concerned:" 92(2): "The Commission may decide to appoint independent experts to carry	*AIA does not require collaboration with "relevant actors across sectors", except perhaps for when the AIO exercises its power to conduct evaluations and delegates this to a third party.*

			out evaluations on its behalf, including from the scientific panel established pursuant to Article 68. Independent experts appointed for this task shall meet the criteria outlined in Article 68(2)."	
14	**Research**	"Organizations making these commitments should also endeavor to advance research and investment on the security, safety, bias and disinformation, fairness, explainability and interpretability, and transparency of advanced AI systems and on increasing robustness and trustworthiness of advanced AI systems against misuse."	*Not found*	*AIA does not require organisations to advance research and investment*
15	**Value chain**	*Not found*	56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): "take into account the specific challenges of tackling those risks in light of the **possible ways in which such risks may emerge and materialise along the AI value chain**."	*Action 3 includes the consideration of deployers and users, but only with respect to transparency reports.*
16	**Definition of risk**	*Not found*	3(2): "'risk' means the combination of the probability of an occurrence of harm and the severity of that harm;"	*CoC appears to only refer to risks of harm. Also, the term "mitigatate" is typically used in risk management when addressing risk of harm, while*

				the term "treatment" is typically used when addressing risk that could result in positive or negative outcomes. That said, neither "severity" nor "probability" appear in the CoC. *The AIA is quite clear on the need to address risk of harm and both its probability and severity.*
17	**Risk sources (hazards)**	*Not found*	55(1)(b): "assess and mitigate possible systemic risks at Union level, **including their sources**, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" 56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (c): "the identification of the type and nature of the systemic risks at Union level, **including their sources**, where appropriate;"	*Neither "risk source(s)" nor "hazard(s)" appear in the CoC.* *However, risk source and hazard identification is typically necessary for risk management.*

SaferAI

Action 2

Identify and mitigate vulnerabilities

High-level findings:

- Action 2 focuses on the monitoring of vulnerabilities and incidents and on concrete measures to do so.
- Action 2 is relatively explicit, while the AI Act is less so. That said, most of these explicit points can reasonably be inferred from the Act. At the same time, while the CoC is relatively specific, the Act is more broad. Thus, the points in Action 2, in conjunction with other measures, can be seen as specific measures that could help fulfil the broader requirements of the Act.
- There is low commonality regarding the facilitation and incentivisation of finding issues and vulnerabilities, in particular through contests or prizes, though bug bounties may be appropriate.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 2)	EU AI Act (GPAI focus)	Comments
1	**Intended use/purpose**	"Organizations should use…AI systems as intended"	53(1)(b): "**draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems**. Without prejudice to the need to observe and protect intellectual property rights and confidential business information or trade secrets in accordance with Union and national law, the information and documentation shall: (i) enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to this Regulation; and	*CoC only.* *Intended use is only in the requirements for high-risk AI systems, not those for GPAI models. However, GPAI model providers have transparency obligations which include documenting the acceptable use policies applicable.*

			(ii) **contain, at a minimum, the elements set out in Annex XII**;" Annex XII(1)(b): "**the acceptable use policies applicable**;"	
2	**Life cycle**	"Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market."	(110): "...In particular, international approaches have so far identified the need to **pay attention to risks from potential intentional misuse**..." 55(1)(b): "**assess and mitigate possible systemic risks at Union level, including their sources, that may stem from** the development**, the placing on the market, or the use** of general-purpose AI models with systemic risk;" 55(1)(c): "**keep track of**, **document**, **and report**, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents and possible corrective measures to address them;**"	*Identifying and mitigating vulnerabilities and misuse is appropriate, even necessary, to comprehensively assess possible risks, keep track of serious incidents, and ensure an adequate level of cybersecurity protection.* *CoC Action 2 appears to limit these efforts to the post-deployment lifecycle stage, whereas the AIA requires them across the lifecycle.*
3	**Corrective action**	"Take appropriate action to address [vulnerabilities, incidents, emerging risks and misuse after deployment]"	(110): "...In particular, international approaches have so far identified the need to **pay attention to risks from potential intentional misuse**..." 55(1)(b): "assess and **mitigate possible systemic risks** at Union level,	*CoC is slightly more explicit.* *Addressing vulnerabilities, emerging risks, and misuse is an appropriate risk mitigation measure.*

			including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" 55(1)(c): "**keep track of**, **document**, **and report**, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents and possible corrective measures to address them;**"	*Serious incidents must be reported under the Act.*
4	**Vulnerabilities**	"[Monitor for] vulnerabilities"	55(1)(b): "**assess** and mitigate **possible systemic risks** at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" 55(1)(c): "**keep track of**, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;" 55(1)(d): "**ensure an adequate level of cybersecurity protection** for the general-purpose AI model with systemic risk and the physical infrastructure of the model."	*CoC is more explicit.* *Monitoring for vulnerabilities is in high-risk AI system requirements, but not explicitly for GPAI models.* *That said, monitoring for vulnerabilities is appropriate, even necessary, to comprehensively assess possible risks, keep track of serious incidents, and ensure an adequate level of cybersecurity protection.*

5	**Incidents**	"[Monitor for] incidents"	55(1)(c): "**keep track of**, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;"	*Equivalent, with the reasonable inference that monitoring for incidents is necessary to keep track of them.*
6	**Emerging risks**	"[Monitor for] emerging risks"	55(1)(c): "**keep track of**, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;" 56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): "the measures, procedures and modalities for the assessment …of the systemic risks shall…take into account the specific challenges of tackling those risks **in light of the possible ways in which such risks may emerge and materialise along the AI value chain.**"	*Equivalent, assuming that assessing emerging risks stemming from use of a model includes monitoring for incidents, which is reasonably inferred from 55(1)(c).*
7	**Misuse**	"[Monitor for] misuse"	(110): "...In particular, international approaches have so far identified the need to **pay attention to risks from potential intentional misuse**..."	*Misuse is in a recital. Monitoring for misuse risks is likely necessary in order to keep track of serious incidents.*

			55(1)(c): “**keep track of**, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;”	
8	**Facilitating third-parties**	“Facilitating third-party and user discovery and reporting of issues and vulnerabilities after deployment such as through bounty systems, contests, or prizes to incentivize the responsible disclosure of weaknesses”	55(1)(c): “**keep track of**, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;” 55(1)(d): “**ensure an adequate level of cybersecurity protection** for the general-purpose AI model with systemic risk and the physical infrastructure of the model.”	*CoC is more explicit.* *In theory, under the Act, it is in the interests of the providers to find all issues and vulnerabilities to fulfil their incident reporting and cybersecurity requirements. One way to do so is through facilitation and incentivisation. However, no explicit mention of facilitation or incentivisation was found in the Act, and contests or prizes could be seen as outside the intention of the Act.*
9	**Documentation**	“Maintain appropriate documentation of reported incidents”	55(1)(c): “keep track of, **document**, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;”	*AIA appears more broad.* *CoC may technically be limited to documenting reported incidents, not all incidents.*
10	**Mitigation**	“Mitigate the identified risks and vulnerabilities, in collaboration with other stakeholders”	55(1)(b): ”assess and **mitigate possible systemic risks** at Union level, including their sources, **that may stem from** the development, the placing on	*AIA appears more broad.* *CoC appears to require “collaboration”, while it seems the*

			the market, or **the use of general-purpose AI models with systemic risk**;"	*AIA leaves it to the providers. CoC also uses the term "vulnerabilities", which could reasonably be assumed to be risks (or at least hazards or risk sources).* *CoC may lean towards only reported incidents here as well, rather than "possible systemic risks", though it is unclear.*
11	**Reporting mechanisms for stakeholders**	"Mechanisms to report vulnerabilities, where appropriate, should be accessible to a diverse set of stakeholders."	55(1)(c): "keep track of, **document**, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them;" 55(1)(d): "**ensure an adequate level of cybersecurity protection** for the general-purpose AI model with systemic risk and the physical infrastructure of the model."	*CoC is more explicit.* *In theory, under the Act, it is in the interests of the providers to find all issues and vulnerabilities to fulfil their incident reporting and cybersecurity requirements. However, no explicit mention of such a reporting mechanism for stakeholders was found in the Act.*

SaferAI

Action 3

Transparency and documentation

High-level findings:

