Responsible AI Policy Framework

Responsible AI

A GLOBAL POLICY FRAMEWORK

Principle 1
Ethical Purpose and Societal Benefit

Organisations that develop, deploy or use AI systems and any national laws that regulate such use should require the purposes of such implementation to be identified and ensure that such purposes are consistent with the overall ethical purposes of beneficence and non-maleficence, as well as the other principles of the Policy Framework for Responsible AI.

1 Overarching principles

1.1 Organisations that develop, deploy or use AI systems should do so in a manner compatible with human agency and the respect for fundamental human rights (including freedom from discrimination).

1.2 Organisations that develop, deploy or use AI systems should monitor the implementation of such AI systems and act to mitigate against consequences of such AI systems (whether intended or unintended) that are inconsistent with the ethical purposes of beneficence and non-maleficence, as well as the other principles of the Policy Framework for Responsible AI set out in this framework.

1.3 Organisations that develop, deploy or use AI systems should assess the social, political and environmental implications of such development, deployment and use in the context of a structured Responsible AI Impact Assessment that assesses risk of harm and, as the case may be, proposes mitigation strategies in relation to such risks.

2 Work and automation

2.1 Organisations that implement AI systems in the workplace should provide opportunities for affected employees to participate in the decision-making process related to such implementation.

2.2 Consideration should be given as to whether it is achievable from a technological perspective to ensure that all possible occurrences should be pre-decided within an AI system to ensure consistent behaviour. If this is not practicable, organisations developing, deploying or using AI systems should consider at the very least the extent to which they are able to confine the decision outcomes of an AI system to a reasonable, non-aberrant range of responses, taking into account the wider context, the impact of the decision and the moral appropriateness of "weighing the unweighable" such as life vs. life.

2.3 Organisations that develop, deploy or use AI systems that have an impact on employment should conduct a Responsible AI Impact Assessment to determine the net effects of such implementation.

2.4 Governments should closely monitor the progress of AI-driven automation in order to identify the sectors of their economy where human workers are the most affected. Governments should actively solicit and monitor industry, employee and other stakeholder data and commentary regarding the impact of AI systems on the workplace and should develop an open forum for sharing experience and best practices.

2.5 Governments should promote educational policies that equip all children with the skills, knowledge and qualities required by the new economy and that promote life-long learning.

2.6 Governments should encourage the creation of opportunities for adults to learn new useful skills, especially for those displaced by automation.

2.7 Governments should study the viability and advisability of new social welfare and benefit systems to help reduce, where warranted, socio-economic inequality caused by the introduction of AI systems and robotic automation.

3 Environmental impact

3.1 Organisations that develop, deploy or use AI systems should assess the overall environmental impact of such AI systems, throughout their implementation, including consumption of resources, energy costs of data storage and processing and the net energy efficiencies or environmental benefits that they may produce. Organisations should seek to promote and implement uses of AI systems with a view to achieving overall carbon neutrality or carbon reduction.

3.2 Governments are encouraged to adjust regulatory regimes and/or promote industry self-regulatory regimes concerning market-entry and/or adoption of AI systems in a way that the possible exposure (in terms of 'opportunities vs. risks') that may result from the public operation of such AI systems is reasonably reflected. Special regimes for intermediary and limited admissions to enable testing and refining of the operation of the AI system can help to expedite the completion of the AI system and improve its safety and reliability.

3.3 In order to ensure and maintain public trust in final human control, governments should consider implementing rules that ensure comprehensive and transparent investigation of such adverse and unanticipated outcomes of AI systems that have occurred through their usage, in particular if these outcomes have lethal or injurious consequences for the humans using such systems. Such investigations should be used for considering adjusting the regulatory framework for AI systems, in particular to develop, where practicable and achievable, a more rounded understanding of how and when such systems should gracefully handover to their human operators in a failure scenario.

3.4 AI has a particular potential to reduce environmentally harmful resource waste and inefficiencies. AI research regarding these objectives should be encouraged. In order to do so, policies must be put in place to ensure the relevant data is accessible and usable in a manner consistent with respect for other principles of the Policy Framework for Responsible AI such as Fairness and Non-Discrimination, Open Data and Fair Competition and Privacy, Lawful Access and Consent.

4 Weaponised AI

4.1 The use of lethal autonomous weapons systems (LAWS) should respect the principles and standards of and be consistent with international humanitarian law on the use of weapons and wider international human rights law.

