# Pilot phase: Reporting Framework for the International Code of Conduct for Organizations Developing Advanced AI Systems

**This document is intended solely for coordination purposes. All responses must be submitted online on www.oecd.ai/g7.**
**If you have questions or technical difficulties, please contact ai@oecd.org.**

Thank you for participating in the pilot phase of the reporting framework for the International Code of Conduct for Organizations Developing Advanced AI Systems (Code of Conduct). The objective of the reporting framework is to contribute to monitoring the voluntary application of the Code of Conduct by organizations developing advanced AI systems. In this pilot phase, organizations are invited to complete the survey below to the best of their ability. Responses to the survey will be collected and recorded by the OECD Secretariat for the sole purpose of refining and improving the reporting framework in the next phase.

Please complete the reporting framework by **September 6, 2024.**

The survey is structured around the 11 Action of the Code of Conduct and includes the following types of questions:

- Yes/No questions that include open text fields for further explanation
- Multiple-choice questions
- Open-ended questions

Given the voluntary nature of the Code of Conduct and that information regarding model development can be sensitive, reported activities should not include confidential or proprietary information.

Please note that responses must be filled and submitted within one session, and organizations are encouraged to consolidate their inputs in advance before filling the online survey.

**A1. Name***

**A2. Organization***

**A3. Email address***

# Section B: Action 1.A: Risk Identification

*Excerpt from the Code of Conduct: Action 1*

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, through a combination of methods for evaluations, such as red-teaming, and implementing appropriate mitigation to address identified risks and vulnerabilities. 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. In support of such testing, developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development. These measures should be documented and supported by regularly updated technical documentation.

This testing should take place in in secure environments and be performed at several checkpoints throughout the AI lifecycle in particular before deployment and placement on the market 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. In designing and implementing testing measures, organizations commit to devote attention to the following risks as appropriate:

Chemical, biological, radiological, and nuclear risks, such as the ways in which advanced AI systems can lower barriers to entry, including for non-state actors, for weapons development, design acquisition, or use. Offensive cyber capabilities, such as the ways in which systems can enable vulnerability discovery, exploitation, or operational use, bearing in mind that such capabilities could also have useful defensive applications and might be appropriate to include in a system. Risks to health and/or Safety, including the effects of system 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. Societal risks, as well as risks to individuals and communities such as the ways in which advanced AI systems or models can give rise to harmful bias and discrimination or lead to violation of applicable legal frameworks, including on privacy and data protection. Threats to democratic values and human rights, including the facilitation of disinformation or harming privacy. 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.

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.

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.

**B1.** **1. How does your organization identify the risks referenced in Action 1 of the Code of Conduct (and possibly others, if appropriate) across the lifecycle of advanced AI systems (AI lifecycle), including before deployment and placement on the market?***

*Please see above the list of risks referenced in Action 1.*

B2. 2. How does your organization define and/or classify different types of risks (e.g., unreasonable risks)?*

# Section C: Action 1.B: Risk Evaluation

*Excerpt from the Code of Conduct: Action 1*

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, through a combination of methods for evaluations, such as red-teaming, and implementing appropriate mitigation to address identified risks and vulnerabilities. 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. In support of such testing, developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development. These measures should be documented and supported by regularly updated technical documentation.

This testing should take place in in secure environments and be performed at several checkpoints throughout the AI lifecycle in particular before deployment and placement on the market 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. In designing and implementing testing measures, organizations commit to devote attention to the following risks as appropriate:

Chemical, biological, radiological, and nuclear risks, such as the ways in which advanced AI systems can lower barriers to entry, including for non-state actors, for weapons development, design acquisition, or use. Offensive cyber capabilities, such as the ways in which systems can enable vulnerability discovery, exploitation, or operational use, bearing in mind that such capabilities could also have useful defensive applications and might be appropriate to include in a system. Risks to health and/or Safety, including the effects of system 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. Societal risks, as well as risks to individuals and communities such as the ways in which advanced AI systems or models can give rise to harmful bias and discrimination or lead to violation of applicable legal frameworks, including on privacy and data protection. Threats to democratic values and human rights, including the facilitation of disinformation or harming privacy. 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.