- Action 3 requires public reporting while the AI Act Art. 53 (and Annexes XI and XII that specify the content of the reports) require three kinds of reports that are targeted to different audiences: one that can be requested by the AI Office and by national authorities, one that should be provided to downstream providers and one public report about the content used for training.
- According to the CoC, reports should be kept up to date, and published for all significant releases, while according to the AIA, they should only be kept up-to-date.
- The CoC report and the AIA reports share one topic with high commonality: technical documentation of the model
- The CoC report and the AIA reports share some contents with medium commonality: detail on evaluations and red-teaming, discussion and assessment of risks to safety or society, instructions for use (only the CoC report and the AIA report to downstream providers share that) and model capacities
- There is no topic from the CoC report that is completely ignored in the AI Act reports.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 3)	EU AI Act (GPAI focus)	Comments
1	**Recipient of technical documentation and information**	"Publicly report" "This should include publishing transparency reports"	53(1)(a): "draw up and keep up-to-date the technical documentation of the model, including [...] **providing it, upon request, to the AI Office and the national competent authorities**;" 53(1)(b): "draw up, keep up-to-date and make available information and documentation **to providers of AI systems** who intend to integrate the general-purpose AI model into their AI systems." 53(1)(d): "draw up and **make publicly available** a sufficiently detailed summary	*The CoC requires public reporting. The AIA requires three kinds of reports:* *1) one report upon request to the AI Office and the national competent authorities (53(1)(a))* *2) one report to providers that integrate the GPAI model. (53(1)(b))* *3) one public report about the content used for training (53(1)(d))* *Similar to action 4*

			about the content used for training of the general-purpose AI model, according to a template provided by the AI Office."	
2	**Keep documentation up-to-date**	"should be **kept up-to-date**" "This should include publishing transparency reports containing meaningful information for **all new significant releases** of advanced AI systems."	53(1)(a): "draw up and **keep up-to-date** the technical documentation of the model" [report to AIO] 53(1)(b): "draw up, **keep up-to-date** and make available information and documentation" [report to downstream providers]	*Keeping up-to-date is the same.* *Updating for all new significant releases is reasonably implied by the AIA, but publishing is not required.* *Substantial modification is considered for high-risk systems in the AIA, but not explicitly for GPAI.*
3	**Documentation of evaluations and red-teaming**[1]	"should include, for example: [...] **Details of the evaluations conducted for potential safety, security, and societal risks, as well as risks to human rights**" "should include, for example: [...] **The results of red-teaming conducted to evaluate** the model's/system's fitness for moving beyond the development stage."	53(1)(a): "draw up and keep up-to-date the technical documentation of the model, including its training and **testing process** and the **results of its evaluation**" [report to AIO] Annex XI(Section 2)(1): "A detailed description of the evaluation strategies, including evaluation results, on the basis of available public evaluation protocols and tools or otherwise of other evaluation methodologies. Evaluation strategies shall include evaluation criteria, metrics and the methodology on the identification of limitations."	*AIA is more detailed and stringent than the CoC. While the CoC simply says "details of the evaluations", the AIA act specifies which details.* *CoC prescribes evaluations for safety, security, and societal risks, as well as risks to human rights. AIA prescribes evaluations for systemic risk (55(1)(a)). Given the definition of systemic risk in 3(65), those are very similar.* *Similar to action 4*

[1] The analysis in this row is only about the commonality between the contents of the CoC report and the AIA reports, and not about commonality between the recipients of the reports. Commonality between the recipients of the CoC report and the AIA reports is already analysed in the first row of this action. High commonality in this row only implies that the contents are the same or very similar, no matter whether there is high or low commonality regarding the recipients.

			Annex XI(Section 2)(2): “Where applicable, **a detailed description of the measures put in place** for the purpose of conducting **internal and/or external adversarial testing (e.g., red teaming),** model adaptations, including alignment and fine-tuning.”	
4	**Documentation contains capacities of the model/system**[1]	“should include, for example: [...] **Capacities of a model/system** and significant limitations in performance that have implications for the domains of appropriate use”	Annex XI(Section 1): “The technical documentation referred to in Article 53(1), point (a) shall contain at least the following information as appropriate to the size and risk profile of the model: (1) A general description of the general-purpose AI model including: (a): **the tasks that the model is intended to perform**” Annex XI(Section 2)(1): “A detailed description of the evaluation strategies, including evaluation results, on the basis of available public evaluation protocols and tools or otherwise of other evaluation methodologies. Evaluation strategies shall include evaluation criteria, metrics and the methodology on the identification of limitations.” 52(1): “Where a general-purpose AI model meets the condition referred to in Article 51(1), point (a), **the relevant provider shall notify the Commission** without delay and in any event within two weeks after that requirement is met or it becomes known that it will be met.” 51(1): “A general-purpose AI model shall be	*1) The technical documentation **for the AI Office** according to Art. 53 & 55 (and Annex XI) does not require a reporting of model capacities/capabilities. There is medium commonality with two AIA sections:* *a) Annex XI requires the provider to report the tasks that the model is intended to perform. It also requires them to report a description of the methodology on the identification of limitations. It must also include the results of that identification of limitations.* *b) Art. 52(1) requires the provider to notify the EC if the GPAI model poses systemic risk, i.e., if it has high impact capabilities. However, the criterion used to classify models as having high impact capabilities is at present based on the amount of computation used for the model’s*

			classified as a general-purpose AI model with systemic risk if it meets any of the following conditions: (a) **it has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks**; (b) based on a decision of the Commission, ex officio or following a qualified alert from the scientific panel, it has capabilities or an impact equivalent to those set out in point (a) having regard to the criteria set out in Annex XIII. (2) A general-purpose AI model shall be presumed to have high impact capabilities pursuant to paragraph 1, point (a), **when the cumulative amount of computation used for its training measured in floating point operations is greater than 10^25.** (3) The Commission shall adopt delegated acts in accordance with Article 97 to amend the thresholds listed in paragraphs 1 and 2 of this Article, as well as to supplement benchmarks and indicators in light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, when necessary, for these thresholds to reflect the state of the art." Annex XIII(e): "the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to;"	*training measured. Annex XIII, in particular (e), gives some broad categories of capabilities, e.g., autonomy.*

			53(1)(b)(i): "**enable providers of AI systems to have a good understanding of the capabilities and limitations** of the general-purpose AI model and to comply with their obligations pursuant to this Regulation" Annex XII: "The information referred to in Article 53(1), point (b) shall contain at least the following: (1): A general description of the general-purpose AI model including: (a) **the tasks that the model is intended to perform** and the type and nature of AI systems into which it can be integrated;"	*2) The technical documentation **for downstream providers** should enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model (53(1)(a)). However, there is no prescription that the capabilities and limitations must be directly included in the documentation. Annex XII (Transparency information referred to in Article 53(1), point (b) - technical documentation for providers of general-purpose AI models to downstream providers that integrate the model into their AI system) says that the documentation must include the tasks that the model is intended to perform. Nothing more specific about capabilities and limitations can be found in Annex XII.* *Similar to Action 4*
5	**Discussion and assessment of risks to safety or society**[1]	"should include, for example: [...] Discussion and assessment of the model's or system's effects and risks to safety and society such as harmful bias, discrimination, threats to protection of privacy or personal data, and effects on fairness"	56(2): "The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: [...] (d) the **measures, procedures and modalities for the assessment and management of the systemic risks at Union level, including the documentation thereof**, which shall be proportionate to the	*AIA has more clarity and requires documentation of the management of risks.* *Art. 56(2)(d) says that the codes of practice should cover the documentation of risk assessment and management of systemic risk. It leaves open what should be*

