

## Task Force 4: Science and Digitalization for a Better Future



# Towards Safe, Secure, and Trustworthy AI: Implementing the G7 AI Hiroshima Policy Framework

**Daniele Gerundino**, Centro Studi per la Normazione, Italian National Standards Body (UNI);  
Research associate, Institut de Gouvernance de l'Environnement et Développement Territorial  
(GEDT), University of Geneva, Switzerland

**Conor Hearn**, CEO Lussolo, Director Vasioe Corp (Holdings)

**Paul Alan McAllister**, President and CEO, Global Leaders in Unity and Evolvment (GLUE),  
United States

**Vidisha Mishra**, Head of Community Engagement Programme, Global Solutions Initiative,  
Germany

**Simona Romiti**, Assistant Program Manager for Global Leaders in Unity and Evolvment  
(GLUE), United States

**Alice Saltini**, Research Coordinator, European Leadership Network (ELN)

**Dennis J. Snower**, President, Global Solutions Initiative, Germany

**Paul Twomey**, Fellow and Core Theme Leader, Global Solutions Initiative, Germany

## Abstract

---

The rapid acceleration in the development of artificial intelligence (AI) poses significant potential for society along with large scale risks. The transformative potential of AI cannot be realised without the protection of fundamental human rights and dignity and the principles of transparency and fairness. G7 countries have an opportunity to take a leading role in the governance, development, and distribution of this technology, paving the way to the engagement of G20 countries and the whole international community. In this brief we explore potential solutions to large scale problems regarding governance, standardisation, and the development of advanced AI. The proposals relate to the development of standards, the use of scientific collaboration as a method for development and governance, the essential role of human-centred education, and digital empowerment of people.

## The challenge

---

The exponential increases in computational power and digital data have created the conditions for the current explosion of artificial intelligence (AI). AI based on artificial neural networks (ANNs) and machine learning/deep learning have led to generative AI such as large language and multimodal models. These demonstrate unprecedented capabilities in a broad variety of areas and are re-shaping research and development activities, business processes, and professions. These technologies are far from mature and the transformational potential and substantial benefits to society are only starting to become clear.

At the same time, they pose substantial threats to the fabric of society by contributing to:

- violating fundamental human rights;
- increasing inequalities;
- spreading disinformation;
- altering democratic processes (e.g. via disinformation and deep fakes targeting electors);
- damaging minorities and disadvantaged people (e.g. through profiling and other practices subject to biases);
- violating intellectual property rights (e.g. by capturing and using content subject to copyright);
- bringing vast and perhaps irreversible harm (autonomous weapons).

The development of AI is driven by a few oligopolistic actors (“Big Tech”) that prioritize profit and consolidation of power leading to dynamics which increase these risks (Yampolskiy 2024; CFRM 2023). Data is essential to train AI models, and the present digital data system is based on a misalignment between digital citizens/consumers and those who collect and monetize their data (Snower and Twomey 2022), contributing to the risks highlighted above. Looking forward, many

researchers working in the field state that there is a possibility of existential risks to humanity.<sup>1</sup> The development of artificial general intelligence (AGI) if not developed properly has the potential to escape our control or lead to significant unpredictable effects on society.<sup>2</sup>

The international community is aware of the potential and risks posed by AI. Initiatives to steer the development of safe, secure, and trustworthy AI are being pursued by Governments, the UN, the G7 and many other organizations.<sup>3</sup>

Taking all this into consideration, this policy brief aims at contributing to the AI governance harmonization process undertaken by the G7, by providing recommendations on key issues and actions including the development of standards, the creation of a scientific collaboration, for the promotion of human-centred education, and rights for personal data.

## The role of the G7

---

The G7 can take a leading role in enabling the development of safe, secure and trustworthy AI. This Policy Brief highlights a set of important actions described in “Recommendations” aiming to operationalize the G7 AI Hiroshima process. They could be endorsed by the G7, implemented by its members and actively promoted to ensure the broader engagement of the G20 and the whole international community.