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.

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.

**C1. 3. What testing measures (internal or external) are employed for evaluating risks across the AI lifecycle, including during development (e.g., regarding system capabilities and limitations or training datasets)?***

*Please select all that apply.*

Red teaming □

Independent external tests □

Quantitative evaluation tests □

Other

**C2.** **Describe how your organization conducts red-teaming to evaluate the model's/system's fitness for moving beyond the development stage?**

**C3.** **Describe how your organization conducts independent external tests to evaluate the model's/system's fitness for moving beyond the development stage?**

**C4. What metrics are employed for quantitative risk evaluation?**

**C5. What are the limitations of the quantitative evaluations your organization conducts, if any?**

**C6. 4. At which checkpoints throughout the AI lifecycle are evaluations performed?***

.

**C7. 5. Does testing take place in secure environments?***

Yes □

No □

C8. **If yes, please elaborate.**

# Section D: Action 1.C: Risk Management

*Excerpt from the Code of Conduct: Action 1*

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, through a combination of methods for evaluations, such as red-teaming, and implementing appropriate mitigation to address identified risks and vulnerabilities. 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. In support of such testing, developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development. These measures should be documented and supported by regularly updated technical documentation.

This testing should take place in in secure environments and be performed at several checkpoints throughout the AI lifecycle in particular before deployment and placement on the market 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. In designing and implementing testing measures, organizations commit to devote attention to the following risks as appropriate:

Chemical, biological, radiological, and nuclear risks, such as the ways in which advanced AI systems can lower barriers to entry, including for non-state actors, for weapons development, design acquisition, or use. Offensive cyber capabilities, such as the ways in which systems can enable vulnerability discovery, exploitation, or operational use, bearing in mind that such capabilities could also have useful defensive applications and might be appropriate to include in a system. Risks to health and/or Safety, including the effects of system 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. Societal risks, as well as risks to individuals and communities such as the ways in which advanced AI systems or models can give rise to harmful bias and discrimination or lead to violation of applicable legal frameworks, including on privacy and data protection. Threats to democratic values and human rights, including the facilitation of disinformation or harming privacy. 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.

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.

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.

**D1.** **6. What steps does your organization take to address risks and vulnerabilities across the AI lifecycle, including throughout the development of advanced AI systems?***

D2. **7. Does your organization's testing measures inform actions to address identified risks?***

Yes □

No □

D3. **If yes, please also describe any actions taken to manage risks including those that may overlap and relate to one another.***

D4. **8. How does your organization collaborate with relevant stakeholders across sectors to assess and adopt risk mitigation measures to address risks, in particular systemic risks?***

D5. **9. Does your organization advance research and investment related to any of the following: security, safety, bias and disinformation, fairness, explainability and interpretability, and transparency of advanced AI systems, increasing robustness and/or trustworthiness of advanced AI systems?***

Yes □

No □

**D6.** **If yes, please elaborate.**

## Section E: Action 1

**E1.** **Implementation documentation Action 1**

**Please share any relevant links or public documentation, including, for example: safety frameworks and/or policies, responsible scaling policies, technical documentation regarding pre-deployment risk identification, assessment and management.**

## E2. Feedback on the reporting framework Action 1

**Please provide any feedback on the reporting framework questions for Action 1 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

# Section F: Action 2: Post-Deployment Monitoring and Reporting

*Excerpt from the Code of Conduct: Action 2*

Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market

Organizations should use, as and when appropriate commensurate to the level of risk, AI systems as intended and monitor for vulnerabilities, incidents, emerging risks and misuse after deployment, and take appropriate action to address these.

Organizations are encouraged to consider, for example, 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.

Organizations are further encouraged to maintain appropriate documentation of reported incidents and to mitigate the identified risks and vulnerabilities, in collaboration with other stakeholders. Mechanisms to report vulnerabilities, where appropriate, should be accessible to a diverse set of stakeholders.