			risks, take into consideration their severity and probability and take into account the specific challenges of tackling those risks in light of the possible ways in which such risks may emerge and materialise along the AI value chain." 91: "The Commission may request the provider of the general-purpose AI model concerned to provide the documentation drawn up by the provider in accordance with Articles 53 and 55, **or any additional information that is necessary for the purpose of assessing compliance of the provider with this Regulation."** 55(1): "In addition to the obligations listed in Articles 53 and 54, providers of general-purpose AI models with systemic risk shall: [...] (b): **assess** and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;"	*documented and who has access to the documentation, though it is assumed here to be the AI Office.* *Apart from that, according to Art. 91+55(1)(b), the EC can request information that is necessary to assess whether the provider does risk assessment. Please note that risk assessment in the CoC is about "risks to safety and society" whereas risk assessment in the AIA is about systemic risks at Union level. The definition of systemic risk in the AIA (Art. 3(65)) is broader than that of "risks to safety and society" in the CoC.* *Similar to action 4*
6	**Clarity of reports to enable interpretation and appropriate usage**	"Organizations should make the information in the transparency reports **sufficiently clear and understandable** to **enable** deployers and users as appropriate and relevant to	Annex XI(Section 1): "The technical documentation referred to in Article 53(1), point (a) shall contain at least the following information [...] (2) [...] (a) the technical means (e.g. **instructions of use**, infrastructure, tools) required for the	*AIA does not require reporting that addresses users in order to make sure that the model is used appropriately. It only requires reporting that addresses downstream providers for the sake of integration and compliance with the AIA.*

		interpret the model/system's output and to **enable users to use it appropriately**;"	general-purpose AI model to be integrated in AI systems;" 53(1)(b): "the information and documentation shall: (i) enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to this Regulation;" Annex XI(section 2)(1): "A **detailed** description of the evaluation strategies, including evaluation results [...]" *Nothing more specific found*	*There are requirements regarding clarity of reports in the CoC, while the AIA has similar requirements enabling a "good understanding" and "to comply with their obligations".*
7	**Technical documentation**	"transparency reporting should be supported and informed by robust documentation processes such as **technical documentation** and instructions for use**."**	53(1)(a): "draw up and keep up-to-date the **technical documentation of the model,** including its training and testing process and the results of its evaluation, which shall contain, at a minimum, the information set out in Annex XI for the purpose of providing it, upon request, to the AI Office and the national competent authorities;"	*In the AIA, the report that is targeted towards the AIO is referred to as technical documentation. Technical information that must be included is specified in Annex XI. Examples are the architecture and the number of parameters (section 1(1)(d)). The CoC does not specify which specific information the technical documentation must contain. We assume it should be similar.*
8	**Reports Contain instructions for use**	"transparency reporting should be supported and informed by robust documentation processes such as technical documentation and **instructions for use**"	Annex XI(Section 1): "The technical documentation referred to in Article 53(1), point (a) shall contain at least the following information [...] (2) [...] (a) the technical means (e.g. **instructions of**	*The AIA, refers only to those instructions of use that are required for the general-purpose AI model to be integrated in AI systems.The CoC refers to instructions for use broadly.*

			use, infrastructure, tools) required for the general-purpose AI model to be integrated in AI systems;”	

SaferAI

Action 4

Incident reporting and information sharing

High-level findings:

- Action 4 requires public reporting while the AI Act Art. 53 (and Annexes XI and XII that specify the content of the reports) require three kinds of report that are targeted to different audiences: one that can be requested by the AI Office and by national authorities, one that should be provided to downstream providers and one public report about the content used for training
- Incident reports are required by both CoC and AIA with high commonality.
- The CoC report and the AIA reports share some contents with medium commonality: evaluation reports, information on security and safety risks and information on dangerous intended or unintended capabilities
- The CoC report contains contents that do not have to be included in the AIA reports: Information on attempts by AI actors to circumvent safeguards

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 4)	EU AI Act (GPAI focus)	Comments
1	**Recipient of the shared information & reporting**	"Work towards responsible information sharing and reporting of incidents **among organizations developing advanced AI systems including with industry, governments, civil society, and academia**." "Organizations should collaborate with other organizations across the AI lifecycle to share and report relevant information **to the public**" "Organizations should also collaborate and share the aforementioned information **with**	55(1)(c): "keep track of, document, and report, without undue delay, **to the AI Office** and, as appropriate, **to national competent authorities**, relevant information about serious incidents and possible corrective measures to address them" 53(3): "shall cooperate as necessary with the **Commission and the national competent authorities**" 53(1): Providers of general-purpose AI models shall: [...] (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the	*The CoC requires public reporting. The AIA requires three kinds of reports:* *1) one report (which includes relevant information about serious incidents) **upon request** to the AI Office and the national competent authorities* *2) one report to providers that integrate the GPAI model. (same as Action 3 Item 1).* *3) one public report about the content used for training* *Similar to action 3*

		relevant public authorities, as appropriate."	general-purpose AI model, according to a template provided by the AI Office.	
2	**Incident reporting**[2]	"Work towards [...] **reporting of incidents** among organizations developing advanced AI systems including with industry, governments, civil society, and academia"	55(1)(c): "keep track of, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, **relevant information about serious incidents** and possible corrective measures to address them"	*CoC: reporting of incidents to many actors including* industry, governments, civil society, and academia*;* *AIA: reporting of relevant information about serious incidents to the AI Office*
3	**Evaluation reports**[2]	"This includes responsibly sharing information, as appropriate, including, but not limited to evaluation reports [...]"	53(1)(a): "draw up and keep up-to-date the technical documentation of the model, including its training and testing process and the **results of its evaluation**" Annex XI(Section 2)(1): "A detailed description of the evaluation strategies, including evaluation results, on the basis of available public evaluation protocols and tools or otherwise of other evaluation methodologies. Evaluation strategies shall include evaluation criteria, metrics and the methodology on the identification of limitations." Annex XI(Section 2)(2): "Where applicable, a detailed description of the measures put in place for the purpose of conducting internal and/or external adversarial testing (e.g., red	*AIA is more detailed and stringent than the CoC. While the CoC simply says "details of the evaluations", the AIA act specifies which details.* *CoC prescribes evaluations for safety, security, and societal risks, as well as risks to human rights. AIA prescribes evaluations for systemic risk (55(1)(a)). Given the definition of systemic risk in 3(65), those are very similar.* *Similar to action 3*

[2] The analysis in this row is only about the commonality between the contents of the CoC report and the AIA reports, and not about commonality between the recipients of the reports. Commonality between the recipients of the CoC report and the AIA reports is already analysed in the first row of this action. High commonality in this row only implies that the contents are the same or very similar, no matter whether there is high or low commonality regarding the recipients.

			teaming), model adaptations, including alignment and fine-tuning."	
4	**information on security and safety risks[2]**	"This includes responsibly sharing information, as appropriate, including, but not limited to [...] information on security and safety risks [...]"	91: "The Commission may request the provider of the general-purpose AI model concerned to provide the documentation drawn up by the provider in accordance with Articles 53 and 55, **or any additional information that is necessary for the purpose of assessing compliance of the provider with this Regulation.**" 55(1): "In addition to the obligations listed in Articles 53 and 54, providers of general-purpose AI models with systemic risk shall: [...] (b) assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;"	*There is no prescription to universally share information on security and safety risks. However, according to Art. 91+55, the EC can request information that is necessary to assess whether the provider does risk assessment. Please note that risk assessment in the CoC is about "risks to safety and society" whereas risk assessment in the AIA is about systemic risks at Union level. The definition of systemic risk in the AIA (Art. 3(65)) is broader than that of "risks to safety and society" in the CoC.* *Similar to action 3*
5	**information on dangerous intended or unintended capabilities[2]**	"This includes responsibly sharing information, as appropriate, including, but not limited to [...] information on [...] dangerous intended or unintended capabilities [...]."	Annex XI(1): "A general description of the general-purpose AI model including: (a) the tasks that the model is intended to perform and the type and nature of AI systems in which it can be integrated" 53(1)(b)(i): "**enable providers of AI systems to have a good understanding of the capabilities and limitations** of the general-purpose AI model and to comply with their obligations pursuant to this	*AIA report to the AI Office: Only the tasks that the model is intended to perform must be included. Nothing about dangerous capabilities. If dangerous capabilities are not intended, then they are excluded from the documentation.* *AIA report to downstream*