- 1) Provide authoritative input on strategic directions and priorities for standardisation and conformity assessment, aiming to operationalize the policy framework approved by the G7 in 2023 and contributing to regulatory harmonization among the G7 countries and beyond.
- 2) Launch an international scientific collaboration with the goal of developing safe and transparent advanced AI as a coordinated, large-scale effort between leading research and academic organizations. The initiative can lead to the establishment of a Research Organization for General Intelligence (ROGI), acting as a promoter of open and responsible innovation and a counterbalance to current oligopolistic trends.

---

<sup>1</sup> About half of the respondents to the Expert Survey on AI 2023 (involving 2,778 AI professionals) gave at least 10 per cent chance that the impact of advanced AI would be catastrophic for humanity (e.g. human extinction). See Grace 2024.

<sup>2</sup> “Let an ultraintelligent machine [a concept close to what we now call AGI] be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion,’ and the intelligence of man would be left far behind.” See Good 1966: 33.

<sup>3</sup> There is a multitude of policy proposals issued by a variety of organizations and a few regulatory frameworks adopted or in the process of being adopted by some jurisdictions. Here we provide as reference the latest set of documents published by a selection of authoritative organizations that we think are of utmost importance for the G7: G7 2023b; UN AI Advisory Body 2023; European Parliament and Council of the EU 2024; White House 2023.

3) Promote a broad education-oriented effort aiming to ensure that human-centred values be foundational components of AI university curricula, and of the various existing forms of vocational training. Contribute to open a dialogue among academic institutions and leading technology companies aiming to spread a humanistic culture and embed human values into the inner fabrics of AI systems.

4) Recommend the adoption by G7 member countries of policies and implementation of measures aimed to ensure the fundamental right of “digital citizens” to full ownership and control of personal data. Considering the key role of data for AI systems, this action is considered an essential element for just digital societies – and is correlated to intellectual property protection measures.

## Recommendations to the G7

---

The policy framework developed by the G7 in 2023 (notably the International Principles and Code of Conduct) is taken as a key reference: the recommendations focus on strategic developments and specific actions aiming at operationalising the Principles and making further progress towards the development of safe, secure, and trustworthy AI. They refer directly to the G7 and its member countries: but their role is intended as seminal to reach out to and engage the G20 countries and the whole international community.

Taking stock of the various perspectives and experiences of the communities represented by the authors, the recommendations address a set of essential, interrelated matters:

### Recommendation #1

---

Strategically promote standardisation and conformity assessment frameworks as fundamental instruments to operationalise the 11 principles of AI through an open, multi-stakeholder process. This recommendation is directly related to Principle 10 (“Advance the development of and, where appropriate, adoption of international technical standards”), with broader implications concerning the implementation of most of the 11 principles.

Most policy proposals<sup>4</sup> for AI governance recognize international standards as essential instruments for the development of trustworthy AI and for supporting innovation for good.

---

<sup>4</sup> Here we give just a few references: G7 AI Principle 10 (“Advance the development of and, where appropriate, adoption of international technical standards”) (G7 2023c); the G7 Digital & Tech Ministers’ Statement of 7 September 2023 (“We reaffirm our commitment to promote the development and adoption of international standards and interoperable tools for trustworthy AI that enables innovation”) (G7 2023a); the UN AI Advisory Body interim report, points 60, 63, 64, 65 and 66 (UN AI Advisory Body 2023).

International standardization (IS) is extremely effective at fostering the adoption of good practices through the development of clear and shared description of characteristics, criteria, rules and guidelines regarding materials, products, systems, and processes. This can be seen in a broad variety of fields where voluntary standards complement or give a basis for regulations, helping to “operationalize” regulatory goals, such as chemical and mechanical industries, construction, electrotechnology, health care, transportation, and many more.

Standards also provide the basis for conformity assessment activities, such as testing, inspection and certification. These activities are deemed explicitly or implicitly essential by all the AI governance systems under development.

International standardisation in the AI field is already taking place with works such as the ISO/IEC committee JCT 1/SC 42<sup>5</sup> and substantial standardisation work will be undertaken in Europe by CEN/Cenelec,<sup>6</sup> once the AI Act enters into force. However, it is non-trivial to ensure the contribution of standardisation is as effective and timely as required. This is due to the broad variety of technical, societal, and ethical issues to be addressed, the breakneck speed of AI development and the “bottom-up” nature of standardisation’s work.