**F1.** **1. How does your organization monitor vulnerabilities, incidents, emerging risks and misuse across the AI lifecycle, including after the deployment of advanced AI systems?***

**F2.** **2. Does your organization have mechanisms to receive reports of incidents and vulnerabilities by third parties?***

Yes □

No □

**This document is intended solely for coordination purposes. All responses must be submitted online on www.oecd.ai/g7.**

**F3.** **If yes, please elaborate.**

**F4.** **3. Does your organization make vulnerability reporting mechanisms accessible to a diverse set of stakeholders, as appropriate?***

Yes □

No □

**F5.** **If yes, please elaborate.**

**F6.** **4. Does your organization have incentive programs for the responsible disclosure of vulnerabilities?***

Yes □

No □

F7. If yes, please elaborate.

F8. 5. Are steps taken to address reported incidents documented and maintained internally?*

Yes □

No □

F9. If yes, please elaborate.

F10. 6. Does your organization take actions to address identified risks and vulnerabilities, including in collaboration with other stakeholders?*

Yes □

No □

F11. **If yes, please elaborate.**

F12. **Implementation documentation**

**Please share any relevant links or public documentation, including, for example: how to report vulnerabilities and patterns of misuse; documentation on reported incidents and mitigation; or responsible disclosure policies / relevant safety frameworks.**

F13. **Feedback on the reporting framework Action 2**

**Please provide any feedback on the reporting framework questions for Action 2 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

This document is intended solely for coordination purposes. All responses must be submitted online on www.oecd.ai/g7.

# Section G: Action 3: Transparency Reporting

*Excerpt from the Code of Conduct: Action 3*

Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability.

This should include publishing transparency reports containing meaningful information for all new significant releases of advanced AI

systems. These reports, instruction for use and relevant technical documentation, as appropriate as, should be kept up-to-date and should

include, for example;

Details of the evaluations conducted for potential safety, security, and societal risks, as well as risks to human rights, Capacities of a model/system and significant limitations in performance that have implications for the domains of appropriate use, 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, and The results of red-teaming conducted to evaluate the model's/system's fitness for moving beyond the development stage.

Organizations should make the information in the transparency reports sufficiently clear and understandable to enable deployers and users as appropriate and relevant to interpret the model/system's output and to enable users to use it appropriately; and that transparency reporting should be supported and informed by robust documentation processes such as technical documentation and instructions for use.

**G1. 1. How does your organization publish clear and understandable reports related to the capabilities, limitations and domains of appropriate and inappropriate use of advanced AI systems?***

G2. 2. How often are such reports usually updated, if relevant?*

G3. 3. How are new significant releases reflected in such reports?*

G4. 4. Which of the following information is included in your organization's publicly available documentation?*

*Please select all that apply.*

□ Details of the evaluations and results conducted for potential safety, security, and societal risks including risks to the enjoyment of human rights

Comment

□ Assessment of the model's or system's effects and risks to safety and society (such as those related to harmful bias, discrimination, threats to protection of privacy or
personal data, fairness)

Comment

**This document is intended solely for coordination purposes. All responses must be submitted online on www.oecd.ai/g7.**

The results of red-teaming conducted to evaluate the model's/system's fitness for moving beyond the development stage □

Comment

Capacities of a model/system and significant limitations in performance that have implications for the domains of appropriate use □

Comment

Other technical documentation if appropriate □

Comment

Instructions for use if relevant □

Comment

Other (open field) □

Comment

**G5.** **5. Does your organization demonstrate transparency related to advanced AI systems though any other methods in addition to such reports, for example through any technical documentation and/or instructions for use?***

Yes □

No □

**G6.** **If yes, please describe.***

**G7.** **Implementation documentation**

**Please share any relevant links or public documentation, including, for example: transparency reports, published red- teaming results, responsible use guidelines, and technical documentation regarding risk identification, assessment and management**

**G8.** **Feedback on the reporting framework Action 3**

**Please provide any feedback on the reporting framework questions for Action 3 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

# Section H: Action 4: Incident Management and Reporting

*Excerpt from the Code of Conduct: Action 4*

Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia

This includes responsibly sharing information, as appropriate, including, but not limited to evaluation reports, information on security and safety risks, dangerous intended or unintended capabilities, and attempts by AI actors to circumvent safeguards across the AI lifecycle.

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.

