DECEMBER 2020

THE
FUTURE
SOCIETY

AREAS FOR FUTURE ACTION IN THE RESPONSIBLE AI ECOSYSTEM

PRODUCED BY:

The Future Society

IN COLLABORATION WITH:

Global Partnership on AI (GPAI) Responsible Development, Use and Governance of AI Working Group
&
International Centre of Expertise in Montreal for the Advancement of Artificial Intelligence (CEIMIA)

Acknowledgments

This report truly is a product of collective intelligence. It has been produced by **The Future Society,** with the input and collaboration of the wider **GPAI Responsible Development, Use and Governance of AI Working Group ("Responsible AI Working Group" or "RAIWG")**, in a period of two months leading to the first Global Partnership for AI (GPAI) plenary. Transparency and open collaboration have been the underlying values behind a process which included three working group meetings, four steering committee meetings, three targeted feedback sessions, seven interviews, one questionnaire, and over 400 comments on interactive documents online.

GPAI Responsible AI Co-Chairs
Yoshua Bengio
Raja Chatila

GPAI Responsible AI Steering Committee
Dyan Gibbens
Kate Hannah
Toshiya Jitsuzumi
Vincent Müller
Wanda Muñoz
Dino Pedreschi
Catherine Régis
Francesca Rossi

GPAI Responsible AI Members
Carolina Aguerre
Genevieve Bell
Ivan Bratko
Joanna Bryson
Partha Pratim Chakrabarti
Jack Clark
Virginia Dignum
Alistair Knott
Pushmeet Kohli
Marta Kwiatkowska
Christian Lemaître Léon
Alice Hae Yun Oh
Luka Omladič
Julie Owono
V K Rajah
David Sadek
Rajeev Sangal
Matthias Spielkamp
Osamu Sudo
Roger Taylor

GPAI Responsible AI Observers
Amir Banifatemi
Vilas Dhar
Marc-Antoine Dilhac
Adam Murray
Karine Perset
Stuart Russell
Cédric Wachholz

CEIMIA Secretariat
Mathieu Marcotte
Christophe Mariage-Beaulieu
Jacques Rajotte
Réjean Roy

The Future Society Project Team
Niki Iliadis, Project Manager
Nicolas Miailhe, Research Director
Nicolas Moës, Analyst
Sacha Alanoca, Analyst
Delfina Belli, Analyst
Caroline Galvan, Advisor
Suzanne Bays, Advisor
Sarah Cairns-Smith, Advisor

Table of contents

Acknowledgments 2

Table of contents 3

Foreword 5

Executive Summary 6

Introduction 11

Section 1. Responsible AI Landscape 12

1.1 Context and Objective of the Catalogue 12

1.2 The Catalogue 13

1.2.1 Categories 13

1.2.2 Attributes 15

1.2.3 Catalogue Insights 16

1.3 Methodology and Limitations 17

Section 2. A Sample of Promising Initiatives 18

2.1 Common Assessment Framework 18

2.1.1 Objective of the Assessment Framework 18

2.1.2 Methodology 18

2.2 Shortlisted Initiatives 22

Section 3. The Road Ahead 33

3.1 Area for Future Action 1: Prioritize resources towards the most pressing global issues 33

3.1.1 Challenge 33

3.1.2 Recommendations 35

3.2 Area for Future Action 2: Ensure initiatives are designed for impact 38

3.2.1 Challenge 38

3.2.2 Recommendations 39

3.3 Area for Future Action 3: Strengthen the ecosystem to accelerate change 40

3.3.1 Challenge 40

3.3.2 Recommendations 41

3.4 Area for Future Action 4: Respect and champion diversity and inclusion 44

3.4.1 Challenge 44

3.4.2 Recommendations 45

Conclusion 47

Appendix 48
Appendix 1: Catalogue 48
Appendix 2: Criteria Assessment Framework 48
Appendix 3: Analysis of 30 Shortlisted Initiaitives 48
1. AI Commons 48
2. AI Explainability 360 50
3. AI for SDGs Think Tank 52
4. AI for Good 53
5. AI-Based Referral System 54
6. AI Now Report 55
7. Algorithm Charter for Aotearoa New Zealand 57
8. Artificial Intelligence Against Modern Slavery (AIMS) 59
9. Artificial Intelligence and Blockchain for Healthcare Initiative in Africa 60
10. Artificial Intelligence Standardization White Paper 61
11. Asilomar AI Principles 62
12. Assessment List for Trustworthy Artificial Intelligence (ALTAI) 64
13. CDEI Review of Online Targeting 65
14. CEPEJ Ethical Charter on the Use of AI in Judicial Systems and their Environment 67
15. Draft AI R&D Guidelines for International Discussions 68
16. Elements of AI 70
17. Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) 72
18. Global Governance of AI Roundtable 73
19. HumanE AI Net 75
20. iGamma 76
21. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 77
22. ISO/IEC JTC 1/SC 42 79
23. Machine Learning Quality Management Guidelines 80
24. Montreal Declaration: Responsible AI 81
25. Observatory from the fAIr LAC Initiative 83
26. OECD Recommendation of the Council on Artificial Intelligence 85
27. Open Kinyarwanda 87
28. Partnership on AI Issue Area on Safety-Critical AI (SCAI) 88
29. UNESCO Recommendation on the Ethics of Artificial Intelligence & AI Decision Makers' Toolkit 89
30. UNICEF AI for Children 90

Foreword

The Global Partnership on AI (GPAI) was founded in 2020 to undertake applied AI projects and provide a mechanism for sharing multidisciplinary analysis, foresight and coordination - with the objective of facilitating international collaboration and synergies and reducing duplication in the area of AI systems.

We have the privilege of co-chairing one of GPAI's four expert working groups, the Responsible Development, Use and Governance of AI Working Group. The working group's mandate is to foster and contribute to the responsible development, governance, and use of human-centered AI systems, while seeking to address the UN Sustainable Development Goals.

For the first GPAI's plenary (December 2020) and as a first step to achieve this mandate, we wanted to identify areas for future actions, including practical projects, to help bridge the gap between responsible AI principles and implementation.

Given the plethora of initiatives around the world led by a multitude of stakeholders (academia, governments, private sector, civil society and international organizations) and the limited time available to analyze them, this was an ambitious objective. We therefore decided to mandate The Future Society (TFS) to help us with this work, which will be key to anchor our working group's relevance by orienting its choices of projects, thematic priorities and deliverables.

The Future Society (TFS) acted independently from the working group, but consulted its members, as well as its Steering Committee, in the course of its mandate. We are thankful for their dedicated work. Despite limited time, TFS was able to capture over 200 diverse initiatives in the ecosystem, to develop an assessment framework allowing the analysis of a subset of promising initiatives that have potential to contribute to the internationally coordinated development, governance and use of beneficial AI systems and applications, and to identify opportunities for future action and collaboration. The ecosystem benefits today from a report that offers a thoughtful catalogue of key initiatives, an analysis of the most promising avenues as well as recommendations for intervention.

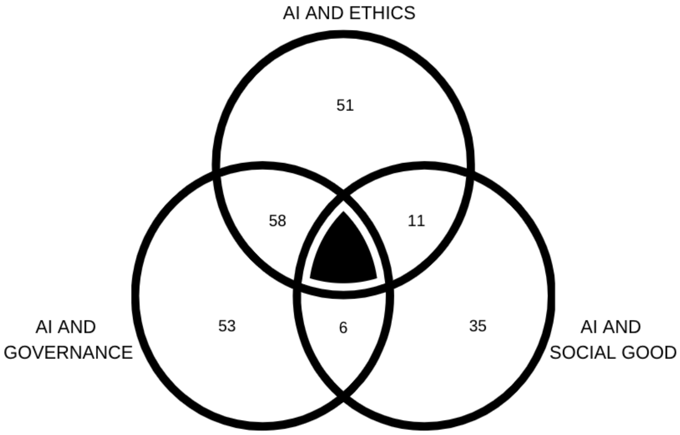
Our work is grounded in a vision of AI that is human-centered, fair, respectful of human rights and democracy, aiming at contributing positively to public good and that is equitable and inclusive. This report will undoubtedly help us translate this vision into action and, thanks to the breadth of knowledge and dedication that our expert members colleagues bring to this working group, we hope to contribute to the development of international coordination mechanisms whereby AI systems support global efforts to implement the UNSDGs, as well as to the identification of appropriate governance frameworks.

Best regards,
Yoshua and Raja

Executive Summary

Over the past decade, Artificial Intelligence (AI) has prominently entered the public conscience and debate. As a general-purpose technology, it has unprecedented potential to advance societal well-being, economic progress and address many of the most pressing challenges of our times. Yet, it also comes with significant risks, and decisions and mechanisms that are taken and designed *today* will set the course for decades to come.

It is in this context, that the Global Partnership on AI (GPAI) was formed —to help promote and foster the responsible development and use of AI globally— in line with societal values, preferences and needs. This report, produced by The Future Society in collaboration with the GPAI Responsible Development, Use and Governance of AI Working Group ("Responsible AI Working Group" or "RAIWG"), serves as a first step in supporting GPAI's mission. In preparation for the first GPAI plenary in December 2020, it aims to provide a high-level review of the landscape, an analysis of opportunities and gaps by delving into a subset of diverse initiatives, and recommendations that may steer the future agenda of the GPAI Responsible Development, Use and Governance of AI Working Group.



Recent years have seen the emergence of a plethora of initiatives that seek to define universal principles of what responsible AI constitutes, conceive of mechanisms to govern its responsible use, or deploy its potential to advance the agenda on social good. Yet, it is exactly the multitude of and parallel efforts that also make it difficult to maintain a high-level overview of opportunities, risks and shortcomings. **Section 1. Responsible AI Landscape** captures an overview of these diverse initiatives across geographies, sectors and actors. It includes a total of 214 initiatives clustered across three categories:

- **AI and Ethics:** Ethical frameworks and guidelines promoting & fostering Responsible AI
- **AI and Governance:** Governance mechanisms operationalizing Responsible AI
- **AI and Social Good:** Applied projects advancing SDGs responsibly

Through a common assessment framework designed specifically for this report, **Section 2. A Sample of Promising Initiatives** analyzes a subset of these initiatives (a total of 30) that are particularly relevant for GPAI. They are analyzed on their potential to help GPAI deliver on its objectives; their effectiveness and alignment with the OECD AI Principles and the UN SDG agenda; their scalability across geographies and sectors; and whether they are as representative as possible across geographies, sectors, stakeholders and target groups.

Section 3. The Road Ahead draws on the opportunities and gaps identified in the previous sections to propose four areas for future action and nine recommendations that will help inform GPAI's agenda going forward. An overview of them is provided below.

<table>
<tr><th colspan="2">Area for Future Action 1</th></tr>
<tr><td colspan="2">Prioritize resources towards the most pressing global issues.</td></tr>
<tr><th colspan="2">Challenge</th></tr>
<tr><td colspan="2">AI has wide-ranging applicability and hence has the potential to influence many of the most pressing issues humanity is facing: it can be a force for good to mitigate climate change or predict the next pandemic, and it can also exacerbate global challenges as evidenced by the rise of misinformation. The breadth of potential applications of Responsible AI creates a prioritization challenge.</td></tr>
<tr><th>Recommendation</th><th>Explanation</th></tr>
<tr><td><b>1. Build a systematic process to ensure efforts are targeted at the most pressing global issues.</b></td><td>Using the SDG framework as a critical prioritization tool, GPAI Responsible AI should identify pressing issues and channel efforts to where they are most effective in a given context. Four main factors should be considered when identifying areas: i) impact; ii) urgency; iii) feasibility; and iv) relevance (Table 7).</td></tr>
<tr><td><b>2. Create focused committees to address identified pressing issues.</b></td><td>GPAI Responsible AI should form focused committees to provide advice and recommendations to GPAI governments that reflect the interdisciplinary scientific consensus related in a pressing issue. Initial pressing issues could be:
i) Committee on Governance and Transparency of Social Media
ii) Committee on AI and Education
iii) Committee on Drug Discovery and Open Science
iv) Committee on Climate Change and Biodiversity</td></tr>
</table>

Area for Future Action 2	
Ensure initiatives are designed for impact.	
Challenge	
Many initiatives within the Responsible AI ecosystem have unclear metrics for tracking progress. This makes it challenging to formulate standardized impact definitions. Beyond measurement, many initiatives also lack clear impact pathways; thus, making it difficult to evaluate performance, specifically when it comes to advancing progress towards the UN SDGs.	
Recommendation	Explanation
3. Develop a common taxonomy and international measurement system among GPAI governments.	GPAI Responsible AI should champion and initiate an international agreement defining Responsible AI with a concrete, efficient and effective measurement system. This would be accompanied by an evidence-based and agreed-upon taxonomy of concepts pertaining to AI itself, how responsible it is, and its impact. Each focused committee could develop performance benchmarks that permit consistent assessment of AI system capabilities globally.
4. Construct a widely applicable and coherent impact assessment methodology.	GPAI Responsible AI should develop an impact assessment methodology aiming at the operationalization of its taxonomy, concrete auditing and evaluation mechanisms, and propose a path of how these guidelines could be standardized across governments. The methodology should address two fundamental aspects of Responsible AI: Governance and AI for Social Good.

<table>
<tr><th colspan="2">Area for Future Action 3</th></tr>
<tr><td colspan="2">Strengthen the ecosystem to accelerate change.</td></tr>
<tr><th colspan="2">Challenge</th></tr>
<tr><td colspan="2">The cultivation of a strong ecosystem with the ability to support and stimulate change is a third challenge. To build this ecosystem, there is a need for governance tools and frameworks that promote transparency and alter incentives and behaviors throughout society to help the adoption of Responsible AI practices. There is also a need for systematic collaboration and cooperation across the ecosystem as well as a mechanism to connect cross-cutting initiatives on the domain level. Finally, for governments to implement these tools and frameworks at scale, there is a need to build capacity amongst policymakers as well as feedback loops between governments and other actors in the ecosystem.</td></tr>
<tr><th>Recommendation</th><th>Explanation</th></tr>
<tr><td><b>5. Create a focused committee on governance issues and governance means.</b></td><td>GPAI Responsible AI should form a cross-cutting committee focusing on key governance issues. Amongst others, this committee’s aims should be to ensure AI systems are designed and used by organizations in an accountable and transparent manner to ensure fairness, safety, robustness, respect for human rights and the promotion of equity. A particular focus should be on AI in high-stake decision making.</td></tr>
<tr><td><b>6. Facilitate coordination within the ecosystem.</b></td><td>GPAI Responsible AI should set up a coordination mechanism to facilitate communication across initiatives, thereby enabling initiatives to leverage each other’s learnings and good practices. Focus should be on supporting those initiatives that are already mindful of increasing coordination and minimizing duplication. In areas identified as pressing, the creation of Public Private People Partnerships across geographies should be initiated to help address the key issues related to those priority areas while at the same time testing the governance tools in deployment.</td></tr>
<tr><td><b>7. Build capacity for policymakers to govern Responsible AI.</b></td><td>GPAI Responsible AI should assist governments in building capacity for the governance of Responsible AI. This could include capacity to engage in international standardization processes, to link accountability tools with metrics and taxonomy, to make fiscal incentives conditional on specific performance, to organize dialogue and coordination among various AI stakeholders, and to better assess and deploy relevant tools to govern AI.</td></tr>
</table>

<table>
<tr><th colspan="2">Area for Future Action 4</th></tr>
<tr><td colspan="2">Respect and champion diversity and inclusion.</td></tr>
<tr><th colspan="2">Challenge</th></tr>
<tr><td colspan="2">Many initiatives in the Responsible AI ecosystem have struggled to collect representative input to inform their activities. This lack of inclusiveness points to a lack of capacity by initiatives, stakeholders and governments to involve a wider group in the technological transition and, hence, to co-shape innovative solutions for addressing the opportunities and the risks. Ultimately, this lack of inclusiveness risks undermining the effectiveness and credibility of many Responsible AI initiatives as well as their ability to scale.</td></tr>
<tr><th>Recommendation</th><th>Explanation</th></tr>
<tr><td><b>8. Develop and disseminate good Diversity & Inclusion (D&I) practices.</b></td><td>GPAI Responsible AI should help shape and spread good D&I practices, including a strategy that helps gauge the extent to which segments of society or geographies are currently underrepresented or excluded in the Responsible AI ecosystem. Furthermore, additional steps could be to encourage open-access information and infrastructure, that is widely available in accessible and user-friendly manner, and to break down communications barriers between geographies, social groups, and disciplines.</td></tr>
<tr><td><b>9. Initiate strategic partnerships with platforms collecting representative input.</b></td><td>GPAI should work with international organizations like OECD, WHO and UNESCO to collect representative input from marginalized groups and the Global South. GPAI's role in these partnerships can be to proactively bring historically marginalized groups into these dialogues and to support initiatives that foster basic AI literacy so the public can be empowered to participate.</td></tr>
</table>

Introduction

In the past decade, Artificial Intelligence (AI) has captured worldwide attention. Its realized and potential impact on the thriving of nations, on the wellbeing of individuals, and on addressing today's global challenges has placed it high on both national and global agendas alike. The key question stakeholders worldwide are trying to answer is what are the right pathways to capture AI's positive impacts and mitigate its negative ones.