The G7, by means of a coordinated effort of its members, can provide an invaluable contribution by forwarding clear input consistently with the G7 policies helping to define strategic directions and prioritization for standardisation in AI, respecting the independence and development processes of standards organizations.

The G7 could undertake the following actions:

a) Establish a special advisory group to provide input and interface key standards organizations. (Standards developing organizations such as the Italian National Standards Body (UNI) are available to provide secretariat services to such group starting in 2024, under the G7 Italian Presidency).

b) The group could select and consult with key actors including organizations such as the Global Partnership on Artificial Intelligence (GPAI) and groups from the OECD AI Policy Observatory, top international experts with diverse backgrounds, and representatives of stakeholders able to express the expectations and concerns of the Global South. With this aiming to define and

---

<sup>5</sup> The ISO/IEC JTC 1/SC 42 Artificial intelligence (<https://www.iso.org/committee/6794475.html>) has already published 25 international standards addressing various matters related to AI and 31 are currently under development – with many of them of significant interest from the G7 perspective.

<sup>6</sup> This is related to a distinctive aspect of the EU policy framework for products (in a broad sense). EU binding legal acts (directives, regulations and decisions) set compulsory rules addressing “essential requirements”. These can be complemented by voluntary standards, among which particularly important are the “harmonized standards” developed by the European Standardisation Organisations (ESOs), upon request (“mandate”) of the European Commission. A comprehensive description of the relevant EU policies can be found in European Commission 2022.



promote a set of high priority standardisation projects that would help operationalize the G7's proposed comprehensive policy framework.

For example, the first clause of the G7 Code of conduct (Take appropriate measures [...] to identify, evaluate, and mitigate risks), indicates that "This includes employing diverse internal and independent external testing measures, through a combination of methods for evaluations".

A technical standard or suite of standards addressing this could provide a precise description agreed with a double level of consensus (among stakeholders and across countries), of:

- The types of risks posed by AI systems and the criteria to be used to assess them.
- The methods to be applied for risks assessment including types of testing, inspection and/or certification.
- Required technical documentation; use of stakeholders' feed-back, and more).
- The appropriate modalities for evaluating results.
- The conformity assessment schemes to be applied.

Similar considerations apply to most of the Principles and elements of the Code of Conduct.

c) One or more focused workshops could be organized in 2024, to gather multistakeholder input, help establish priorities and precisely characterize the standards projects advocated.

d) The group could then actively interact with the main standards organizations, ensuring that the G7-backed proposals would be well understood and, if appropriate, taken up expeditiously in their work programmes.

## **Recommendation #2**

---

Scientific collaboration as a method for building secure and trustworthy AI and ensuring its equitable distribution. This recommendation is related to the G7 AI Principle 8 ("Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures") as well as to several others.

A new entity should be established: the Research Organization for General Intelligence (ROGI), with the goal of developing, understanding, steering, and distributing human and superhuman level AI. The proposal takes inspiration from previous successful scientific collaborations, such as the European Organization for Nuclear Research (CERN) and the Human Genome Project. The G7 and its member countries (or some of them) can take an essential role as initiators of the initiative, aiming to engage in due course the G20 and all other countries, with particular attention to those of the Global South.

An international scientific project can ensure that this technology is developed safely and responsibly and is distributed in an equitable manner. Scientific collaboration can create dynamics ensuring safe development without hindering technological advancements as other regulatory measures could. The distribution of research can lead to large scale economic benefits by providing the technology for use to a range of companies and organizations, avoiding issues stemming from consolidation and oligopolistic market conditions (Lynn et al. 2023). Other initiatives and experts have called for the establishment of a similar model based on international collaboration, underscoring the relevance and ongoing discussion about this governance model (Bojic et al. 2023). By thoroughly developing definitions, methods, and benchmarks the organization can oversee development ensuring transparency and safety (Blili-Hamelin et al. 2024).