This should also include ensuring appropriate and relevant documentation and transparency across the AI lifecycle in particular for advanced AI systems that cause significant risks to safety and society.

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 relevant public authorities, as appropriate.

Such reporting should safeguard intellectual property rights.

**H1. 1. Does your organization share information with other stakeholders (other organizations, governments, civil society and academia, etc.) regarding the outcome of the evaluation of risks to an advanced AI system?***

Yes □

No □

**H2. If yes, if appropriate, please elaborate on the type of information shared, the frequency, and the ways in which it is communicated.**

**H3. 2. How does your organization share information, as appropriate, with relevant other stakeholders regarding advanced AI system incidents?***

**H4. 3. Are information sharing activities supported by relevant documentation?***

Yes □

No □

**H5. If yes, please elaborate.**

**H6.** **4. Does your organization share and report incident-related information publicly?***

Yes □

No □

**H7.** **If yes, please elaborate.**

**H8.** **5. Does your organization use shared incident reports of other organizations to help identify risks?***

Yes □

No □

**H9.** **If yes, please**

### H10. Implementation documentation

**Please share any relevant links or public documentation, including, for example: policies to report incidents to public authorities where appropriate, and to developing frameworks and platforms to enable such reporting, in particular for serious incidents, and / or responsible disclosure policies.**

### H11. Feedback on the reporting framework Action 4

**Please provide any feedback on the reporting framework questions for Action 4 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

# Section I: Action 5: Organizational Governance

*Excerpt from the Code of Conduct: Action 5*

Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures.

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.

This includes disclosing where appropriate privacy policies, including for personal data, user prompts and advanced AI system outputs. Organizations are expected to establish and disclose their AI governance policies and organizational mechanisms to implement these policies in accordance with a risk-based approach. This should include accountability and governance processes to evaluate and mitigate risks, where feasible throughout the AI lifecycle.

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.

Organizations should establish policies, procedures, and training to ensure that staff are familiar with their duties and the organization's risk management practices.

**I1. 1. What AI risk management and governance policies has your organization developed and/or implemented?***

**I2.** **2. When and under what circumstances are these policies updated?***

**I3.** **3. Does your organization communicate its risk management policies and practices or otherwise have a means for sharing this information with users and/or the public?***

Yes □

No □

**I4.** **If yes, please provide references.**

**I5.** **4. Does your organization disclose privacy policies addressing the use of personal data, user prompts, and/or the outputs of advanced AI systems?***

Yes □

No □

**I6.** **If yes, please elaborate.**

**I7.** **5. Are relevant staff trained on your organization’s governance policies and risk management practices?***

Yes □

No □

**I8.** **If yes, please provide appropriate references/examples.***

**I9.** **Implementation documentation**

**Please share any relevant links or public documentation, including, for example: responsible AI / AI governance policies, privacy policies and policies outlining adoption of risk management frameworks.**

**I10. Feedback on the reporting framework Action 5**

**Please provide any feedback on the reporting framework questions for Action 5 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

# Section J: Action 6: Information Security

*Excerpt from the Code of Conduct: Action 6*

Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.

These may include securing model weights and, algorithms, servers, and datasets, such as through operational security measures for information security and appropriate cyber/physical access controls.

This also includes performing an assessment of cybersecurity risks and implementing cybersecurity policies and adequate technical and institutional solutions to ensure that the cybersecurity of advanced AI systems is appropriate to the relevant circumstances and the risks involved. 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. This includes a commitment to have in place a vulnerability management process and to regularly review security measures to ensure they are maintained to a high standard and remain suitable to address risks.

This further includes 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.