The Global Partnership on AI (GPAI) was created in this context, as an international and multistakeholder initiative to undertake applied AI projects and provide a mechanism for sharing multidisciplinary analysis, foresight and coordination - with the objective of facilitating international collaboration and synergies and reducing duplication in the area of AI systems. The initiative was launched in June 2020 by Canada and France - along with Australia, the European Union, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom and the United States of America.

Following its launch, GPAI brought together experts from diverse sectors into four working groups: Data Governance; Responsible Development, Use and Governance of AI; Future of Work; and Commercialization and Innovation. These working groups have been given the same task: to help advance GPAI's mission "to support the development and use of AI based on human rights, inclusion, diversity, innovation, and economic growth, while seeking to address the United Nations Sustainable Development Goals."[2]

This report, produced by The Future Society in collaboration with the GPAI Responsible Development, Use and Governance of AI Working Group ("Responsible AI Working Group" or "RAIWG"), is a first step towards achieving GPAI's mission. It provides a high-level review of the Responsible AI landscape, a deeper analysis of a subset of promising initiatives, and a proposal of areas for future action and recommendations to inform the agenda of the Responsible AI Working Group. The overall aim of these areas for future action is to ensure that all AI is developed and operated in a responsible manner. As such, they also aim to illustrate pathways to facilitate cross-sectoral and international collaborations that ensure applications of AI are used to advance the Social Good agenda, as aligned with the UN Sustainable Development Goals (SDGs).

The report is structured in three sections. The first section reviews the landscape by presenting a catalogue of diverse Responsible AI initiatives worldwide. The second section offers an analysis of a sample of these initiatives and provides a basis to understand the current gaps and opportunities in the Responsible AI ecosystem. The third section builds on both the landscape review and the analysis to plan the road ahead: identifying four areas for future action and nine concrete recommendations for GPAI.

Section 1. Responsible AI Landscape

1.1 Context and Objective of the Catalogue

The field of Responsible AI as defined by the working group's mandate is vast. The risks and opportunities of AI have resulted in an ecosystem filled with a plethora of initiatives that seek to provide guidance on how AI is to be developed and adopted, or how to use it to advance the AI for Social Good agenda.

In order to identify potential areas for future action within this ecosystem, this report seeks an understanding of what this landscape looks like as well as its opportunities and gaps. Multiple inventories and landscape analyses have already been published that hint at the size and scale of the ecosystem.[1] [2] [3] [4] [5] [6]

This report leverages these inventories and landscape analyses as starting points for building its own catalogue. The aim of the report's catalogue is to list diverse initiatives by academia, the public sector, the private sector, civil society, and international organizations that promote the responsible research and development of AI systems and its applications for Social Good. Although it is not meant to be comprehensive, the catalogue illustrates the diversity and number of existing initiatives in the ecosystem —each with their own approaches, societal objectives, concerns, achievements, hopes and ideas. By taking stock of the collective efforts by society, it shows the importance of the challenges and opportunities brought about by AI.

Furthermore, the catalogue serves as the basis to select the promising initiatives identified in Section 2 and, consequently, to better understand opportunities and gaps in the ecosystem by comparing initiatives, best practices and lessons learned.

[1] Jobin, A., Ienca, M. & Vayena, E. (2019) "Artificial Intelligence: the global landscape of ethics guidelines" Nature Machine Intelligence 1, 389-399

[2] Zeng, Y., Lu, E. & Huangfu, C. (2019) "Linking Artificial Intelligence Principles" In the Proceedings of the AAAI Workshop on Artificial Intelligence Safety

[3] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar M. (2020) "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI" Berkman Klein Center Research Publication No. 2020-1

[4] Hagendorff, T. (2020) "The Ethics of AI Ethics: An Evaluation of Guidelines" Minds and Machine 30, 99-120

[5] Research Center for AI Ethics and Sustainable Development & China-UK Research Centre for AI Ethics and Governance. (2020) "Project under specific SDGs topics", retrieved from http://www.ai-for-sdgs.academy/topics

[6] Rotenberg, M. (2020) "The AI Policy Sourcebook 2020", Electronic Privacy Information Center

1.2 The Catalogue

The catalogue can be found at the following address: https://docs.google.com/spreadsheets/d/1utA1ug1nKF3B4MMmC_AQeyo1IK5jH_abMK9oFHEJekk/edit#gid=8672956

Each row in the catalogue corresponds to an initiative, as defined by the Institutional Engineering approach[7] [8] [9] [10]. In summary, an initiative for this purpose refers to an attempt to create formal or informal mechanisms affecting the behavior of some individuals in a deliberate way within a given context and community. This can, for example, be a taskforce, a set of guidelines, a movement, a research agenda, a piece of legislation, a framework, an alliance of stakeholders, a regular conference, a tool, or an organization.

Each column of the catalogue corresponds to an attribute of the initiative or to its category, as explained below.

1.2.1 Categories

The catalogue separates the landscape of Responsible AI initiatives into three categories:

- **AI and Ethics:** Ethical frameworks and guidelines promoting & fostering Responsible AI.
- **AI and Governance:** Governance mechanisms operationalizing Responsible AI, including auditing mechanisms, risk assessments, standards, certifications, corporate governance frameworks, etc.
- **AI and Social Good:** Applied projects advancing SDGs responsibly.

These three categories are complementary on the spectrum from principles to practice to applied projects. Category A contains normative initiatives that describe an aspired to future. Category B contains prescriptive initiatives that go beyond descriptions of the end goal, and that instead indicate concrete tools or processes to be implemented to reach this future. Category C contains initiatives that aim to implement these tools and processes for advancing the Social Good agenda.

Table 1 further illustrates this categorization. Note that many initiatives have output related to more than one category, in which case it falls under multiple categories, as illustrated in Figure 1.

[7] Aligica, Paul Dragos, & Boetke, Peter J., 2009, *Challenging institutional analysis and development - The Bloomington School,* Routledge.

[8] Ostrom, Elinor, 1986, *An Agenda for the Study of Institutions*, Public Choice, Vol 48, No 1, pp. 3-25

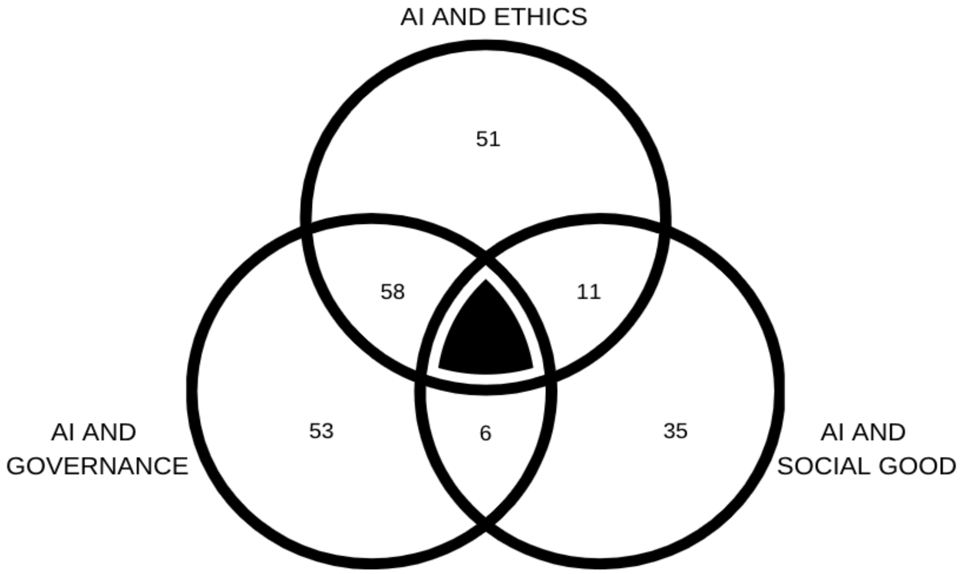
[9] Ostrom, Elinor, 2011, *Background on the Institutional Analysis and Development Framework,* The Policy Studies Journal, Vol. 19, Issue 1

[10] Crawford, Sue E. S., & Ostrom, Elinor, 1995, *A Grammar of Institutions,* The American Political Science Review, Vol 89, Issue 3, pp. 582-600

Table 1: Overview of Responsible AI categories

Category A: AI and Ethics	**Category B: AI and Governance**	**Category C: AI and Social Good**
Description		
Ethical frameworks and guidelines promoting & fostering Responsible AI	Governance mechanisms operationalizing Responsible AI	Applied projects advancing SDGs responsibly
Inclusion criterion		
Does the initiative describe the future we aspire in the context of responsible AI?	Does the initiative indicate how to reach the future we aspire to in the context of Responsible AI?	Does the initiative aim to implement what is indicated we should do to reach the future we aspire in the context of responsible AI; and does it explicitly aim at promoting social good as defined by UN SDGs?
Generic examples		
Ethical guidelines Codes of conduct Macro frameworks	Risk assessment frameworks Certifications Corporate governance frameworks Auditing mechanisms Standards	AI for Social Good applications International platforms Repositories
Specific examples		
CEPEJ European Ethical Charter on the Use of Artificial Intelligence in Judicial systems and Their Environment (Europe)	Machine Learning Quality Management Guideline (Japan)	Open Kinyarwanda (Rwanda)
Number of initiatives (as of November 11th, 2020)		
120	117	52

Figure 1: Number of initiatives across categories



1.2.2 Attributes

For the catalogue to be useful for the ulterior analysis, specific attributes are assigned to each initiative.

- **Name:** The name to help identify the initiative online, whether it is its official name or simply a common way of referring to the initiative.
- **Link:** Access to further information on the initiative, whether it is its official webpage or relevant and thorough news coverage.
- **Organization:** The group(s) that have launched, produced, developed or undertaken the initiative.
- **Brief Description:** A brief summary of the initiative explaining what it is about and what it aims to achieve.
- **Sector:** The sector from which the initiative and its authors originate, such as academia, private sector, civil society, public sector, or international organization. Note an initiative can be mixed, meaning cross-sectoral.
- **Geographical scope:** The country or region that the initiative targets.
- **Target Audience:** The group or type of individuals that the initiative targets.
- **Stage of development:** The phase an initiative is currently in.
- **Date started:** Time when the initiative has become public —regardless of the stage of development.
- **Country/region of origin:** Area from which the initiative has started, regardless of its geographical scope.

1.2.3 Catalogue Insights

As of November 11th, 2020, 214 initiatives are listed in the catalogue. These initiatives represent 38 different countries and regions and come from 189 different organizations or authors.[11]

With regards to the three categories, 120 initiatives fall under the category of AI and Ethics, 117 under AI and Governance and 52 under AI and Social Good. The earliest initiative originated in April 2011, and the latest in September 2020. Initiatives in the AI and Social Good category are more recent (with an average starting date in February 2019) than AI and Governance initiatives (with an average starting date in November 2018), which in turn are more recent than AI and Ethics initiatives (with an average starting date in April 2018). This trend illustrates how the ecosystem as a whole has shifted its focus over time from principles to practice to applied projects.

Figure 2: Geographic Distribution of Initiatives in the Catalogue

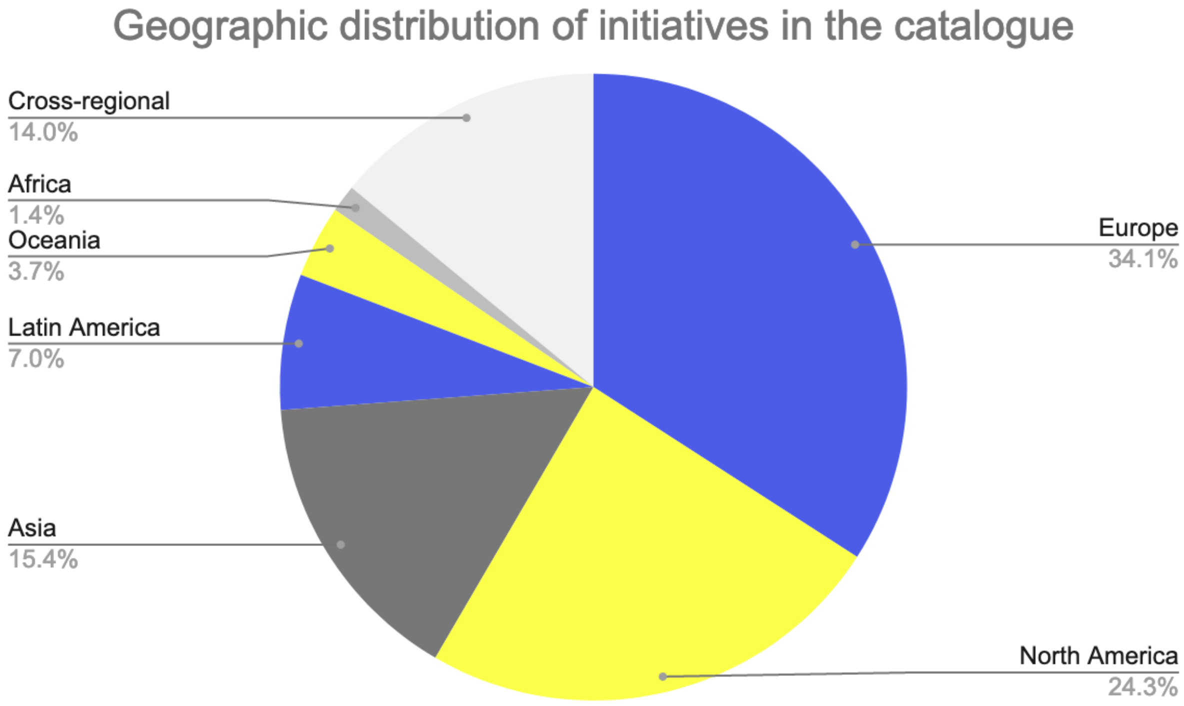


Figure 2 illustrates the distribution of initiatives by region of origin. Over 58% of the initiatives in the catalogue are from Europe and North America. Only 1.4% of initiatives are from Africa. Furthermore, the catalogue reveals that initiatives in emerging and developing economies (excl. China) overwhelmingly focus on AI and Social Good (14 out of a total of 19 initiatives, that is 74% compared to otherwise 24% for the entire population of the catalogue). On the flipside, only five out of 179 initiatives (or 3%) under AI and Ethics and AI and Governance originate from emerging

[11] Note that as the catalogue continues to be regularly updated, these figures evolve.

and developing economies.[12] This could hinder adoption of responsible AI for Social Good in these economies.

1.3 Methodology and Limitations

The underlying methodology for cataloguing was designed to capture a diverse list of initiatives at a specific point in time that favored breadth over depth throughout, with the rationale to use it as a basis to shortlist the most promising initiatives. This was to ensure that the catalogue captures the diversity of initiatives across geographies, sectors and scope whilst equally encompassing both well-known international initiatives and national ones with niche topics.

As a starting point, the project team leveraged existing public inventories of initiatives related to AI (Jobin et al., 2019; Zeng et al., 2019; Fjeld et al., 2020).[13][14][15] The project team then leveraged the Responsible AI Working Group background materials and past work of The Future Society that mapped the AI ecosystem. This initial raw sample was further supplemented with dedicated research for relevant initiatives in countries that were heavily underrepresented at that stage of the process. Finally, members of the Responsible AI Working Group were asked to submit additional initiatives that had not yet been included. Through desktop research (using news coverage, official websites and existing literature), each initiative was assigned attributes and bucketed in one or more of the three aforementioned categories.