The ROGI can be structured in various ways, consisting of a mixture of existing entities with a central body that balances research activities and policy issues with state led representation (CERN 2024). Working collaboratively with existing academic institutions, companies and other organizations provides members with greater access to research talent and computing resources incentivizing both states and other entities to join. With commitments from state parties and industry there is the possibility to expand computing and research capabilities within the central body for ROGI. There is a precedence for ROGI with organizations such as the AI alliance: committed to open innovation, the alliance relies on 80 billion US dollars in annual R&D funding and a diversified membership (large companies, start-ups, universities, research and government organizations, and non-profit foundations) (Markman 2024). This gives an idea of the incredible scale and possibilities of collaboration.

The process leading to ROGI would require an initial study to determine benchmarks and comprehensive testable definitions for AI systems. The ROGI will then be established to perform collaborative work, based on the Organization's charter or the blueprint agreed by the initial collaborating parties and founding members. It should be equipped with dedicated large-scale computational resources and foresee an in-depth interaction among research institutions and specialized AI regulatory agencies across the globe.

Key research goals are the creation of steerable human level AI systems (Park et al. 2022), research into interpretability and alignment, and the development of strategies for the equitable distribution of such systems in a way that maximizes societal benefits.

To establish a large-scale scientific collaboration several key milestones are needed, as outlined below.

*Milestone 1:* Initial feasibility study. The G7 could set-up a dedicated task force or request an authoritative organization (such as the OECD AI policy observatory, or a selection of think tanks specialized on AI and its long-term impacts) to perform a study to define:

- The key objectives, scope, and benefits of ROGI including a tentative work plan, key issues

and actors.

- Structure and modality of operation of the collaboration, possible partnerships with existing organizations such as the AI alliance.
- Determine resources required to establish the project (financial and in kind).

*Milestone 2:* Establishment of the ROGI. Based on the recommendations of the feasibility study:

- Launch the initiative establishing a charter and body for ROGI, oversight mechanisms, and collaborations.
- Ensure the commitment of a set of initial collaborating partners/founding members.
- Secure resources (financial or in-kind) supporting the operations.
- Publish its research plan and develop strategies to bolster the reach of ROGI.

*Milestone 3:* Dissemination of results and links with regulatory policy.

- Develop strategies, benchmarks, assessments for analysing AI models and research outputs produced through the collaboration.
- Develop strategies for dissemination of research outputs and models including methods for open distribution and limited distribution for models deemed to pose threats.

## **Recommendation #3**

---

Educate AI developers to advance AI understanding and human-centred values within the culture via whole education frameworks that address all AI actors at various levels (development, deployment, and use).

This recommendation is related to the G7 AI Principle 9: Prioritize the development of advanced AI systems to address the world’s biggest challenges, in particular, but not only, the climate crisis, global health and education.

While there are many areas of risk associated with AI implementation, i.e. programme bias, privacy violations, disinformation and more, there is perhaps no risk greater than the perceived threat of AI to human agency. Human agency is “an individual’s capacity to determine and make meaning from their environment through purposive consciousness, and reflective and creative action” (Parsell et al. 2017). Human agency is experienced in many ways, including through employment.<sup>7</sup> As many organizations in G7 countries seek to upskill their employees on how to develop, deploy and use AI, the number of job losses due to AI in areas of low complementarity between humans and machines continues to rise. AI has become a threat to work as a traditional

---

<sup>7</sup> Human agency involves four core properties: intentionality, forethought, self-reflectiveness, and self-reactiveness, all of which are essential components of employment. Human agency is a mindset plus a set of learnable actions that help persons attain their goals in life. The 2018 Pew Research Center survey indicates concerns about human agency, evolution and survival due to AI.



paradigm by which human agency is valued. It is vital for the G7 to identify a means within AI to affirm human agency and instil human-centric values with the culture. It is recommended that this be done via a commitment to the Design, Construction and Management of hybrid training or multidisciplinary education of persons responsible for the next generation of AI systems.<sup>8</sup>

The aim of this political commitment is an upstream, forward-thinking, and collaborative approach to meet the escalating ethical demand for a more talented human workforce. Talent, in this sense, hinges not only upon AI engineering skills, i.e. AI security, big data analysis, programming, but a deep respect for humanity. When facilitated by the G7, this method of instruction offers a guarantee that technological innovation will drive AI system design and development within ethical guardrails.<sup>9</sup> This ethical dimension seeks answers to the critical issues that will influence future work, like the boundaries of humanistic thinking, the meaning of personal privacy, the moral design of AI systems, how to build a radically more efficient planet, and how to protect democratic institutions. These and many other novel scenarios will require an in-depth knowledge of how AI systems are designed to work, and how humanity functions, both in relation to AI machines and other human and biological species. This ethical dimension also goes to the question of who is in charge and how the voices of human aspiration will be heard above the sounds of dedicated servers, fans, and electric cooling systems with the power to implement decisions that can either enhance or put human lives at risk.