4.2 Governments should implement multilateral mechanisms to define, implement and monitor compliance with international agreements regarding the ethical development, use and commerce of LAWS.

4.3 Governments and organisations should refrain from developing, selling or using lethal autonomous weapon systems (LAWS) able to select and engage targets without human control and oversight in all contexts.

4.4 Organisations that develop, deploy or use AI systems should inform their employees when they are assigned to projects relating to LAWS.

5 The weaponisation of false or misleading information

5.1 Organisations that develop, deploy or use AI systems to filter or promote informational content on internet platforms that is shared or seen by their users should take reasonable measures, consistent with applicable law, to minimise the spread of false or misleading information where there is a material risk that such false or misleading information might lead to significant harm to individuals, groups or democratic institutions.

5.2 AI has the potential to assist in efficiently and pro-actively identifying (and, where appropriate, suppressing) unlawful content such as hate speech or weaponised false or misleading information. AI research into means of accomplishing these objectives in a manner consistent with freedom of expression should be encouraged.

5.3 Organisations that develop, deploy or use AI systems on platforms to filter or promote informational content that is shared or seen by their users should provide a mechanism by which users can flag potentially harmful content in a timely manner.

5.4 Organisations that develop, deploy or use AI systems on platforms to filter or promote informational content that is shared or seen by their users should provide a mechanism by which content providers can challenge the removal of such content by such organisations from their network or platform in a timely manner.

5.5 Governments should provide clear guidelines to help Organisations that develop, deploy or use AI systems on platforms identify prohibited content that respect both the rights to dignity and equality and the right to freedom of expression.

5.6 Courts should remain the ultimate arbiters of lawful content.

Principle 2
Accountability

Organisations that develop, deploy or use AI systems and any national laws that regulate such use shall respect and adopt the eight principles of this Policy Framework for Responsible AI (or other analogous accountability principles). In all instances, humans should remain accountable for the acts and omissions of AI systems.

1 Accountability

1.1 Organisations that develop, deploy or use AI systems shall designate an individual or individuals who are accountable for the organisation's compliance with those principles.

1.2 The identity of the individual(s) designated by the organisation to oversee the organisation's compliance with the principles shall be made known upon request.

1.3 Organisations that develop, deploy or use AI systems shall implement policies and practices to give effect to the principles if the Policy Framework for Responsible AI or other adopted principles (including analogous principles that may be developed for a specific industry), including:

- **i.** establishing processes to determine whether, when and how to implement a "Responsible AI Impact Assessment" process;
- **ii.** establishing and implementing "Responsible AI by Design" principles;
- **iii.** establishing procedures to receive and respond to complaints and inquiries;
- **iv.** training staff and communicating to staff information about the organisation's policies and practices; and
- **v.** developing information to explain the organisation's policies and procedures.

2 Government

2. Governments that assess the potential for "accountability gaps" in existing legal and regulatory frameworks applicable to AI systems should adopt a balanced approach that encourages innovation while militating against the risk of significant individual or societal harm.

2.1 Any such legal and regulatory frameworks should promote the eight principles of the Policy Framework for Responsible AI or encompass similar considerations.

2.2 Governments should not grant distinct legal personality to AI systems, as doing so would undermine the fundamental principle that humans should ultimately remain accountable for the acts and omissions of AI systems.

3 Contextual approach

3.1 The intensity of the accountability obligation will vary according to the degree of autonomy and criticality of the AI system. The greater the level of autonomy of the AI system and the greater the criticality of the outcomes that it may produce, the higher the degree of accountability that will apply to the organisation that develops, deploys or uses the AI system.

Principle 3
Transparency and Explainability

Organisations that develop, deploy or use AI systems and any national laws that regulate such use shall ensure that, to the extent reasonable given the circumstances and state of the art of the technology, such use is transparent and that the decision outcomes of the AI system are explainable.

1 Definitions

1.1 **Transparency** is an obligation for organisations that use AI in decision-making processes to provide information regarding: a) the fact that an organisation is using an AI system in a decision-making process; b) the intended purpose(s) of the AI system and how the AI system will and can be used; (c) the types of data sets that are used by the AI system; and (d) meaningful information about the logic involved.

1.2 **Explainability** is an obligation for organisations that use AI in decision-making processes to provide accurate information in humanly understandable terms explaining how a decision/outcome was reached by an AI system.

2 Purpose

2.1 Transparency and Explainability aim to preserve the public's trust in AI systems and provide sufficient information to help ensure meaningful accountability of an AI system's developers, deployers and users, and to demonstrate whether the decisions made by an AI system are fair and impartial.