The catalogue is as of yet not a comprehensive list of Responsible AI initiatives worldwide. Rather it served as a basis to arrive at a set of shortlisted initiatives as mentioned above. As such, it has three key limitations. First, the profiles of team and working group members alike are likely to have influenced which regions are currently underrepresented in the catalogue (e.g. Middle East, North Africa, sub-Saharan Africa, Asia and Latin America and the Caribbean). Second, the languages spoken by team members are likely to have influenced the analysis of those initiatives that were not in Chinese, English, French, German, Greek, Italian, Japanese or Spanish. Notably, significant gaps remain for African, Arabic, Central and Eastern European, and South-East Asian initiatives. Third, a tight deadline and project milestones allowed only for a brief few days to update the catalogue before the initiatives passed the common assessment framework. The set of shortlisted initiatives has, however, the potential to grow through crowdsourcing additional initiatives worldwide going forward.

[12] Figures as of November 11th 2020

[13] Jobin, A., Ienca, M. & Vayena, E. (2019) "Artificial Intelligence: the global landscape of ethics guidelines" Nature Machine Intelligence 1, 389-399

[14] Zeng, Y., Lu, E. & Huangfu, C. (2019) "Linking Artificial Intelligence Principles" In the Proceedings of the AAAI Workshop on Artificial Intelligence Safety

[15] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar M. (2020) "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI" Berkman Klein Center Research Publication No. 2020-1

Section 2. A Sample of Promising Initiatives

2.1 Common Assessment Framework

2.1.1 Objective of the Assessment Framework

While the catalogue provides a mapping of the current AI landscape, the assessment criteria and framework aim to identify a sample or subset of initiatives that have the potential to contribute to the development and use of AI systems, and applications that could benefit from international and cross-sectoral collaboration. Specifically, the aim of the assessment framework is to identify areas for future actions and collaborations, including tangible projects, in order to bridge the gap between responsible AI principles and implementation.

The sample of initiatives must therefore fulfill five objectives: They must represent the breadth of Responsible AI initiatives that could contribute to GPAI's objectives; contribute effectively to the advancement of Social Good (as described by the SDGs); have potential for greater adoption; provide geographical representation; and be inclusive each on their own.

2.1.2 Methodology

The assessment framework was designed to fulfill these five objectives as well as two further functions. First, it develops an evaluation funnel to identify the most promising initiatives by analyzing their attributes, and to pre-select a diverse and inclusive subset among the most promising ones in terms of their potential to contribute to GPAI's work. Second, it supports an analytical approach to extract more information (lessons learned, good practices, challenges for implementing ethical AI principles).

Evaluation funnel

For the first stage of the assessment, 214 initiatives were considered through a two-stage waterfall process: preliminary and granular evaluation.[16] As shown in Table 2, the process leverages four assessment criteria across both phases. Each criterion is a composite of concrete and comparable component indicators, as explained in Table 2. There are 23 indicators in total. These indicators have been selected to ensure breadth of coverage (i.e. easy enough to measure for all or most initiatives) without sacrificing informativeness.

[16] These were the 207 initiatives present on the catalogue by October 22nd 2020

Table 2: Assessment Criteria Framework

Criteria	Definition	Indicators
Potential to contribute to GPAI objectives	Degree to which the initiative aligns with the GPAI mandate ("foster and contribute to the responsible development, governance, and use of human-centered AI systems, in congruence with the UN Sustainable Development Goals.") and shows potential for international and cross-sectoral cooperation and collaboration.	● Geographical scope Target audience ● Sector ● Interdisciplinary team ● Cross-sectoral and cross-regional collaboration
Diversity & Inclusiveness	Diversity of overall team profile, degree to which the initiative is accessible within different countries and population segments, and degree to which it serves marginalized and/or underserved communities and/or countries.	● Country of origin ● Category (e.g. Ethics, Governance, or Social Good) ● Profiles and composition of core team ● Inclusion of marginalized groups ● Availability to the originally targeted sectors ● Ease of access for different population segments and levels of bandwidth/connectivity ● Potential for benefiting marginalized groups or countries in the Global South
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Degree to which the initiative meets its objectives and/or the extent to which it advances the Sustainable Development Goals and aligns with OECD AI principles.	● Clarity of objectives and own metrics ● Ability to achieve its objectives (when applicable) ● Number of SDGs served by initiative ● Extent to which SDG(s) is/are served by the initiative ● Alignment with OECD AI Principles
Maturity & Potential for Adoption	Degree of maturity and adoption of the initiative, and scalability.	● Stage of development ● Level of national, regional and/or international adoption/usage ● Scalability ● Required resources for implementation ● Stakeholder buy-in ● Budget

Note: Blue font color marks indicators used for the Preliminaries onwards; Green font color marks indicators used for the Granular evaluation.

The assessment framework is based on the development aid evaluation literature.[17] The literature was adjusted to be applicable to projects that are i) heterogeneous in terms of approach and issues addressed, and ii) in the implementation phase. Thus, the selected indicators refer mostly to inputs and activities ("planned work" indicators) and outputs and outcomes (intended results).[18]

Inputs are the resources available and leveraged for the initiative (e.g. time spent by specific stakeholders, assets at disposal, strategy). The indicators that fall within this category are:

- Geographical scope
- Sector
- Interdisciplinary team
- Country of origin
- Profiles and composition of core team
- Clarity of objectives & own metrics
- Required resources for implementation
- Budget

Activities are the utilization of these resources in the creation of outputs (e.g. conferences, writing up reports). The indicators that fall within this category are:

- Target audience
- Cross-sectoral and cross-regional collaboration
- Inclusion of marginalized groups
- Ease of access for different population segments and levels of bandwidth/connectivity
- Alignment with OECD AI Principles
- Stage of development
- Stakeholder buy-in

Outputs are the proximate results of the initiative (e.g. reports published, some stakeholders implementing new governance mechanisms). The indicators that fall within this category are:

- Availability to the originally targeted sectors
- Ability to achieve its objectives (when applicable)
- Extent to which SDG(s) is/are served by the initiative
- Level of national, regional and/or international adoption/usage
- Scalability
- Number of SDGs served by initiative

Outcomes are the ultimate results (e.g. greater awareness of the issue the initiative is trying to address, more cost-effective procedure to analyze company statements on their compliance with

[17] World Bank's Independent Evaluation Group, *Designing a Results Framework for Achieving Results: A How-To Guide,* International Bank for Reconstruction and Development/World Bank Group

[18] Pp. 24-25, World Bank's Independent Evaluation Group, *Designing a Results Framework for Achieving Results: A How-To Guide,* International Bank for Reconstruction and Development/World Bank Group

measures to combat modern slavery in their supply chain). The indicators that fall within this category are:

- Category (e.g. Ethics, Governance, or Social Good)
- Potential for benefiting marginalized groups or countries in the Global South

As initiatives assessed are mostly in the implementation phase, outcomes have not yet occurred. This is why most indicators reflect inputs and activities rather than output and outcomes. A few indicators pertaining to outputs and outcomes were kept to ensure that —to the extent it is possible to do so consistently— results are factored into the assessment framework. Note that outcomes, particularly the extent to which an effort can be attributed to the improvement of an outcome variable, are notoriously difficult to assess consistently, even in the field of development aid.

Preliminary evaluation

To begin with, all 214 initiatives were scored on their “Potential to contribute to GPAI’s objectives” and their "Diversity & Inclusiveness”, leveraging the indicators in blue in Table 2. Each criterion was scored between 0 (low) and 5 (high). Initiatives that had a combined score of below 7 (out of 10) were excluded. This first filter allowed the team to focus on a more comprehensive analysis of promising initiatives.

64 initiatives passed this exclusion criterion (30% of initial sample).

Granular evaluation

The 64 initiatives proceeded to a more granular analysis, including research on team composition, challenges, success factors, and the extent to which they address the UN SDGs and OECD AI principles. The 64 initiatives were scored on their “Potential to contribute to GPAI objectives”, “Diversity & Inclusiveness”, “Effectiveness/Alignment with UN SDGs and OECD AI Principles” and “Maturity & potential for adoption”. For this, the full set of indicators listed in Table 2 were applied. Each initiative was scored against all four criteria, from 0 (low) to 5 (high), providing a combined score out of 20 for the granular evaluation. Those that scored above 15 were preserved.

35 initiatives passed this filter (16% of initial sample).

The final step was to ensure that the sample was as mutually exclusive yet collectively as exhaustive as possible. First, going from the highest to the lowest scoring initiative (or top to bottom), each initiative was compared with its adjacent ones (i.e. comparing the second highest scoring initiative with the highest scoring initiative, the third highest scoring with both the second and first highest scoring, etc.). If one initiative was too similar (in terms of country of origins, sector, category, organization, etc.) to another that scored higher, it was excluded.

As a result, 30 initiatives passed this filter (14% of initial sample).

As expected, the correlation between the preliminary and the granular scores is positive, but the scores are far from identical to each other (Pearson's r = 0.36).[19] This provides confidence in the validity of the preliminary scores, but also shows the usefulness of having collected additional information.

Analytical approach

In parallel to the evaluation funnel, a more analytical approach aimed to map good practices and lessons learned across these initiatives.

The first round of analysis was based on the attributes in the catalogue and focused on the top 64 initiatives (as explained in the section above), while additional data was collected to further evaluate and analyze them. Concretely, during the shortlisting process of the top 64, more attributes (namely: specific SDGs addressed, specific OECD AI principles addressed, signs of cross-sectoral collaboration, profiles and configuration of the team, key success factors, main challenges, resources and budget available, scalability and availability & accessibility) were identified. These efforts helped identify a first set of areas for future action as well as common threads across initiatives to inform key recommendations.

Upon completion of the evaluation funnel (i.e. shortlisting of the 30 initiatives), a questionnaire was sent to key stakeholders from each initiative to help fill information gaps about good practices, team profiles, challenges, and potential areas for future action. 25 out of the thirty initiatives provided further information to support the respective analysis.

Validation of the methodology and output

The design and implementation of this methodology was guided by the input of the GPAI Responsible AI Steering Committee.

Specifically, the process ensured built-in junctures for feedback and confirmation while enabling progress in parallel on other fronts. For example, once the first step of progress in the evaluation funnel was set up, the Steering Committee provided input while the analysis of the smaller sample continued. This enabled smooth feedback integration, evaluation and analysis in parallel throughout the project - without one activity becoming the bottleneck for the other. It also allowed for increased feedback loops with the Steering Committee.

2.2 Shortlisted Initiatives

The common assessment framework and the methodology described in the former section led to a sample of thirty promising Responsible AI initiatives. Each of the initiatives showed potential to

[19] Pearson's r measures the correlation between two variables i.e. how two dimensions change together. A value of 1 means the two dimensions are changing in perfect synchrony; a value of -1 means that they systematically change in opposite ways; a value of 0 means that changes are purely random and that you cannot predict how one dimension changes based on the other's change. In this case, an intermediary value of 0.36 means that the dimension changes overall in the same direction as the other dimension, but that there still is some asynchrony.

contribute to GPAI’s objective, positive signs of diversity and inclusion, effectiveness in regard to UN SDGs, and maturity or potential for wider adoption.

Table 3: List of 30 Shortlisted Initiatives

AI & Ethics	Both AI & Ethics and AI & Governance	AI & Governance	AI & Social Good	Both AI & Ethics and AI & Social Good
Asilomar AI Principles CEPEJ Ethical Charter on the Use of AI in Judicial Systems and their Environment Draft AI R&D Guidelines for International Discussions Montréal Declaration: Responsible AI	Algorithm Charter for Aotearoa New Zealand Artificial Intelligence Standardization White Paper Assessment List for Trustworthy AI (ALTAI) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems OECD Recommendation on AI Partnership on AI Issue Area on Safety-Critical AI (SCAI) UNESCO Recommendation on the Ethics of AI & AI Decision Makers’ Toolkit	AI Explainability 360 AI Now Report CDEI Review of Online Targeting Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) Global Governance of AI Roundtable ISO/IEC JTC 1/SC 42 Machine Learning Quality Management Guidelines	AI-Based Referral System AI Commons AI for Good AI for SDGs Think Tank Artificial Intelligence against Modern Slavery (AIMS) Artificial Intelligence and Blockchain for Healthcare Initiative in Africa Elements of AI iGamma Open Kinyarwanda Observatory from the fAIr LAC Initiative	HumanE AI Net UNICEF AI for Children

More details on the distribution of the shortlisted initiatives is available in Table 4 and Table 5.

Table 4: Shortlisted Initiatives by Category and Region

	Total per category	Cross-regional	Africa	Asia	Europe	Latin America	North America	Oceania
Total per region	**30**	**11**	**2**	**5**	**4**	**3**	**3**	**2**
AI & Social Good	**10**	2	2	1	1	3		1
AI & Governance	**7**	2		2	1		2	
Both AI & Governance and AI & Ethics	**7**	4		1	1			1
AI & Ethics	**4**	2		1			1	
Both AI and Ethics & AI and Social Good	**2**	1			1			

Table 5: Shortlisted Initiatives by Category and Sector

	Total per category	Mixed	Academia	Civil Society	International orgs	Private Sector	Public Sector
Total per sector	**30**	**10**	**5**	**3**	**7**	**1**	**4**
AI & Social Good	**10**	4	2	2	2		
AI & Governance	**7**	3	2			1	1

Both AI & Governance and AI & Ethics	**7**	2			3		2
AI & Ethics	**4**		1	1	1		1
Both AI and Ethics & AI and Social Good	**2**	1			1		

Table 6 provides a short description of each initiative, listed in alphabetical order.

Appendix 3 provides a detailed analysis of each shortlisted initiative, with information relevant to all four criteria in the criteria assessment framework. The collective analysis —an understanding of key success factors as well as challenges faced by the initiatives— surfaced opportunities and gaps for future action and collaboration identified in Section 3.