			Regulation" *[information for downstream providers]* *52(1): "Where a general-purpose AI model meets the condition referred to in Article 51(1), point (a), the relevant provider shall notify the Commission without delay and in any event within two weeks after that requirement is met or it becomes known that it will be met."* *Annex XIII(e): "the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to;"*	*providers: As in the report to the AI Office, the tasks that the model is intended to perform must be included. Additionally there is a more indirect prescription: The report must enable downstream providers to have a good understanding of the capabilities and limitations of the system. However, it is not specified how this should be achieved.* *Furthermore, 52(1) requires the provider to notify the EC if the GPAI model poses systemic risk, i.e., if it has high impact capabilities. However, the criterion used to classify models as having high impact capabilities is at present based on the amount of computation used for the model's training measured. Annex XIII, in particular (e), gives some broad categories of capabilities, e.g., autonomy.* *Similar to action 3*
6	**Information on attempts by AI actors to circumvent safeguards[2]**	"This includes responsibly sharing information, as appropriate, including, but not limited to [...] information on [...] attempts by AI	55(1)(c): "keep track of, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, relevant information about	*According to the AIA, only serious incidents must be reported. Serious incidents are defined by having certain*

		actors to circumvent safeguards [...]"	**serious incidents** and possible corrective measures to address them" 3(49): "'serious incident' means an incident or malfunctioning of an AI system that directly or indirectly leads to any of the following: (a) the death of a person, or serious harm to a person's health; (b) a serious and irreversible disruption of the management or operation of critical infrastructure; (c) the infringement of obligations under Union law intended to protect fundamental rights; (d) serious harm to property or the environment;" *Nothing specific found about attempts to circumvent safeguards or attempts to cause incidents*	*consequences. So, attempts do not count as serious incidents. Therefore, this CoC obligation is not covered by the AIA. However, while it's not strictly in the AIA, one could interpret that providers should still keep track of circumvention attempts, as a best practice for risk assessment.*
7	**Development of shared standards, tools, mechanisms, and best practices**	Organizations should establish or join mechanisms to develop, advance, and adopt, where appropriate, shared standards, tools, mechanisms, and best practices for ensuring the safety, security, and trustworthiness of advanced AI systems.	(139): "the AI regulatory sandboxes should aim to enhance legal certainty for innovators and the competent authorities' oversight and understanding of the opportunities, emerging risks and the impacts of AI use, to facilitate regulatory learning for authorities and undertakings, including with a view to future adaptions of the legal framework, to support cooperation and the sharing of best practices with the authorities involved in the AI regulatory sandbox"	*The AIA encourages the sharing of best practices in regulatory sandboxes.* *While the AIA does not strictly require providers to work to develop standards, it is in their interest to do so and they are, in theory, invited and facilitated to do so by authorities.*

			40(3): “The participants in the standardisation process shall seek to promote investment and innovation in AI, including through increasing legal certainty, as well as the competitiveness and growth of the Union market, to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are consistent with Union values, fundamental rights and interests, and to enhance multi-stakeholder governance ensuring a balanced representation of interests and the effective participation of all relevant stakeholders…” 56(3): “The AI Office may invite all providers of general-purpose AI models…to participate in the drawing-up of codes of practice.” 62(1)(d): “[Member States shall] facilitate the participation of SMEs and other relevant stakeholders in the standardisation development process.”	
8	**Across the AI lifecycle**	“ensuring appropriate and relevant documentation and transparency across the AI lifecycle”	55(1)(b): “assess and mitigate possible systemic risks at Union level, including their sources, **that may stem from the development, the placing on the market, or the use** of general-purpose AI models with systemic risk;” Annex XI(Section 1)(2): “A detailed description of the elements of the model	*Annex XI(section 1)(2) covers elements from the whole lifecycle: design, training, testing, validation, integration, and deployment* *Furthermore, risk assessment of systemic risks must be done*

			referred to in point 1, and relevant information of the process for the development, including the following elements: (a) the technical means (e.g. instructions of use, infrastructure, tools) required for the general-purpose AI model **to be integrated in AI systems**; (b) the **design specifications** of the model and **training process**, including **training methodologies and techniques**, the **key design choices** including the rationale and assumptions made; what the model is designed to optimise for and the relevance of the different parameters, as applicable; (c) information on the **data used for training, testing and validation**, where applicable, including the type and provenance of data and curation methodologies (e.g. cleaning, filtering, etc.), the number of data points, their scope and main characteristics; how the data was obtained and selected as well as all other measures to detect the unsuitability of data sources and methods to detect identifiable biases, where applicable; (d) the computational resources used to train the model (e.g. number of floating point operations), training time, and other relevant details related to the training; (e) known or estimated **energy consumption** of the model.	*during all stages of the lifecycle*

			With regard to point (e), where the energy consumption of the model is unknown, the energy consumption may be based on information about computational resources used.”	
9	**Advanced AI systems that cause significant risks to safety and society**	“in particular for advanced AI systems that cause significant risks to safety and society”	55: “(1) In addition to the obligations listed in Articles 53 and 54, providers of general-purpose AI models with systemic risk shall: (a) perform model evaluation in accordance with standardised protocols and tools reflecting the state of the art, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating systemic risks; (b) assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk; (c) keep track of, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, relevant information about serious incidents and possible corrective measures to address them; (d) ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model. 2. Providers of general-purpose AI models	*Both focus more on riskier models* *CoC: "in particular" for risks to safety and health* *AIA: more exhaustive reporting for GPAI models with systemic risk*

			with systemic risk may rely on codes of practice within the meaning of Article 56 to demonstrate compliance with the obligations set out in paragraph 1 of this Article, until a harmonised standard is published. Compliance with European harmonised standards grants providers the presumption of conformity to the extent that those standards cover those obligations. Providers of general-purpose AI models with systemic risks who do not adhere to an approved code of practice or do not comply with a European harmonised standard shall demonstrate alternative adequate means of compliance for assessment by the Commission. 3. Any information or documentation obtained pursuant to this Article, including trade secrets, shall be treated in accordance with the confidentiality obligations set out in Article 78."	
10	**Collaboration**	"Organizations should collaborate with other organizations across the AI lifecycle to share and report relevant information to the public with a view to advancing safety, security and trustworthiness of advanced AI systems. Organizations should also collaborate and share the aforementioned information with	53(3): "Providers of general-purpose AI models shall cooperate as necessary with the Commission and the national competent authorities in the exercise of their competences and powers pursuant to this Regulation." (114) In addition, providers of general-purpose AI models with systemic risks should continuously assess and mitigate systemic risks, including for example by putting in	*AIA:* *1) cooperation "as necessary" with EC and national competent authorities* *2) cooperating with relevant actors along the AI value chain in risk mitigation and assessment (see Action 3, entry 6 on the reporting of risk assessment)*

		relevant public authorities, as appropriate."	place risk-management policies, such as accountability and governance processes, implementing post-market monitoring, taking appropriate measures along the entire model's lifecycle and **cooperating with relevant actors along the AI value chain**.	
11	**Safeguarding intellectual property rights**	"Such reporting should safeguard intellectual property rights."	78(1): "The Commission, market surveillance authorities and notified bodies and any other natural or legal person involved in the application of this Regulation shall, in accordance with Union or national law, respect the confidentiality of information and data obtained in carrying out their tasks and activities in such a manner as to protect, in particular: (a) the intellectual property rights and confidential business information or trade secrets of a natural or legal person, including source code, except in the cases referred to in Article 5 of Directive (EU) 2016/943 of the European Parliament and of the Council" 52(6): "The Commission shall ensure that a list of general-purpose AI models with systemic risk is published and shall keep that list up to date, without prejudice to the need to observe and protect intellectual property rights and confidential business information or trade secrets in accordance with Union and national law."	*AIA includes everything from CoC and is even broader*

			(167): “In order to ensure trustful and constructive cooperation of competent authorities on Union and national level, all parties involved in the application of this Regulation should respect the confidentiality of information and data obtained in carrying out their tasks, in accordance with Union or national law. They should carry out their tasks and activities in such a manner as to protect, in particular, intellectual property rights, confidential business information and trade secrets, the effective implementation of this Regulation, public and national security interests, the integrity of criminal and administrative proceedings, and the integrity of classified information.”	
12	**Whistleblowing**	*Not found*	(172): “Persons acting as whistleblowers on the infringements of this Regulation should be protected under the Union law. Directive (EU) 2019/1937 of the European Parliament and of the Council (54) should therefore apply to the reporting of infringements of this Regulation and the protection of persons reporting such infringements.”	*Nothing in the CoC on whistleblower reporting protection.*

SaferAI

Action 5

Risk management framework

High-level findings:

