Hiroshima Process International Guiding Principles for All AI Actors

Hiroshima AI Process. (2023, October 30). Engaging with artificial intelligence. Soumu.go.jp. <https://www.soumu.go.jp/hiroshimaaiprocess/en/documents.html>

Document contains 12 principles for all AI actors

1. Take appropriate measures throughout the development of advanced AI systems to identify, evaluate, and mitigate risks across the AI lifecycle.
2. Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market.
3. Publicly report advanced AI systems’ capabilities, limitations and domains of appropriate and inappropriate use, for transparency and accountability.
4. Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems.
5. Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures.
6. Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.
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.
8. Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.
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.
10. Advance the development of and, where appropriate, adoption of international technical standards.
11. Implement appropriate data input measures and protections for personal data and intellectual property.
12. Promote and contribute to trustworthy and responsible use of advanced AI systems