Table 6: Brief Descriptions of 30 Shortlisted Initiatives

Initiative	**Brief Description**
AI Commons	A global knowledge hub bringing diverse stakeholders together to address the world's greatest challenges using AI. Its key objectives are to identify how beneficial AI can be designed and implemented in an inclusive and distributed manner, and to create open source blueprints for global usage. The initiative strives to advocate and make possible the concept of AI being a public good.
AI Explainability 360-	IBM's AIX360 is an open-source software toolkit that explains AI models and the data they operate on. It also provides a taxonomy of explainable AI techniques and educational materials, including a web demo, glossary, and tutorials illustrating its use in application domains. AIX360 aims to bridge the gap between the AI community and society at large. For data scientist users who are not AI experts, it helps them select an appropriate technique and successfully deploy it in their domain. For policymakers, it provides education on explainable AI technology to promote appropriate regulatory actions. For AI researchers, it points out understudied areas and provides a vehicle for disseminating new techniques.
AI for SDGs Think Tank	An online global repository compiling and analyzing AI projects and proposals that impact the UN SDGs both positively and negatively. It also includes a detailed evaluation of each initiative featured. The initiative's

	mission is to 'promote the positive use of AI for Sustainable Development and investigate the negative impact of AI on sustainable development.'
AI for Good	The AI for Good Global Summit is a United Nations platform, centered around annual global summits, that foster the dialogue on the beneficial use of AI, by developing and identifying concrete projects. The series aims to bring forward AI research topics that contribute towards more global problems, through accelerating the United Nations' Sustainable Development Goals (SDGs). Close to 40 UN organizations are partners of the AI for Good Global Summit and it also bring together experts from industry, government, civil society, academia, etc. It includes the AI Repository, a catalogue of AI initiatives which accelerate progress towards the seventeen UN SDGs.
AI Based Referral System	A diabetic retinopathy screening program for early detection and treatment through convolutional neural networks, based on Mexican clinical guidelines, that will be implemented in three hospitals in Mexico - for early detection and treatment of diabetic retinopathy. Healthcare is one of the most dynamic and challenging sectors in Mexico and the LAC region. Nevertheless, the response to Diabetic Retinopathy (DR) faces three main problems: i) High prevalence of diabetes, the WHO reported that the prevalence of diabetes in Mexico is around 10.4% in 2016; ii) shortage of ophthalmologists, Mexico reports 42.5 ophthalmologists per millions of people (OPM), in contrast with other countries such as Spain with 105.5 OPM or Argentina 103.6 OPM, Brazil 67.4 OPM; and iii) lack of eye care services in primary health care. This initiative aims to address all three.
AI Now Report 2018	The AI Now 2018 Report addresses key governance issues, including i) the growing accountability gap in AI, which favors those who create and deploy these technologies at the expense of those most affected; ii) the use of AI to maximize and amplify surveillance, iii) increasing government use of automated decision systems that directly impact individuals and communities without established accountability structures; iv) unregulated and unmonitored forms of AI experimentation on human populations; and v) the limits of technological solutions to problems of fairness, bias, and discrimination. It includes AI Now's algorithmic impact assessment framework which gives public sectors more tools for critically deciding if an algorithmic system is appropriate, and for ensuring more community input and oversight.
Algorithm Charter for Aotearoa New Zealand	The Algorithm Charter is a commitment by government agencies to improve consistency, transparency, and accountability in their use of algorithms. Signatories commit to a range of actions in the areas of transparency, partnership, focus on people, data, privacy, ethics, human rights, and oversight. The Charter follows a recommendation by the Government Chief Data Steward and Chief Digital Officer, who said that the safe and effective use of operational algorithms required greater consistency across Government. It was developed through consultation with the public and forms a part of the New Zealand Government's Open

	Government Partnership action plan. The Charter draws on the Principles for the Safe and Effective Use of Data and Analytics co-designed by the Government Chief Data Steward and the Privacy Commissioner.
Artificial Intelligence Against Modern Slavery (AIMS)	Project AIMS uses AI to combat modern slavery. It creates the first AI tool for the scalable analysis of company statements on how they are eradicating slavery from their supply chains. The tool builds on the work of Walk Free, WikiRate and the Business & Human Rights Resource Centre (BHRRC) to speed up the statement review process and increase transparency for consumers and businesses.
Artificial Intelligence and Blockchain for Healthcare Initiative in Africa	An initiative accelerating drug discovery and drug development by continuously inventing and deploying AI technologies. The leading short to long-term applications of AI in pharma is more towards reducing the time and hence the cost of drug development. This would not only enhance the return on investment and reduce the costs for users but would be helpful in making useful products available faster, especially where it matters most. With the aid of advances in tech, especially AI, scientists and developers in Africa can be more productive and innovative towards achieving better drug discovery outcomes. This would likely transform pharma and healthcare in the region and globally.
Artificial Intelligence Standardization White Paper	This paper describes China's approach to standards-setting for artificial intelligence. The white paper recommended that "China should strengthen international cooperation and promote the formulation of a set of universal regulatory principles and standards to ensure the safety of artificial intelligence technology." This recommendation was corroborated by previous CESI policies, e.g., its 2017 Memorandum of Understanding with the IEEE Standards Association to promote international standardization.
Asilomar AI Principles	Asilomar AI Principles are 23 guidelines for the research and development of artificial intelligence (AI). The Asilomar principles outline AI developmental issues, ethics and guidelines for the development of beneficial AI and to make beneficial AI development easier. The tenets were created at the Asilomar Conference on Beneficial AI in 2017 in Pacific Grove, California. The conference was organized by the Future of Life Institute. The Asilomar AI Principles are subdivided into 3 categories: Research, Ethics and Values and Longer-Term Issues. Often, the principles are a clear statement of possible undesirable outcomes, followed by a recommendation to prevent such an event.
Assessment List for Trustworthy Artificial Intelligence (ALTAI)	A practical tool that helps business and organizations to self-assess the trustworthiness of their AI systems under development. The initiative follows the High-Level Expert Group on AI's publication: Ethics Guidelines for Trustworthy AI, which proposes seven requirements that AI systems should meet in order to be deemed trustworthy. The

	initiative's mission is 'to guide the development and application of AI in a human-centered approach and to be trustworthy.'
CDEI Review of Online Targeting	A review of online targeting in the UK, proposing three sets of recommendations that relate to increased accountability, transparency and user empowerment with the aim of helping to build public trust and ensuring society and the economy benefit from online targeting.
CEPEJ Ethical Charter on the Use of AI in Judicial Systems and their Environment	The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe has adopted the first European text setting out ethical principles relating to the use of artificial intelligence (AI) in judicial systems. The Charter provides a framework of principles that can guide policy makers, legislators and justice professionals when they grapple with the rapid development of AI in national judicial processes. The initiative's mission is: 'to ensure that AI remains a tool in the service of the general interest and that its use respects individual rights.'
Draft AI R&D Guidelines for International Discussions	The DAI R&D Guidelines for International Discussions and AI Utilization Guidelines were prepared to protect users' interests, prevent spread of risks, and realize a human-centered AI society by promoting the benefits of AI systems and controlling the risks through the sound progress of AI networking, and they are intended for AI developers and users, respectively. They collect the principles and explanations regarding the elements to which developers and users, respectively, are expected to pay attention. They were elaborated as proposed guiding principles to serve as draft non-regulatory and non-binding soft laws to be shared and discussed internationally.
Elements of AI	The Elements of AI is a series of free online courses created by Reaktor and the University of Helsinki. Their aim is to encourage as broad a group of people as possible to learn what AI is, what can (and can't) be done with AI, and how to start creating AI methods. The courses combine theory with practical exercises and can be completed at the user's own pace. It explains the implications of Artificial Intelligence (AI) in real everyday situations with interactive exercises, so that students can make informed decisions as workers, as voters, and as media and product consumers.
Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)	The initiative develops comprehensive suites of objective and verifiable criteria for ethical Transparency, Accountability, Reduction in Algorithmic Bias and Privacy in products, services and systems. So far, it has developed large suites of (roughly 200) criteria for each dimension cited with the exception of Ethical Privacy that's currently under development. The scope of work is generic and universal in that the criteria can be applied to any product/service/system to identify the strength and shortfalls in so far as ethicality is concerned. It also has provisions for customization towards specific priorities, idiosyncrasies and requirements of a given application, industry, discipline or sector.

Global Governance of AI Roundtable	Held yearly in Dubai on the occasion of the World Government Summit (WGS) under the aegis of the UAE State Minister for AI, the Global Governance of AI Roundtable (GGAR) is a revolving international multi-stakeholder governance process that brings together a diverse community of 250 global experts and practitioners from government, business, academia, international organizations, and civil society. GGAR has been envisioned and designed as a unique collective intelligence exercise to help shape and deploy global, but culturally adaptable, norms for the governance of AI. It has no panels, no keynotes; only curated breakout sessions to maximize productivity and outcome. The insights and recommendations have been captured into a comprehensive report, which includes an action-oriented summary for policymakers.
HumanE AI Net	An inter-disciplinary EU research initiative specifically aimed at technical/methodological breakthroughs to operationalize the full spectrum of OECD and European AI principles. It leverages the synergies between the involved centers of excellence to develop the scientific foundations and technological breakthroughs needed to shape the AI revolution in a direction that is beneficial to humans both individually and societally, and that adheres to European ethical values and social, cultural, legal, and political norms. The aim is to facilitate AI systems that enhance human capabilities and empower individuals and society as a whole while respecting human autonomy and self-determination.
iGamma	An AI system to assess the condition of an ecosystem and its benefits. The initiative applies the Ecosystem Integrity Concept which, like human health diagnosis, informs a latent variable through measurable attributes. It has successfully processed data under a unified computational framework based on Bayesian networks, to estimate the condition of terrestrial ecosystems for multiple timesteps, and the crisscross relations of variables that deliver ecosystem services. It is also producing information services (dashboards, reports, and infographics) and disseminates them.
IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	The mission of the IEEE Global Initiative on Ethics of A/IS mission is to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity. It includes the Ethically Aligned Design, First Edition - a comprehensive report that combines a conceptual framework addressing universal human values, data agency, and technical dependability with a set of principles to guide A/IS creators and users through a comprehensive set of recommendations. EAD inspired the IEEE P7000 series: a series of standards projects that address specific issues at the intersection of technological and ethical considerations.
ISO/IEC JTC 1/SC 42	A standardization program made up of eight project working groups aiming to standardize technologies in the area of AI. It also provides

	guidance to JTC 1, IEC, and ISO committees developing AI applications. One committee, ISO/IEC TR 24028, focuses on improving trustworthiness in AI systems as well as identifying standardization gaps in AI. Another committee, ISO/IEC WD TS 4213, is working on an assessment of machine learning classification performance.
Machine Learning Quality Management Guidelines	The Machine Learning Quality Management Guidelines provides a method to enable consistent quality management for AI-based product developments. Its mission is to "manage the quality of products and services using AI safely and securely". The primary output will be a guideline document that provides guidance for goal-definitions and methods for AI developers. Specifically, it builds a quality assessment framework (such as setting levels of quality) associated with some technical guidance (similar to checklists) that allows developers to objectively evaluate quality with aims for international standardization. The initiative also develops tools, publishes reference documents and undertakes academic research on AI quality.
Montreal Declaration: Responsible AI	The Montréal Declaration is a collective endeavor that aims to steer the development of AI to support the common good and guide social change by making recommendations with strong democratic legitimacy. The Declaration's first objective consists of identifying general ethical principles and values, applied to the digital and AI field, that promote the fundamental interests of people and groups. Its mission is to spark public debate and encourage a progressive and inclusive orientation to the development of AI. More specifically, the initiative aims to: (i) Develop an ethical framework for the development and deployment of AI; (ii) Guide the digital transition so everyone benefits from this technological revolution; and (iii) Open a national and international forum for discussion to collectively achieve equitable, inclusive, and ecologically sustainable AI development.
Observatory from the fAIr LAC Initiative	The Inter-American Development Bank (IDB) is leading fAIr LAC with the aim of promoting the responsible development and application of AI to improve the delivery of social services and eventually reducing growing inequalities in Latin America and the Caribbean. fAIr LAC initiative has three main objectives: i) Promote the dialogue around the responsible use of AI focused on citizens from a perspective of diversity and inclusion, through the promotion of a diverse ecosystem of experts, discussion tables, and conferences; ii)) Develop tools to guide the ethical and reliable use of AI in Latin America and the Caribbean through manuals, algorithmic audits, and specific guides; and iii) Encourage responsible AI adoption through pilot projects and the creation of regional hubs. fAIr LAC includes a map of beneficial AI applications in the region that is easily searchable for initiatives by country, sector, or case study.

	It also runs pilot AI projects to systematize the lessons learned from applications where AI helps create greater social impact and to create a cooperative environment so that projects may be scaled and emulated in the region.
OECD Recommendation of the Council on Artificial Intelligence	The initiative provides a set of internationally agreed principles to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values. The Principles focus on AI-specific issues and set a standard that is implementable and sufficiently flexible to stand the test of time in this rapidly evolving field. The principles identify five complementary values-based principles for the responsible stewardship of trustworthy AI and call on AI actors to promote and implement them, these are: inclusive growth, sustainable development and well-being; human-centered values and fairness; transparency and explainability; robustness, security and safety; and accountability.
Open Kinyarwanda	Open Kinyarwanda voice dataset is an initiative to build a freely publicly available speech to text data in Kinyarwanda (Rwanda's official language spoken by over 12 million people in Rwanda & 40 million in the region). Digital Umuganda in collaboration with the German development agency (GIZ), Mozilla & Government institutions is building a dataset of over 1,200 hours and 1,200,000 sentences through crowd-building. The objective is to give innovators, researchers & developers access to a key infrastructure to develop voice technology in Kinyarwanda. The end goal is to take away barriers to access information & services and build inclusive digital solutions that can be accessed by marginalized communities including areas with low literacy levels as well as people living with disabilities.
Partnership on AI Issue Area on Safety-Critical AI (SCAI)	Safety-Critical AI is an initiative within the PAI multistakeholder organization. PAI's goal is to develop the norms, institutions, and technical best practices necessary to ensure the safe research and deployment of AI technologies - particularly in high-stakes, dual-use, and/or safety-critical domains. It does so through a mix of whitepapers, academic research, workshops and convenings, and institutions and services such as expert committees. Thus far, domains that have been identified as high priority and safety critical are healthcare, finance, and autonomous vehicles.
UNESCO Recommendation on the Ethics of Artificial Intelligence & AI Decision Makers' Toolkit	The UNESCO Recommendation expects to define shared values and principles and identifies concrete policy measures on the ethics of AI. Its role will be to help UNESCO Member States and other stakeholders ensure that they uphold the fundamental rights of the UN Charter and of the Universal Declaration of Human Rights and that research, design, development, and deployment of AI systems take into account the well-being of humanity, the environment and sustainable development. The recommendation will have a strong focus on moving from principles to

	practice, including through UNESCO’s AI Decision Makers’ Toolkit - a collection of knowledge products and tools from across UNESCO’s fields of competence to help decision makers address practical questions they face with respect to the development, use and governance of AI.
UNICEF AI for Children	To explore how to embed child rights in the governing policies of AI, UNICEF’s Office of Global Insight and Policy is exploring approaches to protecting and upholding the rights of children in an evolving AI world. As part of the AI and Children policy project, UNICEF hosted a series of workshops around the world to gain regional perspectives on AI systems and children. These conversations helped UNICEF develop a draft policy guidance on how to promote child development in AI strategies and practices. UNICEF offers this draft policy guidance as a complement to efforts to promote human-centric AI, by introducing a child rights lens. The ultimate purpose of the guidance is to aid the protection and empowerment of children in interactions with AI systems and enable access to its benefits.

Section 3. The Road Ahead

This report has classified Responsible AI initiatives across three categories: Ethics, Governance, and Social Good. Ultimately, all three will need to be integrated to ensure an ecosystem where AI is developed and used responsibly to advance the social good agenda as defined by the UN SDGs. Although there have recently been systematic efforts towards integration, this report identifies a need for additional efforts and, hence, puts forward four areas for future action and nine recommendations. Through its interdisciplinary, cross-sectoral and international group of experts on Responsible AI, GPAI can play a unique role in advancing these recommendations.