**J1. 1. How does your organization implement AI-specific measures in its information security practices pertaining to operational and cyber/physical security?***

J2. **2. How does your organization assess cybersecurity risks and implement cybersecurity policies to enhance the cybersecurity of advanced AI systems?***

J3. **3. What measures are in place to ensure that the storage of and work with model weights, algorithms, servers, datasets or other relevant elements are done in in an appropriately secure environment, with limited access to reduce both the risk of unsanctioned release and the risk of unauthorized access?***

J4. **4. What is your vulnerability management process?***

**J5.** **5. How often are security measures reviewed?***

**J6.** **6. Does your organization have an insider threat detection program?***

Yes □

No □

**J7.** **How does the program protect the most valuable intellectual property and trade secrets, for example, by limiting access to proprietary and unreleased model weights?***

**J8. Implementation documentation**

**Please share any relevant links or public documentation, including, for example: security or governance policies, adapted to AI, vulnerability management process and regular review of security measures, and policies / information on insider threat detection.**

**J9. Feedback on the reporting framework Action 6**

**Please provide any feedback on the reporting framework question for Action 6 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

# Section K: Action 7: Content Authentication and Provenance

*Excerpt from the Code of Conduct: Action 7*

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. The provenance data should include an identifier of the service or model that created the content but need not include user information.

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. Organizations should collaborate and invest in research, as appropriate, to advance the state of the field.

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.

**K1.** **1. Is your organization developing and deploying mechanisms such as labelling or watermarking that enables users to identify content generated by advanced AI systems developed by your organization?***

Yes □

No □

**K2.** **If yes, please elaborate.**

K3. 2. What mechanisms, if any, does your organization put in place, such as disclaimers, to allow users, where possible and appropriate, to know when they are interacting with an advanced AI system developed by your organization?*

K4. 3. How does your organization collaborate on and invest in research to advance the state of content authentication and provenance?*

K5. Implementation documentation

Please share any relevant links or public documentation, including, for example: policies for implementing content authentication, provenance, and/or labeling mechanisms for different forms of AI-generated content, and / or information about the use of industry standards and/or specifications for content authentication and provenance.

**K6.** **Feedback on the reporting framework Action 7**

**Please provide any feedback on the reporting framework questions for Action 7 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

## Section L: Action 8: Research and Investment to Advance AI Safety and Mitigate Societal Risks

*Excerpt from the Code of Conduct: Action 8*

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, as well as investing in developing appropriate mitigation tools.

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 rights and privacy, and avoiding harmful bias, mis- and disinformation, and information manipulation. 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.

Organizations are encouraged to share research and best practices on risk mitigation.

**L1.** **1. Does your organization participate in projects, collaborations and investments in research that support the advancement of AI safety, security, and trustworthiness, as well as risk evaluation and mitigation tools?***

Yes □

No □

**L2. If yes, please select all that apply and elaborate/provide examples.**

Upholding democratic values and respecting human rights □

Comment

Protecting children and vulnerable groups □

Comment

Safeguarding intellectual property rights □

Comment

Safeguarding privacy □

Comment

Avoiding harmful bias □

Comment

Avoiding mis- and disinformation and information manipulation □

Comment

Other

Commen

L3. **2. How does your organization share research and best practices on addressing or mitigating risk?***

L4. **3. What research or investment is your organization pursuing to maximize socio-economic and environmental benefits from AI?***

*If possible, please elaborate/provide examples.*

L5. Implementation documentation

Please share any relevant links or public documentation, including, for example: reports (e.g., human rights or corporate responsibility reports) or overviews of research publications.

L6. Feedback on the reporting framework Action 8

Please provide any feedback on the reporting framework question for Action 8 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.

# Section M: Action 9: Advancing Human and Global Interests

*Excerpt from the Code of Conduct: Action 9*

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 These efforts are undertaken in support of progress on the United Nations Sustainable Development Goals, and to encourage AI development for

global benefit.

Organizations should prioritize responsible stewardship of trustworthy and human-centric AI and also 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. Organizations should work with civil society and community groups to identify priority challenges and develop innovative solutions to address the world's greatest challenges.