- Action 5 is explicit about providers having "AI governance and risk management policies". The AI Act is as well when 55(1)(b)'s requirement to "assess and mitigate possible systemic risks" is combined with Recital 114 stating that providers, "should continuously assess and mitigate systemic risks, including for example by putting in place risk-management policies, such as accountability and governance processes."
- While most of the detailed points in Action 5 are not in the AI Act, they are best practices in risk management and common across risk management standards.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 5)	EU AI Act (GPAI focus)	Comments
1	**AI governance and risk management policies**	"Develop, implement and disclose AI governance and risk management policies…"	55(1)(b): "**assess and mitigate possible systemic risks…**" 56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): "the **measures, procedures and modalities for the assessment and management of the systemic risks** at Union level, including the documentation thereof, which **shall be proportionate to the risks**, take into consideration their severity and probability and take into account the specific challenges of tackling those risks in light of the possible ways in	*Action 5 is very similar to Recital 114.* *Also, standards for assessing and mitigating risks usually necessitate such policies, so there is almost certainly a need for such policies.* *There is no public disclosure of such policies required by the AIA, but evaluations, adversarial testing, and model adaptations must be disclosed to the AI Office if requested and are forms of risk assessment and*

			which such risks may emerge and materialise along the AI value chain." (114): "continuously assess and mitigate systemic risks, including for example by **putting in place risk-management policies, such as accountability and governance processes,** implementing post-market monitoring, taking appropriate measures along the entire model's lifecycle and cooperating with relevant actors along the AI value chain." Annex XI(Section 2)(1): "A detailed description of the evaluation strategies, including evaluation results, on the basis of available public evaluation protocols and tools or otherwise of other evaluation methodologies. Evaluation strategies shall include evaluation criteria, metrics and the methodology on the identification of limitations." Annex XI(Section 2)(2): "Where applicable, a detailed description of the measures put in place for the purpose of conducting internal and/or external adversarial testing (e.g. red teaming), model adaptations, including alignment and fine-tuning."	*mitigation.* *"Develop" could not be found in the AIA.*

2	**Organizational mechanisms**	"Organizations should put in place appropriate organizational mechanisms to develop, disclose and implement risk management and governance policies, including for example accountability and governance processes to identify, assess, prevent, and address risks, where feasible throughout the AI lifecycle."	55(1)(b): "assess and mitigate possible systemic risks…" (114): "continuously assess and mitigate systemic risks, including for example by putting in place risk-management policies, such as accountability and governance processes, implementing post-market monitoring, taking appropriate measures along the entire model's lifecycle and cooperating with relevant actors along the AI value chain."	*CoC is more explicit.* *Action 5 is similar to Recital 114.* *Also, standards for assessing and mitigating risks usually necessitate such policies, so there is almost certainly a need for such policies.* *However, "develop" and "disclose" could not be found in the AIA.* *Also, the AIA does not directly require such mechanisms, though they are needed to fulfil the AIA's risk management requirements.*
3	**Privacy policies**	"This includes disclosing where appropriate privacy policies, including for personal data, user prompts and advanced AI system outputs."	(28): Aside from the many beneficial uses of AI, it can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices. Such practices are particularly harmful and abusive and should be prohibited because they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and **fundamental rights enshrined in the Charter, including the right to** non-discrimination, to data protection	*CoC is more explicit.* *Systemic risks must be mitigated, including risks to fundamental rights, which in turn includes the right to privacy. Also, Recital 110 on GPAI systemic risks explicitly lists harms to privacy. Therefore, disclosing privacy policies could be prudent as a risk mitigation measure under*

			and to **privacy** and the rights of the child. (110): [recital on systemic risks] “the facilitation of disinformation or harming privacy with threats to democratic values and human rights;” 3(65): “‘systemic risk’ means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, **fundamental rights**, or the society as a whole, that can be propagated at scale across the value chain;” 55(1)(b): “**assess and mitigate possible systemic risks** at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;”	*the AIA.* *Other EU regulations may be more relevant to the CoC text on privacy policies.*
4	**Establish and disclose AI governance policies and mechanisms**	“Organizations are expected to establish and disclose their AI governance policies and organizational mechanisms to implement these policies in accordance with a risk-based	55(1)(b): “**assess and mitigate possible systemic risks** at Union level, including their sources, that may stem from the development, the placing on the market, or the use of	*CoC is more explicit.* *The AIA does not require public disclosure of these policies, assuming this is the intention of the CoC.*

		approach. This should include accountability and governance processes to evaluate and mitigate risks, where feasible throughout the AI lifecycle."	general-purpose AI models with systemic risk;" (114): "continuously assess and mitigate systemic risks, including for example by putting in place risk-management policies, such as accountability and governance processes, implementing post-market monitoring, taking appropriate measures along the entire model's lifecycle and cooperating with relevant actors along the AI value chain."	
5	**Development and update of risk management policies**	"The risk management policies should be developed in accordance with a risk-based approach and apply a risk management framework across the AI lifecycle as appropriate and relevant, to address the range of risks associated with AI systems, and policies should also be regularly updated."	55(1)(b): "assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" 56(2): "The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): the measures, procedures and modalities for the assessment and management of the systemic risks at Union level, including the documentation thereof, which shall be proportionate to the risks, take into consideration their severity and	*CoC is more explicit.* *If one assumes that policies are needed to manage risk, then these are essentially the same.* *Only the CoC explicitly requires regular updates, but these are needed to address risk throughout the AI lifecycle, including those that may materialise along the AI value chain.*

			probability and take into account the specific challenges of tackling those risks in light of the possible ways in which such risks may emerge and materialise along the AI value chain."	
6	**Policies, procedures and training for staff**	"Organizations should establish policies, procedures, and training to ensure that staff are familiar with their duties and the organization's risk management practices"	4: [AI literacy] 9(5)(c): [Training to deployers] 14(5): [Human oversight] 26(2): [Deployers - training and authority of human oversight] 60(4)(j): [Testing of high-risk outside of sandboxes - training and authority]	*CoC is more explicit.* *This is good practice in risk management, but not explicitly stated for GPAI model providers.* *There are similar requirements for high-risk AI systems however.*

SaferAI

Action 6

Cybersecurity

High-level findings:

- Overall, Action 6 and the Act have much in common. The overall concerns are the same, and some of the text is very similar. Where one of the two is more specific, these more detailed requirements can be reasonably inferred from the other text.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 6)	EU AI Act (GPAI focus)	Comments
1	**Level of controls/ measures/protection**	"Invest in and implement robust security controls" "ensure that the cybersecurity of advanced AI systems is appropriate to the relevant circumstances and the risks involved" "regularly review security measures to ensure they are maintained to a high standard and remain suitable to address risks"	55(1)(d): "**ensure an adequate level of cybersecurity protection for the general-purpose AI model** with systemic risk and the physical infrastructure of the model." 56(2): The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: (d): "the measures, procedures and modalities for the assessment and **management of the systemic risks** at Union level, including the documentation thereof, **which shall be proportionate to the risks**, take into consideration their severity and probability and take into account the specific challenges of tackling those risks in light of the possible ways in which such risks may emerge and materialise along the AI value chain."	*CoC and AIA both require cybersecurity to be "adequate".* *CoC also uses "robust", "appropriate", and "suitable", while the AIA uses "proportionate" (normative) and "appropriate" (recital).* *The AIA does not explicitly require investment, though this is typically necessary to* "ensure an adequate level of cybersecurity protection"

			(115): “[Cybersecurity protection] could be facilitated by [controls] appropriate to the relevant circumstances and the risks involved.”	
2	**Lifecycle**	“across the AI lifecycle”	55(1)(b): “assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the **development, the placing on the market, or the use** of general-purpose AI models with systemic risk;”	*Equivalent*
3	**Cybersecurity risk assessment**	“performing an assessment of cybersecurity risks”	55(1)(b): “**assess** and mitigate **possible systemic risks** at Union level, **including their sources**, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;” (115): “Cybersecurity protection related to systemic risks associated with malicious use or attacks should duly consider [list of threats]”	*CoC explicitly requires a cybersecurity risk assessment, while the AIA reasonably requires one, as inadequate cybersecurity is a source of risk*
4	**Example controls**	“securing model weights and, algorithms, servers, and datasets, such as through operational security measures for information security and appropriate cyber/physical access controls.” “…cybersecurity policies and adequate technical and institutional solutions”	(115): “securing model weights, algorithms, servers, and data sets, such as through operational security measures for information security, specific cybersecurity policies, adequate technical and established solutions, and cyber and physical access controls”	*Slight differences in adjectives, (e.g., “appropriate”, “specific”)*