The four Areas for Future Action are:

1: Prioritize resources towards the most pressing global issues
2. Ensure initiatives are designed for impact
3. Strengthen the ecosystem to accelerate change
4. Respect and champion diversity and inclusion

3.1 Area for Future Action 1: Prioritize resources towards the most pressing global issues

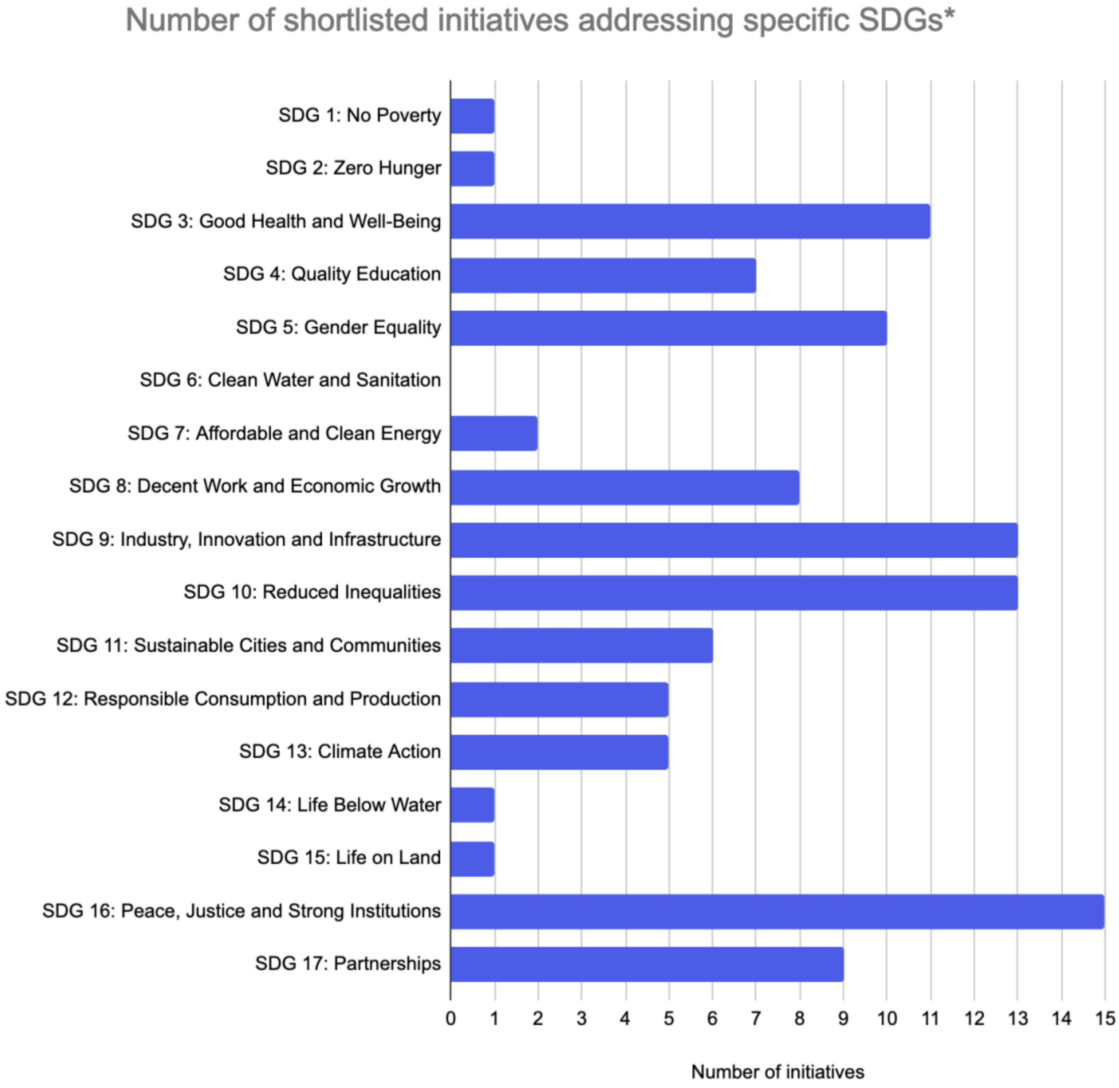
3.1.1 Challenge

AI has wide-ranging applicability and hence has the potential to influence many of the most pressing issues humanity is facing: it can be a force for good to help mitigate climate change or predict the next COVID-19 outbreak, and it can also deepen or give rise to new global challenges as seen through the rise of misinformation. The breadth of potential applications of Responsible AI creates a **prioritization challenge**.

At the initiative level, the analysis reveals that most initiatives have too broad scope and ambition for impact. Notable exceptions to this include the Centre for Data Ethics and Innovation (CDEI) Review of online targeting, which has explored in-depth the hazardous nature of social media targeting, and UNICEF’s AI for Children that, through developing guidance on ensuring children's rights in government and private sector AI policies, has built the policy research capacity and network to place the voice, rights and needs of children on the agenda. However, most initiatives fail to sufficiently target the most important societal issues in national and global agendas today.

The UN SDGs provide an overarching framework to help orient and measure the impact of AI initiatives and the ecosystem at large on the most pressing challenges. Many of the

Figure 3: Distribution of shortlisted initiatives by addressed SDGs



** Assessment is based on the responses gathered from the 30 initiatives. It excludes initiatives that target SDGs agenda as a whole rather than specific SDGs.*

Responsible AI initiatives already rely on the SDGs to the extent that their visions or missions directly or indirectly aim to address them. The AI Commons and the AI for SDGs think tank are two examples of such initiatives, explicitly advancing the SDGs agenda as a whole. Most other initiatives also contribute to a specific subset of SDGs, whether explicitly or implicitly as shown in Figure 3. For example, through its work on biodiversity, iGamma addresses SDG 13 (Climate Action) and SDG 15 (Life on Land). The European Commission for the Efficiency of Justice

(CEPEJ) initiative focuses on the use of AI in Judicial Systems, helping advance SDG 16 (Peace, Justice and Strong Institutions). Some initiatives, however, fail to articulate their activities coherently around the SDG framework or do not liaise with other initiatives and authorities to explore ways to better allocate their efforts.

3.1.2 Recommendations

1. **Build a systematic process to ensure efforts are targeted at the most pressing global issues.**

GPAI should use the SDG framework as a critical prioritization tool and consider the Responsible AI ecosystem in the context of other global objectives and crises that feature high on the agendas of stakeholders worldwide. This contextual approach requires GPAI to update its priority areas on a regular basis and to champion these with actors in the Responsible AI ecosystem. An important aspect would be to detect early warning or weak signals to capture emerging global issues and/or detect shortcomings in addressing them (e.g. unpreparedness for a large-scale pandemic). The process of doing so needs to be rigorous and participatory to uphold credibility and planning certainty, i.e. it should not be erratic on too short a frequency. Conceivable methods could range from qualitative methods (e.g. expert surveys and workshops) to more quantitative methods (e.g. leveraging the SDG indicators). With the multidisciplinary, cross-sectoral, and international group of experts and practitioners it brings together, GPAI is best placed to identify issues, raise awareness about them, and channel efforts to where they are most effective in a given context.

The process for identifying priority areas should acknowledge that not everything can or should be addressed with AI, whether responsible or not, nor at the international level. More generally, GPAI experts should consider four main factors in identifying priority areas for GPAI, as presented in Table 7. The Responsible AI Working Group should periodically (e.g. every year) request its members to take part in a round of priority area identification based on that framework.

Table 7: Factors for Identifying Priority Areas

Impact	**Urgency**	**Feasibility**	**Relevance**
Whether there is a strong need for this priority area to be addressed globally	Whether a window of opportunity exists to make significant progress on the issue at hand	Whether AI can technically be applied to resolve the issue in a way that is cost-effective and safe	Whether GPAI as an international platform is well-placed to address the issue
Sample questions			

Is this issue high on the agendas of stakeholders worldwide? Is there a market failure making the private sector unable to address this?	Are there significant costs in delaying solving this problem?	Are there successful use cases of AI applied to that area? If so, what are the barriers to scalability? If not, are there promising theoretical avenues?	Is international collaboration for that area beneficial, among GPAI members and beyond? Can significant independent and interdisciplinary expertise add value?

To channel efforts towards priority areas, GPAI should leverage its role as an international bridge on Responsible AI between governments and other stakeholders in the ecosystem. GPAI experts should look at building response capacity both among governments and within the broader ecosystem. This could take the form of a procedure to facilitate multistakeholder collaboration or processes to share and raise awareness of the priority areas with governments, major funding bodies, key private sector players, and R&D policy authorities to foster a wider response. The Responsible AI Working Group should discuss the design and official adoption of this procedure in one of its upcoming meetings.

2. **Create focused committees to address identified pressing issues.**

Parallel to building a robust process that assesses the impact of initiatives against the SDGs and identifies new pressing issues, the GPAI Responsible AI Working Group should start advancing on a selected number of areas that have already been proposed and considered relevant as per the four factors in Table 7. Each area should be covered by a focused committee with the mandate to provide advice and recommendations to GPAI governments that reflect the interdisciplinary scientific consensus on the specific issue at hand. This could require additional landscaping to identify related initiatives, surveying progress made in existing literature, a list of major uncertainties and gaps, and actionable recommendations in concrete, unambiguous languages to help policymakers, civil servants and citizens make informed decisions.

The specific areas and corresponding committees which the GPAI RAI Working Group has identified as initial priorities are:

- **Committee on Governance and Transparency of Social Media (SDG 16: Peace, Justice and Strong Institutions):** The committee's work could be divided into three streams, each with its own goals. The first to develop principles and tools drawing the line between what is acceptable or not in terms of profile targeting in advertising and social media. The second to encourage and enable joint research initiatives towards the development of AI tools (e.g., detecting fake news, flagging demagogic messaging, fostering pluralistic views, promoting diversity, bridging controversies) and the assessment of the impact of such tools. This stream could also enable independent research on existing tools used by social media companies to curate and transmit information on their platforms, including content classifiers (e.g. for disinformation or incitements to violence) and recommender systems that disseminate content amongst users. The third to advocate for increased transparency in the operation of social media companies, in relation both to profile targeting and to content classifiers and recommender algorithms. The scope of all three streams should encompass the salient issues in social media (e.g. transparency requirements in political advertising, abuse by terrorist actors) but also less salient issues (e.g. children's vulnerability in social media environments, auditing techniques for modern communications). This committee could leverage the work and findings from, amongst others, initiatives such as the CDEI Review of Online Targeting.

- **Committee on AI and Education (SDG 4: Quality Education):** The committee could define collaborative projects whose implementation would contribute to (a) maximizing the benefits of AI for education management and delivery, empowering teaching and teachers, improving learning outcomes and learning assessment, offering lifelong learning opportunities for all, etc., and (b) addressing cross-cutting themes such as promoting equitable and inclusive use of AI in education; AI literacy and skills as well as training students to become responsible producers and users of AI; monitoring the impact of AI on education; and researching the applicability of AI for education solutions in lower-income countries. This committee could leverage the work and findings from, amongst others, initiatives such as Elements of AI.
- **Committee on Drug Discovery and Open Science (SDG 3: Good Health and Well-Being):** The committee could (a) examine how to create a favorable context for AI to contribute to drug discovery in an open and equitable manner, whereby international public health needs are privileged over profitability. It could also (b) examine how R&D efforts could best be organized and what the rules of engagement should be to ensure licensing of key drugs is attainable for lower-income countries. A particularly pressing example and use case would be that of antibiotics, whose market failure (lack of profitability in developing them) is expected to lead to over 10 million deaths per year by 2050.[20] Collaboration and cross-dialogue with GPAI's Pandemic Response Working Group, and close consultation with other key actors in the field of infectious diseases drug discovery (like for Covid-19, and generally not well addressed currently) will be important to ensure ongoing alignment. This committee could leverage the work and findings from, amongst others, initiatives such as Artificial Intelligence and Blockchain for Healthcare in Africa.
- **Committee on Climate Change & Biodiversity (SDG 13: Climate Action, SDG 14 & 15: Biodiversity):** The committee could (a) elaborate on practical collaborative approaches to fight climate change (e.g. to ensure AI is making zero-carbon renewables

[20] Interagency Coordination Group on Antimicrobial Resistance (2019) "No Time to Wait: Securing the Future From Drug-Resistant Infections - Report To The Secretary-General Of The United Nations", IACG and Tagliabue, A., Rappuoli, R. (2018) 'Changing Priorities in Vaccinology: Antibiotic Resistance Moving to the Top.' https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5992407/

as productive as traditional hydrocarbon suppliers), and (b) examine how to apply AI in a more environmentally-friendly way (e.g. to better evaluate the environmental impact of machine learning), (c) incorporate efforts to halt or possibly reverse the loss of biodiversity as a result of climate change (SDGs 14 & 15: Life below water and Life on land). This will be an important lever and one where AI can play an important value-added role, particularly in reconciling sustainable agriculture practices and economic opportunities. This committee could leverage the work and findings from, amongst others, initiatives such as iGamma.

3.2 Area for Future Action 2: Ensure initiatives are designed for impact

3.2.1 Challenge

Understanding the realities on the ground will be key to making efficient and effective solutions progressing on the responsible development and deployment of AI. Yet, many initiatives **lack compelling impact pathways or measurement metrics** to ensure progress is made in a systematic way.

Impact pathways (eg. theories of change, result matrixes, etc.) are crucial for understanding a problem and, hence, its potential solutions. However, the report finds many initiatives in the ecosystem lack impact pathways which are well-specified and accessible. As a result, it is difficult to measure and reward initiatives' performance on their ability to advance specific SDGs, sub-goals or other key performance indicators. This can be for several reasons. First, a theory of change might not be compelling or explicitly articulated as such. This is particularly the case for initiatives in the Ethics or Governance categories that are, in most cases and almost by nature, further removed from action on the ground. Second, many initiatives do not seem to liaise with relevant authorities at the national or local levels directly, and therefore do not enter any national statistics that track progress towards the SDGs agenda or beyond.

Beyond clearly articulated impact pathways, the ecosystem also suffers from unclear **metrics and performance benchmarks**. In turn, this makes it difficult to develop a measurement system tracking progress on an international level. On an ecosystem level, some initiatives do exist to address this issue and can be leveraged, notably the AI Index Report[21] hosted at Stanford University. However, overall, the absence of metrics (or consensus on what these metrics should be) across the majority of initiatives makes it difficult to assess and get a good sense of realities on the ground, and to inform a coordinated approach to address some of the pressing issues as mentioned in Area for Future Action 1.

[21] AI Index Report. 2019. Stanford HAI. https://hai.stanford.edu/research/ai-index-2019

3.2.2 Recommendations

3. **Develop a common taxonomy and international measurement system among GPAI governments.**

GPAI should champion and initiate an international agreement on defining Responsible AI with a concrete, efficient and effective measurement system. This should include defining measurements of the AI system itself, of the application in deployment, and of the impact the deployed application has on its environment. A set of system-specific metrics would help understand the nature of the system in isolation, prior to deployment, and would enable comparison and assessment of ex ante inherent risks, as well as create financial incentives for private or public organizations towards contributing to the public goods quantified by these metrics. These could include for example agreed-upon definitions and metrics: of fairness, of explainability, of the extent to which the system behaves as expected and avoids corner solutions, of the extent to which the problem at hand can be safely explored, of the system's energy consumption, of the size and heterogeneity of the training dataset used, etc. A set of metrics that assess deployed applications would help assessment and comparison of risks to users, operators, or consumers' safety and fundamental rights. These include agreed-upon metrics: of the extent to which humans intervene, of the number of individuals affected by the decisions, of the criticality and societal and economic costs and benefits associated with failure and adoption, etc. Finally, metrics to capture the ultimate impact to the extent possible would help factoring in the broader societal concerns when deciding to roll out the AI system at scale. These include the 231 indicators supporting the SDG framework.[22]

This should be accompanied by an evidence-based and agreed-upon taxonomy of concepts pertaining to AI itself, how responsible it is, and its impact. This taxonomy, in particular when it comes to the impact of the technology, should be articulable with the SDGs framework and other relevant taxonomies that already inform policies among GPAI governments.[23]

Each focused committee could identify the dimensions relevant to measure the socio-economic and environmental impact related specifically to their priority area, in a coherent way, which highlights the importance for these committees to be sufficiently focused. This could lead to the development of performance benchmarks that permit consistent assessment of AI system capabilities globally. Such discussions and taxonomy should draw upon the work of others in the international arena to avoid duplication of efforts (e.g. OECD ONE.AI taxonomy of AI systems or EU taxonomy on sustainable financing). Moreover, given the fundamental aspect of these definitions, taxonomy, and measurement system for all of its focused committees and potentially for the other Working Groups of GPAI, the Responsible AI Working Group should select some of

[22] United Nations Statistics Division (2020) "SDG Indicators - Global indicator framework for the Sustainable Development Goals and target of the 2030 Agenda for Sustainable Development", United Nations https://unstats.un.org/sdgs/indicators/indicators-list/

[23] For example, New Zealand's Wellbeing Budget or the EU's taxonomy of sustainable activities. https://www.treasury.govt.nz/sites/default/files/2019-05/b19-wellbeing-budget.pdf

its members to form a taskforce to coordinate the drafting of that agreement. In particular, it should appoint "liaison members" with the three other Working Groups to identify and capture synergistic potential on this agreement and, beyond, across its focused committees and activities.

4. **Construct a widely applicable and coherent impact assessment methodology.**

The GPAI Responsible AI Working Group should develop an impact assessment methodology through repeated iterations. This methodology would aim at the operationalization of the taxonomy mentioned above (cf. Recommendation 3), the development of concrete auditing and evaluation mechanisms and the path for how these guidelines could be standardized across governments. The impact assessment methodology would coherently address two fundamental aspects of Responsible AI: governance (dealing with questions of whether the AI socio-technical system indeed is compatible with societal values) and AI for Good (dealing with questions of whether the application of the AI system advances the SDGs or other societal goals). In doing so, it would also unify these often separate pillars, as specific governance rules (such as fiscal incentives) may be necessary to efficiently achieve the AI for good objectives.

Such methodology would promote and facilitate reporting by initiatives of their own impact and further advance refinement of their theory of change. Overall, it would help inform the international agreement on a common taxonomy being forged in parallel and converge towards it (cf. Recommendation 3). The methodology will have to be initially at a high-level, with focused committees helping refine and adapt it for various, more granular priority areas. It should leverage existing efforts in this direction including AI Now’s Algorithmic Impact Assessment Framework and should ensure a participatory and multidisciplinary approach. The Responsible AI Working Group should convene an online workshop with GPAI experts to agree on the objectives and main functions of the methodology and to identify who among its members should be responsible for proposing a first draft.