A multidisciplinary educational framework committed to AI will also explore the problem of equity and ensure that AI's social and economic benefits do not accrue solely to the benefit of the Global North while the Global South lags behind in the AI revolution.<sup>10</sup>

An Oxford Insights assessment of 181 countries found that preparedness in using AI in public services showed the lowest-scoring regions in much of the Global South, such sub-Saharan Africa, some Central and South Asian countries, and some Latin American countries. The survey used 39 indicators across 10 dimensions which make up three pillars: Government, technology sector, and data and infrastructure. Perhaps the most extreme case is Africa, which by the year 2050 will have 25 per cent of the world's population. The AI preparedness score can be linked to literacy as Africa has 25 per cent of the adults worldwide (182 million out of 763 million) who lack literacy skills. Africa's educational attrition rate also increases from 20 per cent (ages of 6-11) to 33 per cent (ages 12-14) and reaches 60 per cent (ages 15-17).

---

<sup>8</sup> The field of artificial intelligence is a multidisciplinary one that spans the intersection of psychology, informatics, and philosophy. Students who are trained in such a rich multidisciplinary environment may have best opportunities for the future (Naidoo 2023).

<sup>9</sup> Nurturing the future through human-centric AI requires guardrails in the multiple areas: bias and ethics; collaboration between humans and AI, user-centred design, transparency and explainability, safety, joint human-AI testing, and training (Endsley 2024).

<sup>10</sup> The survey used 39 indicators across 10 dimensions which make up three pillars (government, technology sector and data and infrastructure) (Oxford Insights 2023).

Therefore, it is recommended that the G7:

- a) use education as a powerful tool to encourage collaboration between BigTech and educational institutions at all levels, starting with (public) universities;
- b) strengthen the bilateral relationship with the African Union to build an Agile Governance model on multidisciplinary educational systems for AI.<sup>11</sup>

## Recommendation #4

---

The implementation of ethical principles to AI requires empowering citizens to control personal data about them, both individually and collectively.

This recommendation aligns with the Hiroshima Guiding Principle 11 (“Implement appropriate data input measures and protections for personal data and intellectual property”), as well as with Principle 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”).

Data is at the core of ethical AI development (Solove 2024), but G7 countries show a lack of regulatory uniformity. A common characteristic among G7 digital governance regimes prevents their citizens from deciding how data about them is bought, sold, shared, and processed. It also allows most misuses of AI. This is rooted not in current frameworks but in the adoption and consolidation of the 50-year-old Fair Information Privacy Principles (King and Meinhardt 2024). In other words, AI challenges can be traced to citizens’ inability to decide the terms of use of personal data about them. This lack of control is the core enabler for unethical AI.

AI systems are trained on data generated by individuals and this use needs to account for the interests and goals of the originators, not just AI developers (Renieris 2023). The vast array of problems concerning the use of personal data for digital and AI systems are treated separately and addressed by separate policymakers, frameworks, and regulations. This siloed approach undermines efforts to address the root and common causes of these digital misfunctions, and instead reacts to their symptoms. Further, the rapid rate of technical progress outpaces policy responses – which continue to be reactive rather than proactive. Granted, policymaking and the development of normative systems are generally slow to catch up with unregulated or novel threats. It is unreasonable to expect otherwise. However, individualistic approaches to data

---

<sup>11</sup> Agile governance connects a series of related activities: multi-stakeholder collaboration; governing for the environment; the importance of values in governing; managing uncertainty; managing technology’s impact; making multilateralism more effective; and governing communication chaos. Agile governance means more than just coordinating effective, efficient, and reliable public and private institutions to effectively manage problems – the term implies a forward-looking approach that seeks to anticipate problems before they materialize. Demands for agile governance have never been greater amid a lingering pandemic, conflict, and ongoing crisis of multilateralism, and the persistent weakness of many national governance systems. See Hertie School 2024.

protection would make any regulatory efforts to protect citizens' personal data unfit. Citizens must be protected well before they enter digital interactions.