5	**Physical infrastructure**	"implement robust security controls, including physical security"	55(1)(d): "**ensure an adequate level of cybersecurity protection for** the general-purpose AI model with systemic risk and **the physical infrastructure of the model**." (115): "That protection could be facilitated by…physical access controls"	*Both require controls for physical security*
6	**Secure environment**	"Organizations should also have in place measures to require storing and working with the model weights of advanced AI systems in an appropriately secure environment with limited access to reduce both the risk of unsanctioned release and the risk of unauthorized access."	55(1)(d): "ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model." (115): "providers should ensure an adequate level of cybersecurity protection for the model and its physical infrastructure, if appropriate" (115): "Cybersecurity protection related to systemic risks associated with malicious use or attacks should duly consider…unauthorised releases…unauthorised access"	*Essentially the same*
7	**Vulnerability management process**	"commitment to have in place a vulnerability management process"	56(2): "The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: [...] (d): "the measures, procedures and modalities for the assessment and management of the systemic risks at Union	*The AIA does not explicitly require or discuss a "vulnerability management process", though it can be considered a specific example of "documentation" and an "adequate technical and established solution" for*

			level, including the documentation thereof…" (115): "That protection could be facilitated by securing model weights, algorithms, servers, and data sets, such as through…adequate technical and established solutions"	*cybersecurity*
8	**Regular review**	"regularly review security measures to ensure they are maintained to a high standard and remain suitable to address risks"	55(1)(d): ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model.	*CoC is explicit about regular review and maintenance of security measures, while it can be reasonably inferred from the AIA as being necessary*
9	**Insider threats**	"implement robust security controls, including…insider threat safeguards" "establishing a robust insider threat detection program consistent with protections provided for their most valuable intellectual property and trade secrets, for example, by limiting access to proprietary and unreleased model weights." "operational security measures for information security"	(115): "protection could be facilitated by…operational security measures for information security"	*Both mention opsec (recital in AIA), but only CoC specifically discusses insider threat controls.*
10	**Malicious use or attacks**	"ensure that the cybersecurity of advanced AI systems is appropriate to the relevant circumstances and the risks involved."	(115): "Cybersecurity protection related to systemic risks associated with malicious use or attacks should duly consider accidental model leakage, unauthorised releases, circumvention of safety	*CoC does not explicitly consider accidental model leakage, circumvention of safety measures, defence against cyberattacks, or*

		“securing model weights” “unauthorised releases…unauthorised access”	measures, and defence against cyberattacks, unauthorised access or model theft.”	*model theft. However, these can be reasonably inferred from the CoC.*

SaferAI

Action 7

Content Authentication and Provenance Mechanisms

High-level findings:

- Both CoC and AIA require content authentication and provenance mechanisms.
- CoC prescribes tools or APIs to allow users to determine if particular content was created with their advanced AI system, such as via watermarks; AIA includes watermarks as a possible technique, but additionally mentions metadata identifications and cryptographic methods
- CoC prescribes collaboration and investments in research, as appropriate, to advance the state of the field of content authentication and provenance mechanisms; the AIA does not contain the obligation for providers to advance the field, it just prescribes the application of such mechanisms

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 7)	EU AI Act (GPAI focus)	Comments
1	**Content authentication and provenance mechanisms**	"Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content This includes, **where appropriate and technically feasible**, content authentication and provenance mechanisms for content created with an organization's advanced AI system. "	50(2): "Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally **acknowledged state of the art**, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the	*High commonality is based on the assumption that "appropriate and technically feasible" is similar to the "generally acknowledged state of the art". "Reliable" is covered by both.* *Providers could in theory use mechanisms already developed, instead of developing their own, though perhaps it would be difficult to implement outside the model.* *AIA is more specific, particularly regarding*

			semantics thereof, or where authorised by law to detect, prevent, investigate or prosecute criminal offences." (133): "...Such techniques and methods can be implemented at the level of the AI system or at the level of the AI model, including general-purpose AI models generating content…"	*machine readability, with further information in Recital 133.* *There are exceptions to certain AI systems in the AIA. No exceptions in CoC.*
2	**Identifier of the service or model that created the content**	"The provenance data should include an identifier of the service or model that created the content, but need not include user information."	50(2): "...Providers shall ensure their technical solutions are effective, interoperable…" (133): "Such techniques and methods should be sufficiently reliable, interoperable, effective and robust as far as this is technically feasible, taking into account available techniques or a combination of such techniques, such as watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints or other techniques, as may be appropriate."	*Solutions, such as metadata identifications and logging methods, would reasonably include an identifier of the service or model in order to be effective and interoperable.*
3	**Provide tools/APIs to check whether content was created by the organisation's AI**	"Organizations should also endeavor to develop tools or APIs to allow users to determine if particular content was created with their advanced AI system, such as via watermarks."	50(2): "[Providers]...shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.| (133): "Such techniques and methods should be sufficiently reliable, interoperable, effective and robust as far as this is technically feasible, taking into account available techniques or a combination of such techniques, such as	*The CoC foresees the development of detection tools for users, while providers could use existing tools under the AIA.* *The AIA requires machine readability, which the CoC does not.* *So, while these are not exactly the same, they are*

			watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints or other techniques, as may be appropriate."	*at least complimentary.*
4	**Collaboration and investment in research**	"Organizations should collaborate and invest in research, as appropriate, to advance the state of the field."	50(2): …Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art… (133): "When implementing this obligation, providers should also take into account the specificities and the limitations of the different types of content and the relevant technological and market developments in the field, as reflected in the generally acknowledged state of the art."	*AIA does not require providers to collaborate nor invest in research to advance the field, but rather that providers take into account the generally acknowledged state of the art, including relevant technological and market developments in the field.*
5	**Labelling/disclaimers for interactions with AI systems**	Organizations are further encouraged to implement other mechanisms such as labeling or disclaimers to enable users, where possible and appropriate, to know when they are interacting with an AI system.	50(1): "Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use."	*CoC: "where possible and appropriate"* *AIA: Applies to AI systems, not GPAI models specifically. Nothing about feasibility and appropriateness; instead: concrete exception (which might fall under the appropriateness condition in CoC)*

SaferAI

Action 8

Investments in Research and Mitigation Measures

High-level findings:

- Commonality: The objectives of the research investments prescribed by the CoC are in line with the objectives of the AIA.
- Discrepancy: The CoC requires research investments and collaboration to promote those objectives. The AIA does not require such research investments and collaboration and does not require providers to share research and best practices on risk mitigation. However, there are some ways for GPAI model providers to share their research with institutions, namely through the advisory forum and the drawing-up of the codes of practice.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 8)	EU AI Act (GPAI focus)	Comments
1	**Development of mitigation tool and proactive risk mitigation**	"Organizations also commit to invest in developing appropriate mitigation tools, and work to proactively manage the risks of advanced AI systems, including environmental and climate impacts, so that their benefits can be realized."	55(1)(b): "assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" 56(2): "The AI Office and the Board shall aim to ensure that the codes of practice cover at least the obligations provided for in Articles 53 and 55, including the following issues: [...] (d) the measures, procedures and modalities for the assessment and **management of the systemic risks** at Union level, including the documentation thereof, which shall be proportionate to the risks, take into consideration their severity and probability and take into account the specific challenges of	*Risk Mitigation is required by both CoC and AIA.* *Development of mitigation tools is not required by AIA. Only risk mitigation in general is required, this could be done with tools provided by other organisations.*

			tackling those risks in light of the possible ways in which such risks may emerge and materialise along the AI value chain." (116): "Codes of practice should also be focused on specific **risk assessment and mitigation measures**." (164): "Compliance with the obligations should be enforceable, inter alia, through requests to take appropriate measures, **including risk mitigation measure**s in the case of identified systemic risks" 112(6): By 2 August 2028 and every four years thereafter, the Commission shall submit a report on the review of the progress on the development of standardisation deliverables on the energy-efficient development of general-purpose AI models, and asses the need for further measures or actions, including **binding measures or actions**. The report shall be submitted to the European Parliament and to the Council, and it shall be made public." (174): "Moreover, by 2 August 2028 and every four years thereafter, the Commission should evaluate and report to the European Parliament and to the Council on [...] the progress on the development of standardisation deliverables on energy efficient development of general-purpose AI models,	*The AIA is concerned about the energy impacts of models. There are no binding requirements on model providers now, however binding measures or actions might be set up in the future according to Art. 112(6).*