3.3 Area for Future Action 3: Strengthen the ecosystem to accelerate change

3.3.1 Challenge

The priority areas from ‘Area for Future Action 1: Prioritize resources towards the most pressing global issues’ provide GPAI with a clear focus. The measurements, taxonomy and impact assessment methodology from ‘Area for Future Action 2: Ensuring initiatives are designed for impact’ will enable a better understanding of the current situation. With these, GPAI will be in a position to catalyze change and alter the trajectory of the ecosystem. However, it needs to go beyond that. There needs to be a focus on building a strong and healthy ecosystem that supports and stimulates change.

First, societies need **governance tools and frameworks** to ensure they systematize the responsible adoption of AI at large. These tools will have to be designed to benefit from the gains in transparency and clarity made through progress on Area for Future Action 2. The objective of these tools should be to alter incentives and behaviors throughout society to help the adoption of Responsible AI practices.

Second, the Responsible AI initiatives need more systematic **collaboration and cooperation**. A symptom of this captured in the catalogue is the numerous initiatives with overlapping scope. While multiplication of efforts can at times be value-additive by enabling competition among initiatives (e.g. to develop better operational practices), it may also be confusing for stakeholders that are directly affected by Responsible AI but not directly part of the ecosystem (e.g. domestic regulators, trade unions, local authorities). An important task for the domain-specific committees proposed in Recommendation 2 is to assess healthy levels of competition versus places where there is room for synergy and cooperation.

Third, at the ecosystem level there is also a need to **connect cross-cutting initiatives to the domain level.** For instance, an ethics framework needs feedback from domains deploying such a framework and domains need to know how to access well-supported ethics frameworks. Further efforts to promote this kind of interaction in the ecosystem is needed to ensure cooperation.

The wider challenge here is not simply the design of tools, but their feasibility of implementation by GPAI governments and society beyond the Responsible AI ecosystem. Attention to this will **promote the ability of initiatives to scale** and reach their impact potential. The shortlisted initiatives show good practices to ensure this is possible —in particular simplicity of the tool and testing. For example, the Algorithm Charter for Aotearoa New Zealand is a short, simple way for individuals unfamiliar with Responsible AI to know how to alter their organization's practice. The Assessment List for Trustworthy Artificial is interesting because of its reliance on a wide "piloting" consultation to better assess the burden of compliance. GPAI could rely on lessons learned from these and other national experiences.

3.3.2 Recommendations

5. **Create a focused committee on governance issues and governance means.**

In parallel to the committees focused on thematic priority areas identified in Recommendation 2, GPAI Responsible AI Working Group should create a committee that focuses on cross-cutting governance issues and means to govern the development and deployment of AI towards Responsible AI.

- **<u>Committee on Governance Issues and Governance Means:</u>** This committee could work on the objectives and mechanisms of governance to make sure that AI systems are designed and used by organizations in an **accountable and transparent** manner **to ensure fairness, safety, robustness, respect for human rights, the promotion of equity and deployment beneficial to the public good.** The scope of this committee could encompass a review, synthesis, research and refinement of existing and promising governance mechanisms (e.g. sector-specific templates of internal policies, procedure for determining scope of audit systems, risk assessment methodology, fiscal incentives, etc.) as well as their promotion among GPAI governments. In doing so, the committee should act with the independence appropriate for a scientific advisory body.

Specifically, the committee should establish practices that ensure stakeholders maintain accountability for their endeavors despite having introduced technologies that might allow for blurred lines of accountability. **Transparency** must be a characteristic of all digital systems. Developers and operators of AI systems must ensure adequate means to trace accountability through their systems so that any problems can be identified as unfortunate, negligent, or of malicious intent, and the intent must be traceable to the accountable entity – whether a developer, an owner/operator, or criminal entry. The burden of providing transparency should be proportionate to the economic impact of the system.

The application of **AI in high-stake decision making** is a particular issue to consider. Decision making is a process whose quality should be assessed in terms of the final outcome—the quality of the decision—rather than assessing only the quality of the decision-support AI tool in isolation (e.g. in terms of its predictive accuracy). It is therefore key to design and develop AI tools that empower the cognitive capacities of the decision maker through a fruitful human-AI collaboration (explainable AI is key here), to the end of recognizing and mitigating bias and discrimination, and ultimately improve the fairness and transparency of the decisions. This goal is quite challenging for the current generation of AI decision-support tools and should be properly addressed at the technical and normative level to foster quick progress and avoid missteps.

On the research side, the committee could sponsor research into what governance or technical tools are needed to implement widely accepted principles, and what additional operationalization and development is needed to make the tools cost-effective for AI for Social Good use cases. This would help the ecosystem move from principles to practice, and from practice to impact. GPAI experts and governments could develop and promote a research agenda for Responsible AI helping the ecosystem move towards implementation. The research agenda could focus on common socio-economic, operational, technical and scientific challenges that need to be overcome to ensure Responsible AI can be implemented broadly in society. Overall, the committee should regularly take stock of progress in the implementation of the governance tools and resolution of the associated challenges.

6. **Facilitate coordination within the ecosystem.**

GPAI should set up a coordination mechanism to facilitate communication across initiatives, thereby enabling initiatives to leverage each other's learnings and good practices. Two aspects of this coordination are important. First, some of these initiatives should increase coordination between existing initiatives developing principles, those developing governance tools, and those having applied projects on AI for Good - not duplicating their efforts. Thanks to the Working Group experts' authority, expertise and familiarity with many of these initiatives, there are some synergies that can be captured through the forum it provides for discussing common issues. The Working Group should encourage spreading these discussions in its own network.

Second, some multistakeholder initiatives should coordinate the operationalization of governance tools that stimulate change. Learning from the Assessment List for Trustworthy AI which worked with a wide range of stakeholders to help operationalize the associated ethical principles, these initiatives must include actors otherwise unfamiliar with concerns posed by Responsible AI, such as national and local authorities, development stakeholders, non-tech private sector, and civil society organizations. These initiatives would help pilot and improve the recommended governance tools before rolling them out among GPAI governments.

In areas identified as promising by GPAI and existing initiatives, Working Group members should coordinate with the focused committees (cf. Recommendation 2) to initiate the creation of Public Private People Partnerships across geographies that would tackle priority areas while at the same time testing the governance tools in deployment. In this way, the experts would ensure that the sometimes novel institutional and technical tools they recommend are effective and efficient.

In one of its upcoming meetings, the Responsible AI Working Group could discuss the establishment of a function helping Working Members address these external relations mentioned above. This could take the form of nominating a "Liaison Member in charge of External Affairs", a taskforce of such members, or even support from the secretariat and GPAI collaborators.

7. Build capacity for policymakers to govern for Responsible AI.

While the focused committee on governance issues and governance tools can help research, develop and test new mechanisms to change incentives, it will often ultimately be governments implementing these mechanisms. GPAI should assist governments in building capacity for the governance of Responsible AI. This is also work that would benefit from collaboration with the OECD, given its history in helping diffuse good practices on governance to its member states.

For example, a specific area that GPAI could help build capacity on is **international standardization processes**. Many of the initiatives shortlisted (eg. IEEE Global Initiative, ISO/IEC JTC 1/ SC 42, and AI Standardization White Paper) are working towards setting standards on various and sometimes overlapping aspects of Responsible AI. For these standards to be effective, it is important for governments to be involved through mechanisms like OCEANIS.[24]

[24] OCEANIS is a global forum for discussion, debate and collaboration for organizations interested in the development and use of standards to further the development of autonomous and intelligent systems. Read more here: https://ethicsstandards.org/.

Other specific areas where GPAI experts could help governments build capacity include:

- Capacity to link accountability tools with metrics and taxonomy developed by GPAI (cf. Area for Future Action 2);
- Capacity to make fiscal incentives conditional on specific performance in line with Responsible AI;
- Capacity to productively organize dialogue and coordination among various AI stakeholders as well as between these stakeholders and traditional social actors, regulators and local authorities; and,
- Capacity to understand, assess, and deploy the relevant tools to govern AI.

The focused Committee on Governance Issues (cf. Recommendation 5) could be mandated to take the leadership within GPAI for these collaborative capacity-building activities.

3.4 Area for Future Action 4: Respect and champion diversity and inclusion

3.4.1 Challenge

All three previous Areas for Future Action have the potential for altering the trajectory of the ecosystem. However, it is fundamental to GPAI's success that the new trajectory embodies the views of everyone and every community. Many initiatives in the shortlist, and many more in the catalogue, have struggled to collect representative input to inform their activities. This **lack of inclusiveness** points to a lack of capacity to involve a wider group in the current technological transition and, hence, co-shape innovative solutions. Widening the ecosystem and obtaining a more representative range of input among initiatives will be necessary to ensure that Responsible AI is by and for everyone.

Specifically, the sample of initiatives in the catalogue suggests an underrepresentation of initiatives from the Global South and marginalized communities, such as people with disabilities, indigenous groups, the LGBTI+ community, persons living below the poverty line and migrants. There are of course notable and recent exceptions, such as the A+ Alliance for Inclusive Algorithms.

This lack of diversity risks undermining the effectiveness and credibility of Responsible AI initiatives as well as their ability to scale. Importantly, it risks perpetuating existing inequalities and biases and misinforming policy priorities. This is particularly problematic for cross-regional collaborations.

3.4.2 Recommendations

8. **Develop and disseminate good Diversity & Inclusion (D&I) practices.**

GPAI should help shape and spread good D&I practices across the ecosystem. As a first step, GPAI could formulate an inclusion strategy that helps gauge the extent to which segments of society or geographies are currently underrepresented or excluded in the Responsible AI debate. This would include specific policies, objectives, activities and a results matrix to monitor and report on progress.

As a second step, GPAI could encourage open-access to information and infrastructure, in a widely accessible and user-friendly manner. Accessibility and user-friendliness of initiatives should be stressed as key mechanisms to include the perspectives of those marginalized, less technically versatile or digitally literate.

As a third step, GPAI could break down communication barriers between geographies, social groups, and disciplines. Amongst others, communication barriers may include language constraints, cultural differences, restricted access to (digital) information, knowledge gaps and suboptimal user-friendliness. To address these, GPAI could support efforts to make material on Responsible AI initiatives widely available in different languages for diverse target groups.

GPAI should build on the good practices demonstrated by several initiatives fostering a multistakeholder diverse process. For example, the Observatory from the fAIr LAC Initiative includes stakeholders from across Latin America and all sectors by design. Digital Umuganda's Open Kinyarwanda has maintained a very open and crowdsourced approach to contribution, connecting Rwandese developers, translators and users in building an open source Kinyarwanda voice data set. The UNESCO Recommendation on the Ethics of Artificial Intelligence relied on a global online survey for experts from 155 countries and 12 regional and sub-regional consultation, which amounted to nothing short of 50,000 comments. The Montreal Declaration involved citizens through deliberations to co-shape AI principles. Elements of AI, a series of online courses that explains the concept of AI as well as its ethical and legal challenges for a layperson, has a wide reach by being free of charge, designed in an attractive manner and widely promoted via various outlets.

Learning from these, the Responsible AI Working Group should convene its members and delegate responsibility to a taskforce to formulate an inclusion strategy for GPAI. It could also discuss the creation of a second taskforce responsible for finding, promoting and implementing accessibility and user-friendliness and finding cost-effective ways to make key materials on Responsible AI initiatives widely available for diverse target audiences.

9. **Initiate strategic partnerships with platforms collecting representative input.**

Beyond being a provider of good D&I practices, GPAI should work with international organizations like OECD, WHO and UNESCO to proactively collect representative input from marginalized groups and the Global South. GPAI’s role in this endeavor can be threefold. First, it can proactively identify and invite historically marginalized groups and underrepresented regions to share their concerns, priorities, and solutions. Second, it can liaise with initiatives whose activities would benefit from a more inclusive and diverse perspective and assist in setting up such a process. Third, GPAI can promote and help create synergies between existing efforts that foster basic digital and AI literacy, as well as civic engagement. Such initiatives are essential to build understanding of AI and its key socio-ethical challenges across a public domain audience. The Responsible AI Working Group should convene its members and mandate a subset of them to design the procedure to conduct these three activities and discuss the official adoption of this procedure.

Conclusion

This report has provided a high-level overview of existing initiatives that are active in the field of Responsible AI. Through a common assessment framework, it hones in on a subset of promising initiatives that could further advance the objectives of GPAI and are aligned with both the UN SDG and OECD AI Principles. A deeper analysis of this subset of initiatives —their challenges, key success factors and lessons learned— has allowed identification of four areas for future action and nine specific recommendations to inform GPAI's agenda-setting going forward.

There is an important role for GPAI to play in strengthening the rules and workings of the game in the overall Responsible AI ecosystem. In doing so, GPAI would facilitate the ecosystem in the move from principles to practice. Establishing priority areas to focus efforts towards would help deliver meaningful progress towards the GPAI mandate. Championing the development of a taxonomy of Responsible AI as well as an impact assessment framework would provide common language and a framework to leverage synergies across initiatives in the ecosystem. Developing a dedicated committee to inform governance processes across geographies, sectors and actors would be important to shape and uphold these rules of engagement over time. Furthermore, advancing collaboration, cooperation and connecting cross-cutting initiatives would be an important part of such efforts and one where GPAI can act as a bridge and traveler between worlds.

Finally, tomorrow's world can only be representative of all its citizens if their voices are heard *today*. A Responsible AI ecosystem will need to reflect preferences and values across all parts of society and geographies to be fully legitimate in the long run and preempt social discontent. Conceiving an inclusion strategy for ensuring representativeness in the development of Responsible AI and serving as a broker for representative input across initiatives is a critical void to fill.

At this critical juncture in time, it will take coordinated efforts across geographies, sectors and actors to set Responsible AI on a sustainable course. Through setting and assessing this course, encouraging a strong governance system and finally ensuring that responsible AI reflects the world's diverse voices, GPAI is uniquely poised to give Responsible AI the sustained momentum to carry it across space and time.

Appendix

Appendix 1: Catalogue

GPAI RAI Living Catalogue of Initiatives is accessible by clicking here or by copying the following link and pasting it into your browser:

https://docs.google.com/spreadsheets/d/1utA1ug1nKF3B4MMmC_AQeyo1IK5jH_abMK9oFHEJekk/edit#gid=0

Appendix 2: Criteria Assessment Framework

GPAI RAI Criteria Assessment Framework is accessible by clicking here or by copying the following link and pasting it into your browser:

https://docs.google.com/document/d/1fziVWtO4VS2RPXp0yoJFEabruvbeI-qaEwvs3Sy8OLg/edit

Appendix 3: Analysis of 30 Shortlisted Initiatives

1. AI Commons

Initiative	AI Commons
Category	AI and Social Good
Brief Description	A global knowledge hub bringing diverse stakeholders together to address the world's greatest challenges using AI. Its key objectives are to identify how beneficial AI can be designed and implemented in an inclusive and distributed manner, and to create open source blueprints for global usage. The initiative strives to advocate and make possible the concept of AI being a public good.
Organization	AI Commons
Geography	The scope of the initiative is global.
Sector	The initiative is a nonprofit organization, but stakeholders involved come from multiple sectors including the private sector, civil society, academia, and international organizations.