A significant source of the digital disempowerment is the “third-party funded digital barter” between digital consumers and third-party funders (such as advertisers and data brokers). The interests of the third-party funders are not aligned with the interests of the digital consumers (Lamdan 2023). This fundamental flaw is responsible for an array of serious problems, including inequities, inefficiencies, manipulation of digital consumers, as well as dangers to social cohesion and democracy.

Building on existing governance structures that promote the ability of digital citizens to exercise wide-ranging control over their individual personal data and over their data commons, we propose for the G7 to formulate and endorse a set of policies that reflect a coherent approach and a unified understanding to the ethical foundations of digital governance:

- **Ensure citizens' control over their personal data and who has access to it:** Citizens should have effective control mechanisms over who accesses data about them and should be able to negotiate the terms and conditions under which such data is processed, through easy-to-use technical tools and supporting institutions. For example, personal data normally used to enter legal contracts, should be maintained in trusted repositories authenticated by a public institution or legally accepted sources. Such data holdings should be open to audit, similar to financial reporting based on Generally Accepted Accounting Principles (GAAP).
- **Enable citizens to negotiate the terms under which their personal data is used:** Provide effective rights of association and representation for citizens to ensure that professionals skilled in digital markets can advise groups of citizens, and negotiate on their behalf, about what terms should apply to the use of their personal data by AI developers.
- **Protect vulnerable citizens by imposing fiduciary obligations on the use of inferred data:** Ensure that entities which process information about individuals abide by duty-of-care rules that commonly apply in the offline world (for instance, between doctors and patients) - the data can only be used in the “best interests” of the data subject. Fiduciary obligations include governments and citizens exerting effective oversight to data agents to ensure that data agents use their skills and expertise to promote and protect the interests of data subjects. The G7 can spearhead processes to ensure high levels of trust in these institutions, for example, through the anti-corruption task force.

While meaningful control over personal data is not a panacea, without it, unethical use of AI will likely grow in scale. The G7 countries boast the most mature digital governance regimes. Therefore, the G7 process is in a venue to spearhead the processes that tackle the systemic failures that have allowed unethical AI.

The proposals also help uphold basic human rights through the incentives generated by a data-just digital governance system. In a data-driven world where physical and digital identities are

inextricably linked, a holistic understanding of bodily autonomy, human agency and rights, that spans across the digital and physical spaces, is fundamental to building ethical AI systems and democratic digital societies.

## References

---

Blili-Hamelin, Borhane, Hancox-Li, Leif, and Smart, Andrew. 2024. Unsocial intelligence: A pluralistic, democratic, and participatory investigation of AGI Discourse. *arXiv* 21 February. <https://doi.org/10.48550/arXiv.2401.13142>

Bojic, Ljubisa, et al. 2023. CERN for AGI: A theoretical framework for autonomous simulation-based artificial intelligence testing and alignment. *arXiv* 14 December. <https://doi.org/10.48550/arXiv.2312.09402>

CERN. 2024. *Our governance*. <https://home.cern/node/5277>

CFRM-Stanford Center for Research on Foundation Models. 2023. *The foundation model transparency index*. <https://crfm.stanford.edu/fmti>

Endsley, Mica. 2024. Nurturing the future through human-centric AI guardrails. *Federal News Network* 11 March. <https://federalnewsnetwork.com/?p=4921900>

European Commission. 2022. *The 'Blue Guide' on the implementation of the product rules*. [https://single-market-economy.ec.europa.eu/node/1796\\_en](https://single-market-economy.ec.europa.eu/node/1796_en)

European Parliament and Council of the EU. 2024. *Final draft of the artificial intelligence act*. <https://artificialintelligenceact.eu/the-act>

G7. 2023a. *G7 Hiroshima AI process: G7 Digital & Tech ministers' statement*, 7 September. <http://www.g7.utoronto.ca/ict/2023-statement.html>