			including the need for further measures or actions." Annex XI(Section 1): "The technical documentation referred to in Article 53(1), point (a) shall contain at least the following information as appropriate to the size and risk profile of the model: [...] 2. [...] (e) known or estimated energy consumption of the model."	
2	**Investments in research to advance AI safety, security, trust and addressing key risks**	"Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures." "This includes conducting, collaborating on and investing in research that supports the advancement of AI safety, security, and trust, and addressing key risks" "Organizations commit to conducting, collaborating on and investing in research that supports the advancement of AI safety, security, trustworthiness and addressing key risks, such as prioritizing research on upholding democratic values, respecting human rights, protecting children and vulnerable groups, safeguarding intellectual property	1(1): "...ensuring a high level of protection of health, **safety**, fundamental rights enshrined in the Charter, including **democracy**, the rule of law and environmental protection, **against the harmful effects of AI systems** in the Union and supporting innovation." (176): "the objective of this Regulation, namely to improve the functioning of the internal market and to promote the uptake of human centric and **trustworthy** AI, while ensuring a high level of protection of health, **safety**, **fundamental rights** enshrined in the Charter, including **democracy**, the rule of law and environmental protection against harmful effects of AI systems in the Union and supporting innovation" (115): "That protection could be facilitated by securing model weights, algorithms, servers, and data sets, such as through operational **security measures for information security,**	*Commonality: The objectives of the research investments prescribed by the CoC are mostly in line with the objectives of the AIA. The only objective that is not explicitly mentioned in the AIA regarding GPAI models is the protection of vulnerable groups (apart from children). Instead, the AIA individually lists the rights of people with disabilities and gender equality.* *Discrepancy: The CoC requires research investments to promote those objectives. The AIA*

		rights and privacy, and avoiding harmful bias, mis- and disinformation, and information manipulation."	**specific cybersecurity** policies, adequate technical and established solutions, and cyber and physical access controls, appropriate to the relevant circumstances and the risks involved." 53(1)(b): "draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Without prejudice to the need to observe and **protect intellectual property rights** and confidential business information or trade secrets in accordance with Union and national law, the information and documentation shall: (110): "In particular, international approaches have so far identified the need to pay attention to risks from potential intentional misuse or unintended issues of control relating to alignment with human intent; chemical, biological, radiological, and nuclear risks, such as the ways in which barriers to entry can be lowered, including for weapons development, design acquisition, or use; offensive cyber capabilities, such as the ways in vulnerability discovery, exploitation, or operational use can be enabled; the effects of interaction and tool use, including for example the capacity to control physical systems and interfere with critical infrastructure; risks from models of making copies of themselves or 'self-replicating' or training other models; **the ways in which models can give rise to**	*does not require research investments.*

			harmful bias and discrimination with risks to individuals, communities or societies; the facilitation of disinformation or harming privacy with **threats to democratic values and human rights**; risk that a particular event could lead to a chain reaction with considerable negative effects that could affect up to an entire city, an entire domain activity or an entire community. (48): “The extent of the adverse impact caused by the AI system on the fundamental rights protected by the Charter is of particular relevance when classifying an AI system as high risk. Those rights include: [...] **protection of personal data, [...] the right to non-discrimination [...] the right to education, [...] the rights of persons with disabilities, gender equality, intellectual property rights [...]** In addition to those rights, it is important to highlight the fact that **children have specific rights** as enshrined in Article 24 of the Charter and in the United Nations Convention on the Rights of the Child [...]” *(even though this recital is about AI systems and not about GPAI models, it is relevant here because it lists fundamental rights. Risks to fundamental rights must be mitigated as part of systemic risk mitigation according to 3(65) and 55(1)(b))*(133): (133) “A variety of AI systems can generate large quantities of synthetic content that becomes increasingly hard for humans to distinguish from human-generated and authentic content. The wide availability and	

			increasing capabilities of those systems have a significant impact on the integrity and trust in the information ecosystem, raising new risks of **misinformation and manipulation** at scale [...]" *No obligations regarding research*	
3	**Share research and best practices on risk mitigation**	"Organizations are encouraged to share research and best practices on risk mitigation."	38(3): "The Commission shall provide for the exchange of knowledge and best practices between notifying authorities." 66: "the Board may in particular: [...] (b) collect and share technical and regulatory expertise and best practices among Member States;" 68(3): "The scientific panel shall advise and support the AI Office, in particular with regard to the following tasks: (a) [...] (ii) contributing to the development of tools and methodologies for evaluating capabilities of general-purpose AI models and systems, including through benchmarks;" 68(2): "The scientific panel shall consist of experts selected by the Commission on the basis of up-to-date scientific or technical expertise in the field of AI necessary for the tasks set out in paragraph 3, and shall be able to demonstrate meeting all of the following conditions: (b) **independence from any provider of AI systems or general-purpose AI models**;"	*There are some obligations and mandates that the scientific panel (Recital (116)), the AI Board (66(b), Recital (20)), and the AIO/EC have related to the development and sharing of best practices. Providers are not directly part of those institutions (65(2)/Recital (149) and 68(2)), but they might give advice to the AI Board and EC through the advisory forum (Recital (150)).*

			65(2): “The Board shall be composed of one representative per Member State.” (149): “such representatives may be any persons belonging to public entities who should have the relevant competences and powers to facilitate coordination at national level and contribute to the achievement of the Board’s tasks.” (148): “Furthermore, a Board composed of representatives of the Member States, a scientific panel to integrate the scientific community and **an advisory forum to contribute stakeholder input to the implementation of this Regulation**, at Union and national level, should be established.” (20): “The European Artificial Intelligence Board (the ‘Board’) should support the Commission, to promote AI literacy tools, public awareness and understanding of the benefits, risks, safeguards, rights and obligations in relation to the use of AI systems. In cooperation with the relevant stakeholders, the Commission and the Member States should facilitate the drawing up of voluntary codes of conduct to advance AI literacy among persons dealing with the development, operation and use of AI.” (150): “With a view to ensuring the involvement of stakeholders in the implementation and application of this Regulation, an **advisory forum should be established to advise and**	

			provide technical expertise to the Board and the Commission. To ensure a varied and balanced stakeholder representation between commercial and non-commercial interest and, within the category of commercial interests, with regards to SMEs and other undertakings, the advisory forum **should comprise inter alia industry, start-ups, SMEs**, academia, civil society, including the social partners, as well as the Fundamental Rights Agency, ENISA, the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI)" 56(3): The AI Office may invite all providers of general-purpose AI models, as well as relevant national competent authorities, to participate in the drawing-up of codes of practice. Civil society organisations, industry, academia and other relevant stakeholders, such as downstream providers and independent experts, may support the process.	*Providers can participate in the drawing-up of the Codes of Practice. Civil society organisations, industry, academia and other relevant stakeholders, such as downstream providers and independent experts, may support the drawing-up of the Codes of Practice. This is another opportunity for those stakeholders to share research and best practices on risk mitigation. (56(3))*

SaferAI

Action 9

Developing AI for the Benefit of the Public

High-level findings:

- CoC and AIA have similar (or at least compatible) end-goals regarding developing AI for the benefit of the public, but the details of their scope and who is responsible differ

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 9)	EU AI Act (GPAI focus)	Comments
1	**Purpose and priority**	"Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education"	1(1): "The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI)" 57(5): "AI regulatory sandboxes established under paragraph 1 shall provide for a controlled environment that fosters innovation and facilitates the development, training, testing and validation of innovative AI systems…"	*Each has a different focus, i.e., CoC addressing the world's greatest challenges vs AIA developing innovative and trustworthy AI*
2	**Scope**	"These efforts are undertaken in support of progress on the United Nations Sustainable Development Goals, and to encourage AI development for global benefit."	1(1): "...ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation."	*Global vs EU scope* *According to the CoC, the world's greatest challenges must be actively addressed. There is no such obligation in the AIA. The AIA is focused on ensuring that no harm is done (to health, safety, and fundamental rights).*