Key Success Factors	Success for the initiative means showcasing of benefits to those that are closest to the problems. Metrics to measure success consists of a series of impact measurement of solution usage, local sourcing of data, involvement of problem owners in the solution design, and sustainability of solutions. The initiative has been able to i) create a unified approach to evaluate problem solving with AI; ii) build a clear definition of problem owners vs. problem solvers, and create a cross-sector collaboration framework; iii) create a knowledge hub of how local solutions can be built; and iv) organize local living labs that can identify and help build local solutions.
Key Hurdles	Challenges include i) finding local funding resources to engage problem stakeholders in communities; ii) sustainable access to quality data; and iii) need for additional outreach to communicate the benefits and hence engage more stakeholders.
Potential to contribute to GPAI objectives	Initiative contributes to the responsible development of AI by serving as a collaboration platform for applied projects advancing SDGs. Furthermore, it already has achieved cross-regional reach and involves stakeholders from various sectors and disciplines, including AI practitioners, entrepreneurs, academia, investors, NGOs, AI industry players and organizations/individuals focused on the common good.
Diversity and Inclusiveness	Founding members, steering committee, and team are composed of individuals from different geographical and sectoral backgrounds. Currently, the team is limited to volunteers relying on local living labs. By design, AI commons includes the perspectives of problem owners and helps identify solutions that are beneficial to people closest to the problems. All resources are provided open source for stakeholders worldwide.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has shown progress towards its objectives and achieved the following impact: i) validation of matching solutions to really benefit problem owners; ii) encouraging data sharing; iii) evaluating Responsible AI design. Furthermore, solving for benefits going back to users is helping provide ways to engage local problem champions and help with inclusive operations and deploying responsible AI. The initiative is clearly aligned with the UN SDGs, aiming to advance progress towards all seventeen of the goals. Furthermore, it supports the implementation of the following OECD principles: inclusive growth, sustainable development and well-being, human-centered values and fairness.
Maturity / Potential for adoption	The initiative was born in 2016 and has launched eight initiatives since 2018. Given its multistakeholder nature and engagement with key partners (eg. ITU, IEEE, Global Pulse, United Nations Interregional Crime and Justice Research Institute [UNICRI]), World Bank) it has high potential for adoption.

2. AI Explainability 360

Initiative	AI Explainability 360
Category	AI and Governance
Brief Description	AIX360 is an open-source software toolkit that explains AI models as well as the data they operate on. It also provides a taxonomy of explainable AI techniques and educational materials, including a web demo, glossary, and tutorials illustrating its use in application domains. AIX360 aims to bridge the gap between the AI community and society at large. For data scientist users who are not AI experts, it helps them select an appropriate technique and successfully deploy it in their domain. For policymakers, it provides education on explainable AI technology to promote appropriate regulatory actions. For AI researchers, it points out understudied areas and provides a vehicle for disseminating new techniques. AIX360 has been donated to the Linux AI & Data Foundation and is part of a broader Trustworthy AI initiative including AI Fairness 360 (aif360.mybluemix.net), Adversarial Robustness Toolbox (adversarial-robustness-toolbox.org), and AI Factsheets 360 (aifs360.mybluemix.net).
Organization	IBM
Geography	The initiative originated in the US but has now had global outreach.
Sector	Private sector
Key Success Factors	AIX360 is distinguished by the principle that "one explanation does not fit all." This is demonstrated by its coverage of more types of explanations than other toolkits, which allows it to serve a wider range of users and use cases. It is also reflected in its emphasis on educational materials, which help foster wider adoption. These materials have already served to inform financial industry users and government regulators about the explainable AI space. A second category of success factors has to do with its nature as an open-source project. AIX360 is designed to be easy to install, use, and extend (as shown by the successful integration of contributions from the community). Active communication channels in GitHub and Slack ensure that users receive the help they need and have their questions answered. This project is part of IBM Research's Trusted AI initiative (https://research.ibm.com/artificial-intelligence/trusted-ai/). It includes research in the areas of bias, transparency, and adversarial robustness.
Key Hurdles	Most of the algorithms (8 out of 10) in the open source toolkit were novel. The team conceptualized, designed, implemented and tested them before including them in the toolkit. Moreover, they built a custom, easy to extend architecture that seamlessly integrates these very different methods, something that had not been done before. They also created runnable, yet highly descriptive Jupyter notebooks for individuals to be able to

	understand and use the provided capabilities. All these aspects presented significant scientific and technical challenges. To be able to resolve them, stakeholders from different time zones (US west coast, east coast, India) as well as skill sets (core AI researchers, developers, human-computer interaction researchers) had to come together. This was a non-trivial operational challenge for the initiative.
Potential to contribute to GPAI objectives	The initiative serves as an example of how explainable AI can be promoted and fostered, supporting both developers and policymakers. Potential for cross-regional collaboration is high as it is crucial for the initiative's success of an open-source project to ensure that its innovations satisfy the diverse needs of explainable AI for all constitutions.
Diversity and Inclusiveness	The original creators of AIX360 are 20 experts from the IMB Research team. These creators are based in the US, India, and Argentina, and have expertise in AI, machine learning, and human-computer interaction. Participation in the AIX360 open-source community is open to the world. In fact, one of the initiative's lessons learnt is that open governance of the toolkit encourages more participation. Its approach focuses on the ultimate consumers of AI explanations, including affected citizens, non-AI domain experts, and regulators. AIX360's educational materials are aimed at different levels of AI expertise, with the web demo being the most accessible.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Success is measured by how widely adopted the toolkit is and how useful organizations and individuals find it to be. This includes but is not limited to GitHub stars (697), forks (151), public slack users (197) as well as comments from individuals with varied expertise. This work is impacting IBM commercial offerings like Watson OpenScale, AI Governance, and Trustworthy AI consulting. The initiative directly addresses SDG 5 (Gender equality) and SDG 16 (Peace, Justice & Strong Institutions) as well as the following OECD principles: human-centered values and fairness, transparency and explainability, robustness, security and safety, accountability.
Maturity / Potential for adoption	Launched in 2019, the initiative started with creators from North America, Asia, and South America and continues with contributions from other regions. Being open source, it has high potential to be adopted and used by developers and policymakers worldwide.

3. AI for SDGs Think Tank

Initiative	AI for SDGs Think Tank
Category	AI and Social Good
Brief Description	An online global repository compiling and analyzing AI projects and proposals that impact the UN SDGs both positively and negatively. It also includes a detailed evaluation of each initiative featured. The initiative's mission is to 'promote the positive use of AI for Sustainable Development and investigate the negative impact of AI on sustainable development.'
Organization	Research Center for AI Ethics and Sustainable Development at the Beijing Academy of Artificial Intelligence.
Geography	The initiative was launched in China but has global scope.
Sector	Civil society
Key Success Factors	Some of the key success factors include a user-friendly interface which allows viewers to search by SDG goal or by a specific topic. Furthermore, the information is crowdsourced, and each initiative is scored based on a common rating scheme.
Key Hurdles	Some challenges include i) lack of comprehensiveness; ii) lack of transparency over rating scheme used to evaluate each initiative; and iii) unclear impact.
Potential to contribute to GPAI objectives	As a repository, it has the potential to advance GPAI's objectives by promoting collaboration and reducing duplication in the area of AI. Furthermore, given its global scope, it has the potential to attract the attention and engagement of cross-national and cross-sectorial stakeholders.
Diversity and Inclusiveness	Thirteen AI experts (from China, UK, US, Japan, Canada, Singapore, Brazil, the Netherlands) serve as the founding members. The initiative aims to feature initiatives from all over the world, although, due to its early stages, it currently has low coverage from specific regions like the Global South and MENA. All of the information is open to the public online and anyone can easily submit an initiative following their guidelines.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has captured over 215 applied projects and has developed a common framework to evaluate them, yet specific metrics on its performance are not specified. Furthermore, it has already launched a research program to delve into the positive and negative impacts of AI on SDGs, and a cooperation network. Given its mission, there is clear alignment to the UN SDGs framework.

Maturity / Potential for adoption	Launched in 2020, the initiative is still in its early stages. However, it shows signs for adoption worldwide as it serves as a repository of applied projects that stakeholders worldwide can learn from and use to build synergies.

4. AI for Good

Initiative	AI for Good
Category	AI and Social Good
Brief Description	The AI for Good Global Summit is a United Nations platform, centered around annual global summits, that foster the dialogue on the beneficial use of Artificial Intelligence, by developing and identifying concrete projects. The AI for Good Global Summit series aims to bring forward Artificial Intelligence research topics that contribute towards more global problems, through accelerating the United Nations' Sustainable Development Goals (SDGs). Close to 40 UN organizations are partners of the AI for Good Global Summit and they also bring together experts from industry, government, civil society, academia, etc. It includes the AI Repository, a catalogue of AI initiatives which accelerate progress towards the seventeen UN SDGs.
Organization	International Telecommunication Union (ITU)
Geography	The initiative is cross-regional and has global coverage.
Sector	International organization, involving stakeholders from various sectors including academia, civil society, private sector and public sector.
Key Success Factors	Key success factors include i) developing concrete projects and identifying practical applications for beneficial use of AI; and ii) engaging an as-wide-as-possible audience with, in particular, machine learning experts and problem owners. Success indicators are: active outreach for inspiring speakers and diverse audience to connect AI innovators with public and private-sector decision-makers in the interests of stimulating the discovery and delivery of “AI for Good” solutions for all, engaging influential speakers and audience, global media coverage, gender balance, regional balance, quantity and quality of contents, and outcomes of the summit (i.e. focus groups, global initiatives, etc.).
Key Hurdles	Challenges have included: i) scaling the AI for Good Global Summit projects; and ii) matching problem owners with providers of relevant solutions. One of the key lessons from the initiative is making concrete advances means having informed discussions with genuine expertise, or it will end up with generic platitudes.

Potential to contribute to GPAI objectives	Initiative contributes to GPAI objectives by promoting collaboration bringing different stakeholders together to build synergies and work together to apply AI for Good. A special focus is put on ensuring the platform includes the perspectives of groups which are underrepresented.
Diversity and Inclusiveness	The team is diverse in terms of region, gender and backgrounds, with education in diverse fields, including physics, economics, and business. One of the aims of the summits is to ensure safe and equitable AI, as such, it tries to ensure an as diverse audience as possible. In past years, scholarships were provided to those needing them. Currently, the summit is online.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The summit has achieved its objectives mainly through the establishment of the following platforms and focus groups on: 'Machine Learning for Future Networks including 5G,' 'AI for Health' (ITU-WHO), 'Environmental Efficiency for AI and other Emerging Technologies,' 'Autonomous and Assisted Driving,' 'AI and Data Commons,' and 'AI for Good Repository.' All of the SDGs are addressed by the initiative as well as the following OECD principles: Inclusive growth, sustainable development and well-being, Human-centered values and fairness, Transparency and Explainability, Robustness, Security and Safety, and Accountability.
Maturity / Potential for adoption	The summit was launched in 2018 and is planned to take place annually in Geneva. Due to the pandemic, the summit is currently taking place online as a series of events throughout the year. The platforms are scalable and can be adopted by stakeholders worldwide for various purposes.

5. AI-Based Referral System

Initiative	AI-Based Referral System for Patients with Diabetic Retinopathy
Category	AI and Social Good
Brief Description	A diabetic retinopathy screening program for early detection and treatment through convolutional neural networks, based on Mexican clinical guidelines, that will be implemented in three hospitals in Mexico - for early detection and treatment of diabetic retinopathy. Healthcare is clearly one of the most dynamic and challenging sectors in Mexico and the LAC region. Nevertheless, the response to Diabetic Retinopathy (DR) faces three main problems: i) High prevalence of diabetes, the WHO reported that the prevalence of diabetes in Mexico is around 10.4% in 2016; ii) shortage of ophthalmologists, Mexico reports 42.5 ophthalmologists per millions of people (OPM), in contrast with other countries such as Spain with 105.5 OPM or Argentina 103.6 OPM, Brazil 67.4 OPM; and iii) lack of eye care services in primary health care.
Organization	Government of the State of Jalisco; Universidad Autónoma de Guadalajara; Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional; Centro Médico de Occidente

Geography	The initiative was initiated in Mexico and has national scope.
Sector	Public sector and academia
Key Success Factors	The key contributions to the success of the first stage were: i) a high-performance technical team; ii) strong infrastructure, supported by the Jalisco government who bought a dedicated AI-server to make training and validations with more than 90k retina fundus images; and iii) expert collaboration. Important metrics are: i) referral model with at least a sensitivity of 80% and specificity of 95%; ii) generation of high-quality datasets; iii) the need to generate a web application to integrate the AI models with the clinical flow; iv) implementation in three first level-hospitals a pilot of this project; and v) evaluation of the contributions of the early detection of DR that, until now, show their CNN models with a sensitivity of 80-89% and a specificity of 85-92% .
Key Hurdles	The pandemic has had strong implications on the initiative. Many departments and hospitals have changed the operations they used to do. On the other hand, the effort to prevent the spread of the virus has made patients affected by different diseases miss their medical appointments. Furthermore, a particular scientific challenge in this project is the understanding and convergence between different professional areas, i.e. physicians, engineers, medical practitioners, nurses, and others, in order to contribute with the best expertise in each area and to reach the main goal, which is, to reduce the blindness caused by retinopathy diabetic.
Potential to contribute to GPAI objectives	The initiative shows the positive impact AI can have in healthcare, a sector that is currently in severe crisis due to the Covid-19 pandemic. It also is an example of anapplie project adopting globally accepted AI ethics principles (OECD, UNESCO, European Union) to make intelligent systems transparent, explainable and safe for the operator and the user.
Diversity and Inclusiveness	There are 3 main teams: Administrative-organizational (AO), clinical and medical operations (C&MO), and scientific and technological (Sc&T). The AO leaders are 3 persons (2 M - 1 F). In the C&MO team are ophthalmologists and clinical researchers (3 F - 4 M). The Sc&T team has 4 males (M) experts in data science and deep learning from the public sector.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has impacted the scientific community and stakeholders - demonstrating that it is possible to create competitive referral models according to the Mexican DR guidelines with a combination of public datasets and local datasets. The initiative directly advances SDG 3 (good health and wellbeing) and the following OECD AI principles: Inclusive growth, sustainable development and well-being.
Maturity / Potential for adoption	The initiative was launched 2019 and is currently in the pilot phase.

6. AI Now Report

Initiative	AI Now Report 2018 (inc. Algorithmic Impact Assessment Framework)
Category	AI and Governance
Brief Description	The AI Now 2018 Report addresses key governance issues, including i) the growing accountability gap in AI, which favors those who create and deploy these technologies at the expense of those most affected; ii) the use of AI to maximize and amplify surveillance, iii) increasing government use of automated decision systems that directly impact individuals and communities without established accountability structures; iv) unregulated and unmonitored forms of AI experimentation on human populations; and v) the limits of technological solutions to problems of fairness, bias, and discrimination. It includes AI Now's algorithmic impact assessment framework which gives public sectors more tools for critically deciding if an algorithmic system is appropriate, and for ensuring more community input and oversight.
Organization	AI Now Institute
Geography	AI Now is based in New York. The research has an international scope.
Sector	Academia
Key Success Factors	The initiative's multi-stakeholder process resulted in a report with concrete and actionable recommendations for better governance mechanisms, including government regulations and corporate accountability structures that go beyond ethical guidelines. The algorithmic impact assessment is a concrete example.
Key Hurdles	The capacity of governments to operationalize the recommendations and/or implement the impact assessment framework is not specified.
Potential to contribute to GPAI objectives	The report aligns with GPAI's objectives providing tools to evaluate AI systems for responsibility and trustworthiness, based on metrics such as safety, robustness, accountability, transparency, fairness, respect for human rights, and the promotion of equity.
Diversity and Inclusiveness	The report and impact assessment framework are the products of multistakeholder consultations including representatives from various sectors and regions. The authors are 50% female and 50% male. All of the research is available for the public online.

Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The report achieves its research objectives and, furthermore, through its outreach strategy, raises awareness of accountability issues related to AI systems. It is aligned with the UN SDGs, having high impact on SDG 16 (Peace, Justice, and Strong Institutions) as well as SDG 3, SDG 5, SDG 8, and SDG 9.
Maturity / Potential for adoption	The report and impact assessment framework were published in 2018 and were well received by stakeholders in the Responsible AI ecosystem. The framework and toolkit can serve as building blocks to be leveraged as the ecosystem moves from principles to practice.