2.2 The Transparency and Explainability principles support the Accountability principle, the Fairness and Non-Discrimination principle, the Safety and Reliability principle and the Privacy, Lawful Use and Consent principles.

3 Gradual and contextual approach

3.1 The intensity of the obligations of transparency and explainability will depend on the context of the decision and its consequences for the person subject to it. The scope and intensity of the obligations of transparency and explainability will increase as the sensitivity of the data sets used by an AI system increases and as the decisional outcome of an AI system increases in materiality.

3.2 The determination of the intensity of the obligations of transparency and explainability must balance the interests of the person subject to the decision and the interests of the organisation making the decision. The ultimate criteria shall be the reasonable expectations of a person subject to that type of decision.

4 Transparency and explainability by design

4.1 Organisations that develop AI systems should ensure that the system logic and architecture serves to facilitate transparency and explainability requirements. In so far as is reasonably practicable, and taking into account the state of the art at the time, such systems should aim to be designed from the most fundamental level upwards to promote transparency and explainability by design. Where there is a choice between system architectures which are less or more opaque, the more transparent option should be preferred.

4.2 Users of AI systems and persons subject to their decisions must have an effective way to seek remedy in the event that organisations that develop, deploy or use AI systems are not transparent about their use.

5 Technological neutrality

5.1 The use of an AI system by a public or private organisation does not reduce the procedural and substantive requirements that are normally attached to a decision when the decision-making process is completely controlled by a human.

Principle 4
Fairness and Non-Discrimination

Organisations that develop, deploy or use AI systems and any national laws that regulate such use shall ensure the non-discrimination of AI outcomes, and shall promote appropriate and effective measures to safeguard fairness in AI use.

1 Awareness and education

1.1 Awareness and education on the possibilities and limits of AI systems is a prerequisite to achieving fairer outcomes.

1.2 Organisations that develop, deploy or use AI systems should take steps to ensure that users are aware that AI systems reflect the goals, knowledge and experience of their creators, as well as the limitations of the data sets that are used to train them.

2 Technology and fairness

2.1 Decisions based on AI systems should be fair and non-discriminatory, judged against the same standards as decision-making processes conducted entirely by humans.

2.2 The use of AI systems by organisations that develop, deploy or use AI systems and Governments should not serve to exempt or attenuate the need for fairness, although it may mean refocussing applicable concepts, standards and rules to accommodate AI.

2.3 Users of AI systems and persons subject to their decisions must have an effective way to seek remedy in discriminatory or unfair situations generated by biased or erroneous AI systems, whether used by organisations that develop, deploy or use AI systems or governments, and to obtain redress for any harm.

3 Development and monitoring of AI systems

3.1 AI development should be designed to prioritise fairness. This would involve addressing algorithms and data bias from an early stage with a view to ensuring fairness and non-discrimination.

3.2. Organisations that develop, deploy or use AI systems should remain vigilant to the dangers posed by bias. This could be achieved by establishing ethics boards and codes of conduct, and by adopting industry-wide standards and internationally recognised quality seals.

3.4 AI systems with an important social impact could require independent reviewing and testing on a periodic basis.

3.3. In the development and monitoring of AI systems, particular attention should be paid to disadvantaged groups which may be incorrectly represented in the training data.

4 A comprehensive approach to fairness

4.1 AI systems can perpetuate and exacerbate bias, and have a broad social and economic impact in society. Addressing fairness in AI use requires a holistic approach. In particular, it requires:

i. the close engagement of technical experts from AI-related fields with statisticians and researchers from the social sciences; and

ii. a combined engagement between governments, organisations that develop, deploy or use AI systems and the public at large.

4.2 The Fairness and Non-Discrimination Principle is supported by the Transparency and Accountability Principles. Effective fairness in use of AI systems requires the implementation of measures in connection with both these Principles.

Principle 5
Safety and Reliability

Organisations that develop, deploy or use AI systems and any national laws that regulate such use shall adopt design regimes and standards ensuring high safety and reliability of AI systems on one hand while limiting the exposure of developers and deployers on the other hand.

1 Require and/or define explicit ethical and moral principles underpinning the AI system

1.1 Governments and organisations developing, deploying or using AI systems should define the relevant set of ethical and moral principles underpinning the AI system to be developed, deployed or used taking into account all relevant circumstances. A system designed to autonomously make decisions will only be acceptable if it operates on the basis of clearly defined principles and within boundaries limiting its decision making powers.