G7. 2023b. *G7 Hiroshima AI process: G7 Digital & Tech ministers' statement: Attachment: Hiroshima AI process comprehensive policy framework*, <http://www.g7.utoronto.ca/ict/2023-statement-2.html>

G7. 2023c. *Hiroshima process international guiding principles for organizations developing advanced AI system*. <http://www.g7.utoronto.ca/summit/2023hiroshima/231030-ai-principles.html>

Good, Irving John. 1966. Speculations concerning the first ultraintelligent machine. *Advances in Computers* 6: 31-88. [https://doi.org/10.1016/S0065-2458\(08\)60418-0](https://doi.org/10.1016/S0065-2458(08)60418-0)

Grace, Katja. 2024. Survey of 2,778 AI authors: six parts in pictures. *AI Impacts blog* 4 January. <https://blog.aiimpacts.org/p/2023-ai-survey-of-2778-six-things>

Hertie School. 2024. Agile governance. *World Economic Forum*. <https://intelligence.weforum.org/topics/a1Gb0000000pTDaEAM>

King, Jennifer, and Meinhardt, Caroline. 2024. Rethinking privacy in the AI era: Policy provocations for a data-centric world. *Stanford HAI White Papers*. <https://hai.stanford.edu/white-paper-rethinking-privacy-ai-era-policy-provocations-data-centric-world>

Lamdan, Sarah. 2023. *Data cartels: The companies that control and monopolize our information*. Stanford: Stanford University Press

Lynn, Barry, von Thun, Max, and Montoya, Karina. 2023. *AI in the public interest: Confronting the monopoly threat*. Open Markets Institute. <https://www.openmarketsinstitute.org/publications/report-ai-in-the-public-interest-confronting-the-monopoly-threat>

Markman, Jon. 2024. Zuckerberg's game-changer: AI Alliance versus Tech Giants. *Forbes* 8 January. <https://www.forbes.com/sites/jonmarkman/2024/01/08/zuckerbergs-game-changer-ai-alliance-vs-tech-giants>

Naidoo, Rajani. 2023. AI study in universities must be broad, multidisciplinary. *University World News* 16 August. <https://www.universityworldnews.com/post-mobile.php?story=20230816121408407>

Park, Deokgun, et al. 2022. A definition and a test for human-level artificial intelligence. *arXiv* 14 December. <https://doi.org/10.48550/arXiv.2011.09410>

Parsell, Cameron, Eggins, Elizabeth, and Marston, Greg. 2017. Human agency and social work research: A systematic search and synthesis of social work literature. *British Journal of Social Work* 47(1): 238-255. <https://doi.org/10.1093/bjsw/bcv145>

Pew Research Center. 2018. *Artificial intelligence and the future of humans*. <https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans>

Oxford Insights. 2023. *Government AI readiness index 2023*. <https://oxfordinsights.com/?p=113>

Renieris, Elizabeth M. 2023. *Beyond data: Reclaiming human rights at the dawn of the metaverse*. Cambridge: MIT Press

Snowder, Dennis J., and Twomey, Paul. 2022. *Empowering digital citizens: Making humane markets work in the digital age*. GIDE. <https://www.global-solutions-initiative.org/programs/digital-empowerment/empowering-digital-citizens-report>

Solove, Daniel J. 2024. Artificial intelligence and privacy. SSRN 15 March. <https://doi.org/10.2139/ssrn.4713111>

UN AI Advisory Body. 2023. *Interim report: Governing AI for humanity*. <https://www.un.org/en/ai-advisory-body>

White House. 2023. *Executive Order on the safe, secure, and trustworthy development and use of artificial intelligence*. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>

Yampolskiy, Roman V. 2024. *AI: unexplainable, unpredictable, uncontrollable*. Boca Raton: CRC Press

## About Think7

---

Think7 (T7) is the official think tank engagement group of the Group of 7 (G7). It provides research-based policy recommendations for G7 countries and partners. The Istituto Affari Internazionali (IAI) and Istituto per gli Studi di Politica Internazionale (ISPI) are the co-chairs of T7 under Italy's 2024 G7 presidency.