3	**Trustworthy and human-centric AI**	"Organizations should prioritize responsible stewardship of trustworthy and human-centric AI"	1(1): "promote the uptake of human-centric and trustworthy artificial intelligence (AI)" (8): "this Regulation supports the objective of promoting the European human-centric approach to AI"	*Similar*
4	**AI literacy**	"Organizations should…support digital literacy initiatives that promote the education and training of the public, including students and workers, to enable them to benefit from the use of advanced AI systems, and to help individuals and communities better understand the nature, capabilities, limitations, and impact of these technologies."	4: "Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used." 66(f) [Tasks of the Board]: "support the Commission in promoting AI literacy, public awareness and understanding of the benefits, risks, safeguards and rights and obligations in relation to the use of AI systems;" 95(2)(c) [Codes of conduct for voluntary application of specific requirements]: "promoting AI literacy, in particular that of persons dealing with the development, operation and use of AI"	*Organizations vs the AI Board (and EC) support public AI literacy.* *Providers must help ensure AI literacy of their staff and certain others.* *Codes of conduct for providers are voluntary and focused more specifically on the development, operation and use of AI.*
5	**Role of civil society in addressing challenges**	"Organizations should work with civil society and community groups to	53(3): "Providers of general-purpose AI models shall cooperate as necessary with	*Global vs EU focus. Act seems to have EU and*

		identify priority challenges and develop innovative solutions to address the world's greatest challenges."	the Commission and the national competent authorities in the exercise of their competences and powers pursuant to this Regulation." 67(1) [Advisory forum to AI Board and EC]: "An advisory forum shall be established to provide technical expertise and advise the Board and the Commission, and to contribute to their tasks under this Regulation." 67(2) [Advisory forum to AI Board and EC]: "The membership of the advisory forum shall represent a balanced selection of stakeholders, including industry, start-ups, SMEs, civil society…" (165): "Providers…of all AI systems, high-risk or not, and AI models should also be encouraged to apply on a voluntary basis additional requirements related, for example, to…inclusive and diverse design and development of AI systems, including attention to vulnerable persons and accessibility to persons with disability, stakeholders' participation with the involvement, as appropriate, of relevant stakeholders such as business and civil society organisations, academia, research organisations, trade unions and consumer protection organisations in the design and development of AI systems…"	*national institutions facilitating*

SaferAI

Action 10

Development and Adoption of Technical Standards

High-level findings:

- Both the CoC and the Act encourage organisations to participate in the development and use of content provenance methods, along with other methodologies and measures more broadly.

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 10)	EU AI Act (GPAI focus)	Comments
1	**Standards and best practices**	"Organizations are encouraged to contribute to the development and, where appropriate, use of international technical standards and best practices…" "...working with Standards Development Organizations (SDOs), also when developing organizations' testing methodologies…cybersecurity policies, public reporting, and other measures."	40(3): "The participants in the standardisation process shall seek to promote investment and innovation in AI, including through increasing legal certainty, as well as the competitiveness and growth of the Union market, to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are consistent with Union values, fundamental rights and interests, and to enhance multi-stakeholder governance ensuring a balanced representation of interests and the effective participation of all relevant stakeholders…" 56(3): "The AI Office may invite all providers of general-purpose AI models…to participate in the drawing-up of codes of practice."	*While the AIA does not strictly require providers to work to develop standards, it is in their interest to do so and they are, in theory, invited and facilitated to do so by authorities*

			62(1)(d): “[Member States shall] facilitate the participation of SMEs and other relevant stakeholders in the standardisation development process.”	
2	**Content authentication and provenance**	“Organizations are encouraged to contribute to the development and, where appropriate, use of international technical standards and best practices, including for watermarking…” “In particular, organizations also are encouraged to work to develop interoperable international technical standards and frameworks to help users distinguish content generated by AI from non-AI generated content.” “content authentication and provenance mechanisms”	50(1): “Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the **natural persons concerned are informed that they are interacting with an AI system**…” 50(2): “Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards….”	*Broadly similar. While the AIA does not strictly require providers to work to develop standards, this is a reasonable way for them to meet the AIA’s interoperability requirement.*

SaferAI

Action 11

Data input measures and protections for personal data and intellectual property

High-level findings:

- The CoC prescribes measures to manage data quality in order to mitigate against harmful biases; in the AIA, such measures are implied to be part of the mitigation of systemic risks and their sources.
- With respect to privacy, personal data, copyright, and intellectual property, the AIA (especially when combined with Union law) is much more detailed, explicit and comprehensive in its requirements.
- Both require assurance of privacy and compliance with other legal frameworks; the AIA explicitly mentions the need to comply with *Directive (EU) 2019/790.*

#	Point of Comparison	Hiroshima Process Code of Conduct (Action 11)	EU AI Act (GPAI focus)	Comments
1	**Existence of measures on data to mitigate against harmful biases**	"Organizations are encouraged to take appropriate measures to manage data quality, including training data and data collection, to mitigate against harmful biases." "Implement appropriate data input measures" "Appropriate measures could include transparency, privacy-preserving training techniques, and/or testing and fine-tuning to ensure that systems do not divulge confidential or sensitive data"	55(1)(b): "assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" (110): "the ways in which models can give rise to harmful bias and discrimination with risks to individuals, communities or societies;" Annex XI(Section 1)(2)(c): "information on the data used for training, testing and validation, where applicable, including the type and provenance of data and curation methodologies (e.g. cleaning, filtering, etc.), the number of data points, their scope and main characteristics; how the data was obtained and selected as well as all other measures to detect	*CoC has a high-level ask to manage data quality, while the AIA implies this must be done (as such measures must be documented).* *CoC data input measures are similar to AIA curation methodologies.* *The CoC is more explicit about not divulging confidential or sensitive data. However, the AIA implies the need for measures to preserve both; divulging confidential data could harm privacy, and*

			the unsuitability of data sources and methods to detect identifiable biases, where applicable"	*divulging sensitive data could be an info hazard leading to various harms.*
2	**Safeguards to respect privacy, personal data, copyright, and intellectual property**	"Organizations are encouraged to implement appropriate safeguards, to respect rights related to privacy and intellectual property, including copyright-protected content." "Implement appropriate [...] protections for personal data and intellectual property"	2(7): "Union law on the protection of personal data, privacy and the confidentiality of communications applies to personal data processed in connection with the rights and obligations laid down in this Regulation. This Regulation shall not affect Regulation (EU) 2016/679 or (EU) 2018/1725, or Directive 2002/58/EC or (EU) 2016/680, without prejudice to Article 10(5) and Article 59 of this Regulation." 3(65): "'systemic risk' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, **fundamental rights**, or the society as a whole, that can be propagated at scale across the value chain; (28): "Aside from the many beneficial uses of AI, it can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices. Such practices are particularly harmful and abusive and should be prohibited because they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and **fundamental rights enshrined in the Charter,**	*Privacy and personal data:* *Union law on privacy and personal data applies to data processed in connection with the rights and obligations laid down in the AIA (Art. 2(7)).* *AIA is more explicit and comprehensive: Systemic risks to fundamental rights (Art. 3(65)), which include harms to privacy (Recital (28)), must be mitigated.*

			including the right to non-discrimination, to data protection and to **privacy** and the rights of the child." (110): "the facilitation of disinformation or harming privacy with threats to democratic values and human rights;" 55(1)(b): "assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;" 53(1)(c): "put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;" Recitals (105), (106), (107), and (108)	*Copyright, intellectual property and related rights:* *CoC: "safeguards"* *AIA: "policy to comply with Union law"* *AIA, combined with Union law, is much more detailed.*
3	**Comply with applicable legal frameworks**	"Organizations should also comply with applicable legal frameworks."	2(7): "Union law on the protection of personal data, privacy and the confidentiality of communications applies to personal data processed in connection with the rights and obligations laid down in this Regulation. This Regulation shall not affect Regulation (EU) 2016/679 or (EU) 2018/1725, or Directive 2002/58/EC or (EU) 2016/680, without prejudice to Article 10(5) and Article 59 of this Regulation."	*CoC is not specific about which legal frameworks should be complied with.* *AIA states: "put in place a policy to comply with Union law on copyright and related rights" and explicitly mentions Directive (EU) 2019/790.*

			53(1)(c): “put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;” Recitals 105, 106, 107, 108	
4	**Public training data summary**	*Nothing found*	53(1): “Providers of general-purpose AI models shall: [...] (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.”	*The CoC does not require the publication of details on training content.*