1.2 Governments and organisations developing, deploying or using AI systems should validate the underpinning ethical and moral principles as defined periodically to ensure on-going accurateness.

2 Standardisation of behaviour

2.1 Governments and organisations developing, deploying or using AI systems should recall that ethical and moral principles are not globally uniform but may be impacted e.g., by geographical, religious or social considerations and traditions. To be accepted, AI systems might have to be adjustable in order to meet the local standards in which they will be used.

2.2 Consider whether all possible occurrences should be pre-decided in a way to ensure the consistent behaviour of the AI system, the impact of this on the aggregation of consequences and the moral appropriateness of "weighing the unweighable" such as life vs. life.

3 Ensuring safety, reliability and trust

3.1 Governments should require and organisations should test AI systems thoroughly to ensure that they reliably adhere, in operation, to the underpinning ethical and moral principles and have been trained with data which are curated and are as 'error-free' as practicable, given the circumstances.

3.2 Governments are encouraged to adjust regulatory regimes and/or promote industry self-regulatory regimes for allowing market-entry of AI systems in order to reasonably reflect the positive exposure that may result from the public operation of such AI systems. Special regimes for intermediary and limited admissions to enable testing and refining of the operation of the AI system can help to expedite the completion of the AI system and improve its safety and reliability.

3.3 In order to ensure and maintain public trust in final human control, governments should consider implementing rules that ensure comprehensive and transparent investigation of such adverse and unanticipated outcomes of AI systems that have occurred through their usage, in particular if these outcomes have lethal or injurious consequences for the humans using such systems. Such investigations should

be used for considering adjusting the regulatory framework for AI systems in particular to develop a more rounded understanding of how such systems should gracefully handover to their human operators.

4 Facilitating technological progress at reasonable risks

4.1 Governments are encouraged to consider whether existing legal frameworks such as product liability require adjustment in light of the unique characteristics of AI systems.

4.2 Governments should support and participate in international co-ordination (through bodies such as the International Organisation for Standardisation (ISO) and the International Electrotechnical Commission (IEC)) to develop international standards for the development and deployment of safe and reliable AI systems.

Principle 6
Open Data and Fair Competition

Organisations that develop, deploy or use AI systems and any national laws that regulate such use shall promote (a) open access to datasets which could be used in the development of AI systems and (b) open source frameworks and software for AI systems. AI systems must be developed and deployed on a "compliance by design" basis in relation to competition/antitrust law.

1 Supporting effective competition in relation to AI systems

1.1 Governments should support and participate in international co-ordination (through bodies such as the OECD and the International Competition Network) to develop best practices and rigorous analysis in understanding the competitive impact of dataset control and AI systems on economic markets.

1.2 Governments should undertake regular reviews to ensure that competition law frameworks and the enforcement tools available to the relevant enforcement authorities are sufficient and effective to ensure sufficient access to necessary inputs, and adequate choice, vibrant rivalry, creative innovation and high quality of output in the development and deployment of AI systems, to the ultimate benefit of consumers.

2 Open data

2.1 Governments should foster and facilitate national infrastructures necessary to promote open access to datasets to all elements of society having a vested interest in access to such datasets for research and/or non-commercial use. In this regard, governments should give serious consideration to two-tier access models which would allow for free access for academic and research purposes, and paid-for access for commercialised purposes.

2.2 Governments should support open data initiatives in the public or private sector with guidance and research to share wide understanding of the advantages to be gained from open access data, the structures through which datasets can be shared and exchanged, and the processes by which data can be made suitable for open access (including API standardisation, pseudonymisation, aggregation or other curation, where necessary).

2.3 Governments should ensure that the data held by public sector bodies are accessible and open, where possible and where this does not conflict with a public sector mandate to recover taxpayer investment in the collection and curation of such data. Private sector bodies such as industry organisations and trade associations should similarly support and promote open data within their industry sector, making their own datasets open, where possible.

2.4 Organisations that develop, deploy or use AI systems are encouraged to open up access to, and/or license, their datasets, where possible via chaperoned mechanisms such as Data Trusts.

2.5 Any sharing or licensing of data should be to an extent which is reasonable in the circumstances and should be in compliance with legal, regulatory, contractual and any other obligations or requirements in relation to the data concerned (including privacy, security, freedom of information and other confidentiality considerations).