**M1.** **1. Does your organization prioritize AI projects for responsible stewardship of trustworthy and human-centric AI in support of the UN Sustainable Development Goals?***

Yes □

No □

**M2.** **If yes, please elaborate/provide examples.**

**M3.** **2. Does your organization support any digital literacy, education or training initiatives to improve user awareness and/or help people understand the nature, capabilities, limitations and impacts of advanced AI systems?***

Yes □

No □

**M4.** **If yes, please elaborate/provide examples.**

**M5.** **3. Does your organization collaborate with civil society and community groups to identify and develop solutions to address the world’s greatest challenges?***

Yes □

No □

**M6.** **If yes, please elaborate/provide examples.**

**M7.** **Implementation documentation**

**Please share any relevant links or public documentation, including, for example: corporate responsibility reports.**

**M8. Feedback on the reporting framework Action 9**

**Please provide any feedback on the reporting framework question for Action 9 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

## Section N: Action 10: International Interoperability and Standards

*Excerpt from the Code of Conduct: Action 10*

Advance the development of and, where appropriate, adoption of international technical standards

Organizations are encouraged to contribute to the development and, where appropriate, use of international technical standards and best practices, including for watermarking, and working with Standards Development Organizations (SDOs), also when developing organizations' testing methodologies, content authentication and provenance mechanisms, cybersecurity policies, public reporting, and other measures. 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.

**N1. 1. Does your organization contribute to the development of international technical standards and best practices relating to AI in standards development organizations (SDOs) and/or other relevant bodies?***

Yes □

No □

**N2.** **If yes, please elaborate.**

**N3.** **2. Has your organization adopted international technical standards relating to AI?***

Yes □

No □

**N4.** **If so, which ones?***

**N5.** **3. Does your organization use international technical standards and best practices when developing testing methodologies, content authentication and provenance mechanisms, cybersecurity policies, public reporting and/or other measures?***

Yes □

No □

**N6.** **If yes, please elaborate.**

**N7.** **Implementation documentation**

**Please share any relevant links or public documentation, including, for example: policies outlining development and/or adoption of technical standard.**

**N8.** **Feedback on the reporting framework Action 10**

**Please provide any feedback on the reporting framework question for Action 10 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

## Section O: Action 11: Data Input Measures and Protections for Personal Data and IP

*Excerpt from the Code of Conduct: Action 11*

Implement appropriate data input measures and protections for personal data and intellectual property

Organizations are encouraged to take appropriate measures to manage data quality, including training data and data collection, to mitigate against harmful biases.

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. Organizations are encouraged to implement appropriate safeguards, to respect rights related to privacy and intellectual property, including copyright-protected content.

Organizations should also comply with applicable legal frameworks.

**O1.** **1. What measures does your organization take to promote data quality and mitigate harmful biases throughout the AI lifecycle, including in training and data collection processes?***

**O2.** **2. What measures does your organization take to guard against systems divulging confidential or sensitive data?***

**This document is intended solely for coordination purposes. All responses must be submitted online on www.oecd.ai/g7.**

O3. **3. How does your organization protect privacy throughout the AI lifecycle?***

O4. **4. How does your organization protect intellectual property, including copyright-protected content throughout the AI lifecycle?***

O5. **5. How do you manage compliance with applicable legal frameworks across different legal jurisdictions?***

O6. **Implementation documentation**

**Please share any relevant links or public documentation, including, for example: privacy policies, intellectual property protections relevant data-related commitments, and / or compliance arrangements with applicable legal frameworks.**

O7. **Feedback on the reporting framework Action 11**

**Please provide any feedback on the reporting framework question for Action 11 including, for example: clarity of the questions, ease of answering the questions, time required to formulate responses, and consultation required to formulate a response.**

## Section P: Any other information

**P1.** **This optional open response section allows organizations to self-report on additional topics that likely pertain to related voluntary initiatives. This supports comprehensive understanding of their efforts and helps to streamline the reporting process for efficiency and comprehensiveness.**

## Section Q: Feedback on Reporting Framework

**Q1.** **Feedback on the reporting framework**

**Please provide any feedback on the reporting framework including, for example, the design, structure, comprehensiveness, ease of responding and overall time to respond.**

**Q2. Public availability of responses**

**It is envisioned that the reporting framework will be finalized and launched officially following the pilot phase. Please share your views on making public individual responses in the next phase of the reporting framework?*** *Reminder: The responses to this current pilot will be collected and recorded by the OECD Secretariat for the sole purpose of refining and improving the reporting framework in the next phase.*

**Thank you for participating in the pilot phase of the reporting framework for the Code of Conduct for Organizations Developing Advanced AI Systems. Please note that we may contact you for further information regarding your responses to the survey. We appreciate your time and contribution.**