3 Open source AI systems

3.1 Organisations that develop AI systems are normally entitled to commercialise such systems as they wish. However, governments should at a minimum advocate accessibility through open source or other similar licensing arrangements to those innovative AI systems which may be of particular societal benefit or advance the "state of the art" in the field via, for example, targeted incentive schemes.

3.2 Organisations that elect not to release their AI systems as open source software are encouraged nevertheless to license the System on a commercial basis.

3.3 To the extent that an AI system can be subdivided into various constituent parts with general utility and application in other AI use-cases, organisations that elect not to license the AI system as a whole (whether on an open source or commercial basis) are encouraged to license as many of such re-usable components as is possible.

4 Compliance by design with competition/antitrust laws

4.1 Organisations that develop, deploy or use AI systems should design, develop and deploy AI systems in a "compliance by design" manner which ensures consistency with the overarching ethos of subsisting competition/antitrust regimes to promote free and vibrant competition amongst corporate enterprises to the ultimate benefit of consumers.

Principle 7
Privacy

Organisations that develop, deploy or use AI systems and any national laws that regulate such use shall endeavour to ensure that AI systems are compliant with privacy norms and regulations, taking into account the unique characteristics of AI systems, and the evolution of standards on privacy.

1 Finding a balance

1.1 There is an inherent and developing conflict between the increasing use of AI systems to manage private data, especially personal data; and the increasing regulatory protection afforded internationally to personal and other private data.

1.2 Governments that regulate the privacy implications of AI systems should do so in a manner that acknowledges the specific characteristics of AI and that does not unduly stifle AI innovation.

1.3 Organisations that develop, deploy and use AI systems should analyse their current processes to identify whether they need be updated or amended in any way to ensure that the respect for privacy is a central consideration.

2 The operational challenges ahead for AI users

2.1 AI systems create challenges specifically in relation to the practicalities of meeting of requirements under a number of national legislative regimes, such as in relation to consent and anonymization of data. Accordingly, organisations that develop, deploy or use AI systems and any national laws that regulate such use shall make provision for alternative lawful bases for the collection and processing of personal data by AI systems.

2.2 Organisations that develop, deploy or use AI systems should consider implementing operational safeguards to protect privacy such as privacy by design principles that are specifically tailored to the specific features of deployed AI systems.

2.3 Organisations that develop, deploy and use AI systems should appoint an AI Ethics Officer, in a role similar to Data Protection Officers under the GDPR, but with specific remit to consider the ethics and regulatory compliance of their use of AI.

3 AI as a tool to support privacy

3.1 Although there are challenges from a privacy perspective from the use of AI, in turn the advent of AI technologies could also be used to help organisations comply with privacy obligations.

Principle 8
AI and Intellectual Property

Organisations that develop, deploy or use AI systems should take necessary steps to protect the rights in the resulting works through appropriate and directed application of existing intellectual property rights laws. Governments should investigate how AI-authored works may be further protected, without seeking to create any new IP right at this stage.

1 Supporting incentivisation and protection for innovation

1.1 Innovation is only of value if it can benefit society. Funding is necessary to develop innovation to a level where it can be disseminated and utilised by society. Those from whom funding is sought require a return on their investment. Consequently, there must be incentivisation and protection for innovation if it is to attract investment and be brought to the greater good of society.

1.2 Organisations must therefore be allowed to protect rights in works resulting from the use of AI, whether AI-created works or AI enabled works.

1.3 However, care needs to be taken not to take steps which will amount to overprotection, as this could prove detrimental to the ultimate goal of IP protection.

2 Protection of IP rights

2.1 At present IP laws are insufficiently equipped to deal with the creation of works by autonomous AI.

2.2 Organisations that develop, deploy or use AI systems should take necessary steps to protect the rights in the resulting works. Where appropriate these steps should include asserting or obtaining copyrights, obtaining patents, when applicable, and seeking contractual provisions to allow for protection as trade secrets and/or to allocate the rights appropriately between the parties.

3 Development of new IP laws

3.1 Governments should be cautious with revising existing IP laws.

3.2 Governments should explore the introduction of appropriate legislation (or the interpretation of existing laws) to clarify IP protection of AI enabled as well as AI created works, without seeking to create any new IP right at this stage.

3.3 When amending existing or implementing new IP laws, governments should seek adequately to balance the interests of all relevant stakeholders.

3.4 Governments should also explore a consensus in relation to AI and IP rights to allow for unhindered data flows across borders and the rapid dissemination of new technologies and seek to address these issues through an international treaty.