7. Algorithm Charter for Aotearoa New Zealand

Initiative	Algorithm Charter for Aotearoa New Zealand
Category	AI and Ethics; AI and Governance
Brief Description	The Algorithm Charter is a commitment by government agencies to improve consistency, transparency, and accountability in their use of algorithms. Signatories commit to a range of actions in the areas of transparency, partnership, focus on people, data, privacy, ethics, human rights, and oversight. The Charter follows a recommendation by the Government Chief Data Steward and Chief Digital Officer, who said that the safe and effective use of operational algorithms required greater consistency across Government. It was developed through consultation with the public and forms a part of the New Zealand Government's Open Government Partnership action plan. The Charter draws on the Principles for the Safe and Effective Use of Data and Analytics co-designed by the Government Chief Data Steward and the Privacy Commissioner.
Organization	New Zealand Government, Stats NZ
Geography	The initiative originated in New Zealand and has national scope.
Sector	Public sector
Key Success Factors	The report found that there are opportunities to increase collaboration and sharing of good practice across government to ensure that all of the information that is published explains, in clear and simple terms, how algorithms are informing decisions that affect people in significant ways. As agencies continue to develop new algorithms, it is important to preserve appropriate human oversight and ensure that the views of key stakeholders, notably the people who will receive or participate in services, are given appropriate consideration. The Charter is intended to be one part

	of the response to these findings and improve the overall transparency and accountability of government algorithm use, particularly where algorithms are being used in ways that could significantly impact people or groups.
Key Hurdles	Initially, the team proposed a strict definition of algorithms, but many submitters felt that a fixed definition could artificially constrain the work, or not work for some group of agencies. Within government, there's a tension between how to ensure that the Charter responds to the Algorithm Assessment report recommendations around improving consistency and transparency, without stifling innovation.
Potential to contribute to GPAI objectives	This Charter is one example of how a government can demonstrate transparency and accountability in the use of data. Ethics and algorithm transparency are global issues impacting both the public and the private sector. Trust in how governments use data will be essential to ensuring that governments have the necessary social capital to use new and innovative technologies to deliver services and support people. Its implementation helps operationalize principles behind the responsible development and deployment of AI.
Diversity and Inclusiveness	The System Policy team bring together experience from other government agencies (including Te Tāhuhu o Te Mātauranga — The Ministry of Education, Te Manatū Mō Te Taiao — The Ministry for the Environment and Te Puni Kōkiri — The Ministry of Māori Development) as well as working for local government, non-government organizations and the private sector. The team are currently all women (except the manager) and many of them were born in countries other than New Zealand. Public consultation on a draft Algorithm Charter ran from 17 October to 31 December 2019 and this was reported by domestic and international media. Beyond central government agencies, submissions were also solicited from a range of key stakeholders including academics, non-government organizations, civil society representatives, and regulators. Consultation on the charter was also promoted at the All-of-Government Innovation Showcase on 3 December 2019 and Internet NZ facilitated an online discussion session on Twitter on 20 November 2019. All information is available to the public online.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Twenty-six government agencies have signed the Charter so far. The initiative's success is measured, in the short-term, based on agencies making a commitment and then applying the risk matrix to their algorithms to determine the extent they need to apply the Charter commitments. The mid-term measure of success will be about agencies implementing these commitments and ideally applying them to all new algorithms they develop. The long-term aim is to earn increased public trust in government use of data. The initiative is not explicitly intended to address SDGs but, in applying it, New Zealand government agencies who are working toward SDGs may find additional alignment with these aims, particularly those around reducing inequalities and partnership. Furthermore, the New

	Zealand Government has introduced an all-of-government wellbeing budgeting approach in 2018 in response to calls from the OECD principles.
Maturity / Potential for adoption	The framework was published July 2020. After twelve months, a review of the Algorithm Charter will be conducted to ensure it is achieving its intended purpose of improving government transparency and accountability without stifling innovation or causing undue compliance burden. A review of progress is scheduled for 2021 to ensure the Charter is achieving its aims. It has potential to be replicated by other countries and contextualized. Solutions will need to respond to the social and cultural context in each nation. It is also essential to iterate and be prepared to be flexible.

8. Artificial Intelligence Against Modern Slavery (AIMS)

Initiative	Artificial Intelligence Against Modern Slavery (AIMS)
Category	AI and Social Good
Brief Description	Project AIMS uses AI to combat modern slavery. It creates the first AI tool for the scalable analysis of company statements on how they are eradicating slavery from their supply chains. The tool builds on the work of Walk Free, WikiRate and the Business & Human Rights Resource Centre (BHRRC) to speed up the statement review process and increase transparency for consumers and businesses.
Organization	Walk Free, The Future Society, Business Human Rights Resource Centre, WikiRate
Geography	The initiative originated in Australia and currently is piloted in Australia and the United Kingdom.
Sector	Civil society
Key Success Factors	Key success factors have included a promising prototype, strong partnerships and proven commitment from the partners, deep domain knowledge of AI and modern slavery, flexibility and scalability of the solution, and the adaptability of a growing team. Specific metrics for the tool are: 90% decrease in the time taken to assess a report, from one hour per volunteer per report, 90% accuracy rate of the tool from a 68% baseline set by the prototype.

Key Hurdles	Some of the challenges the initiative has encountered are unstructured data caused by the lack of machine readability format of the statements, shortage of talent with the appropriate skills, the lack of awareness about modern slavery and the importance to innovate against it (operational, social economic and scientific challenge).
Potential to contribute to GPAI objectives	A specific project using AI to address modern slavery, this initiative exemplifies an area where AI can make a big difference and requires cross-national and cross-sectorial collaboration. Furthermore, built taking into account ethical principles on AI, it shows how principles can be operationalized in practice to achieve impact.
Diversity and Inclusiveness	The team is global and gender balanced, supported by a multistakeholder and multidisciplinary advisory board. Project AIMS's AI is designed to benefit the most vulnerable, aiming to show the opportunities for inclusive growth, sustainable development and well-being. Its open source GitHub and publications (especially the handbook) will ensure that people understand the tools' outcomes and can challenge them.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	From its launch, the initiative was effective in advancing towards its objectives having built a prototype to analyze statements produced by businesses under the UK Modern Slavery Act (MSA), and the Australian Modern Slavery Act. Directly, the initiative addresses SDG 8 (Decent work & economic growth). It respects all OECD principles, including Inclusive growth, Sustainable development and well-being, Human-centered values and fairness, Transparency and explainability, Robustness, Security and safety, and Accountability.
Maturity / Potential for adoption	The initiative launched in June 2020 and is currently in the prototype phase in Australia and the United Kingdom. It has high potential to be scaled to other nations as well as on a global level.

9. Artificial Intelligence and Blockchain for Healthcare Initiative in Africa

Initiative	Artificial Intelligence and Blockchain for Healthcare Initiative in Africa
Category	AI and Social Good
Brief Description	An initiative accelerating drug discovery and drug development by continuously inventing and deploying AI technologies. The leading short to long-term applications of AI in pharma is more towards reducing the time and hence the cost of drug development. This would not only enhance the return on investment and reduce the costs for users but would be helpful in making useful products available faster, especially where it matters most. With the aid of advances in tech, especially AI, scientists and developers in Africa can be more productive and innovative towards achieving better drug discovery

	outcomes. This would likely transform pharma and healthcare in the region and globally.
Organization	Insilico Medicine
Geography	The initiative's scope is in Africa. Insilico Medicine is based in Hong Kong and has global reach.
Sector	Private sector
Key Success Factors	Key success factors have included processing of large clinical and medical data. The company and its scientists are dedicated to extending human productive longevity and transforming every step of the drug discovery and drug development process through excellence in biomarker discovery, drug development, digital medicine, and aging research.
Key Hurdles	Challenges include developing interest and talent in the region and educate the local scientists on the value and the possible uses of their data
Potential to contribute to GPAI objectives	The initiative illustrates how AI can be used to accelerate drug discovery, a priority area identified by the GPAI RAI working group in which international cooperation and collaboration can have strong impact.
Diversity and Inclusiveness	Their team includes over 120 scientists (structural biologists, medical chemists, and machine learning experts), with about 70% appointed through hackathons and worldwide competitions in Rockville, Oxford, Warsaw, Brussels, Shanghai, Taipei, and Hong Kong.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The company suggests that its AI solutions have the potential to reduce costs in medical treatments by half across the board within the next couple of years. It can serve as a growth engine for African biopharmaceutical industry and transform the R&D in local pharmaceutical companies. The initiative directly impacts SDG 3 (Good health and wellbeing).
Maturity / Potential for adoption	The company was founded in 2014 and the initiative in Africa was launched in 2018. With Covid-19, the company has seen tremendous growth.

10. Artificial Intelligence Standardization White Paper

Initiative	Artificial Intelligence Standardization White Paper
Category	AI and Ethics; AI and Governance
Brief Description	The paper describes China's approach to standards-setting for artificial intelligence. The white paper recommended that "China should strengthen international cooperation and promote the formulation of a set of universal regulatory principles and standards to ensure the safety of artificial intelligence technology." This recommendation was corroborated by previous CESI policies, e.g., its 2017 Memorandum of Understanding with the IEEE Standards Association to promote international standardization.
Organization	China Electronics Standardization Institute (CESI) within the Ministry of Industry and Information Technology
Geography	Initiative published in China with international scope.
Sector	Public sector

Key Success Factors	The initiative includes all standardization protocols and applications examples of AI by China's leading tech companies. The authors of the initiative dedicate several sections to contextualizing their work historically. They also explain the technological, economic, commercial, and international contexts.
Key Hurdles	One of the hurdles has been the integration of the efforts of all stakeholders and of the four tasks of nurturing commercial applications, enabling breakthroughs, supporting fundamental research and deepening smart manufacturing.
Potential to contribute to GPAI objectives	The initiative includes standardization protocols and application examples of AI by China's leading tech companies. It reflects the perspectives of China in what standards can look like for the responsible development and deployment of AI systems. It highlights the role of standards in promoting and fostering the responsible development and deployment of AI.
Diversity and Inclusiveness	One of the four leaders of the four institutions involved in the launch of the report is female. The white paper was published in Chinese (it has been translated in English by the US-based Center for Security and Emerging Technology). The host website is available in Chinese. There were more than 400 people including committee members, experts, scholars and representatives from related standardization technical committees, universities, research organizations, and companies attending the launch event.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	It directly addresses SDG 3, SDG 9, and SDG 16 and several of the OECD principles are mentioned as values in the report. The initiative was launched by four institutions and is part of the Chinese government's State Council's plan for the development of next generation AI. As it falls within a broader plan, the purpose and impact pathway of the initiative is clearly defined and scoped.
Maturity / Potential for adoption	Initiative was published in Chinese in January 2018 and translated in May 2020. It will be revised constantly in the future based on the developing requirements of technologies, industries, and standardization.

11. Asilomar AI Principles

Initiative	Asilomar AI Principles
Category	AI and Ethics
Brief Description	Asilomar AI Principles are 23 guidelines for the research and development of artificial intelligence (AI). The Asilomar principles outline AI developmental issues, ethics and guidelines for the development of beneficial AI and to make beneficial AI development easier. The tenets were created at the Asilomar Conference on Beneficial AI in 2017 in Pacific Grove, California. The conference was organized by the Future of Life Institute. The Asilomar AI Principles are subdivided into 3 categories: Research, Ethics and Values and Longer-Term Issues. Often, the

	principles are a clear statement of possible undesirable outcomes, followed by a recommendation to prevent such an event.
Organization	Future of Life Institute
Geography	The initiative originated in the US but has global reach, including stakeholders from all around the world.
Sector	Civil society
Key Success Factors	Future success of the initiative will be driven in large part by the "snowball effect": a large set of noteworthy endorsers helps drive wider adoption. The principles were also crafted to be both relatively non-controversial (while still having "teeth" in the sense that they will by no means be fulfilled by default). While other sets of principles have endorsement at the national level, probably no set of principles has endorsement by as wide a range of individual high-profile stakeholders.
Key Hurdles	The Asilomar Principles are unusual or even unique in addressing some of the longer-term issues in AI governance, which are left aside or even actively avoided by some other sets. The dynamics leading to this exclusion are complex, but their net effect is that some extremely crucial — but longer horizon — issues are left for indefinite "future consideration." Another obstacle is that they take an explicit position on militarization of AI (against an arms race in AI weapons). This has raised opposition from some governments.
Potential to contribute to GPAI objectives	Initiative exhibits exhibit widespread consensus on issues related to the responsible development and deployment of AI, even if certain vested interests are in opposition. Stakeholders who have endorsed the principles are from several sectors and hence shows how stakeholders from different backgrounds can come to consensus.
Diversity and Inclusiveness	Core team is strongly scientifically- and academically based rather than from a corporate or policy background. The principles are available online and accessible in six languages: Chinese, German, Japanese Korean, Russian, and English.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	This initiative includes as a first stage the creation of the Asilomar AI principles with a large number of high-level (primarily individual) signatories, their announcement and publicity. A second, ongoing, stage is to push for the adoption of the Asilomar principles — either as a whole or in part — by many other organizations, and for their influence on or inclusion in other AI initiatives. The largest success of this phase was the endorsement of the Asilomar AI principles by the California State Government. The initiative affects SDG1 (No poverty), SDG 2 (Zero hunger), SDG 8 (Decent work & economic growth), SDG 10 (Reduced Inequalities) and SDG 16 (Peace, justice & strong institutions). It helps

	implement the following OECD principles: Inclusive growth, sustainable development and well-being, Human-centered values and fairness, Transparency and explainability, Robustness, security and safety, and Accountability.
Maturity / Potential for adoption	Principles were launched January 2017 and have since been signed by 1677 AI/Robotics researchers and 3662 other members of the field. In its second stage, the initiative continues to draw attention and gain new stakeholder endorsement. There is also potential to pilot these principles to see the extent to which they are implemented.

12. Assessment List for Trustworthy Artificial Intelligence (ALTAI)

Initiative	Assessment List for Trustworthy Artificial Intelligence (ALTAI)
Category	AI and Ethics; AI and Governance
Brief Description	A practical tool that helps business and organizations to self-assess the trustworthiness of their AI systems under development. The initiative follows the High-Level Expert Group on AI's publication: Ethics Guidelines for Trustworthy AI. which proposes seven requirements that AI systems should meet in order to be deemed trustworthy. The initiative's mission is 'to guide the development and application of AI in a human-centered approach and to be trustworthy.'
Organization	European Commission High Level Expert Group on AI
Geography	The initiative originated in the EU and its reach is global.
Sector	International organization, including experts from multiple sectors
Key Success Factors	Key success factors have been engagement with potential users to the tool to ensure the 'checklist' is beneficial and practical for assessing the trustworthiness of AI. Furthermore, after users apply the tool, it provides them with a visualization of the self-assessed level of adherence of the AI system and its use with the 7 requirements for Trustworthy AI, as well as recommendations based on the answers to particular questions. Success is measured based on the number of organizations that will use the tool as well as the impact it has on promoting the responsible development of AI.
Key Hurdles	The impact ALTAI has had after the revised version was launched July 2020 is not clear. Since it is not sector-specific, the tool is to be used in a flexible manner, meaning organizations can focus on some elements more than others depending on their particular industry or sector.

Potential to contribute to GPAI objectives	ALTAI is an example of a 'soft governance' mechanism that business and organizations can use to self-assess the trustworthiness of their AI systems. Its own mission aligns with GPAI objectives to further strengthen the benefits that AI yields to the economy and society as a whole. While the ALTAI is voluntary, it is an important step on the path to formal regulation of AI, as it enables companies to signal compliance with it, and thus foster consumer trust.
Diversity and Inclusiveness	Team composed of multidisciplinary AI experts, most of which are from Europe. The revised version was a result of collected feedback via three channels: an online survey filled in by registered participants, the European AI alliance sharing best practices, and a series of in-depth interviews. The tool is available for the public.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has been effective in reaching its objectives, with over 350 stakeholders having already piloted the checklist to self-assess the trustworthiness of their AI systems. Furthermore, the initiative is aligned with SDG 3, SDG 5, SDG 9, SDG 10, SDG 13, SDG 16, and SDG 17. It also respects all of the OECD principles.
Maturity / Potential for adoption	The tool was first presented with the Ethics Guidelines in June 2019 and was revised July 2020 following a piloting process that involved more than 350 stakeholders. It has high potential to be implemented by business and organizations, and also be scaled globally.

13. CDEI Review of Online Targeting

Initiative	CDEI Review of online targeting
Category	AI and Governance
Brief Description	A review of online targeting in the UK, proposing three sets of recommendations that relate to increased accountability, transparency and user empowerment with the aim of helping to build public trust and ensuring society and the economy benefit from online targeting.
Organization	Centre for Data Ethics and Innovation
Geography	The initiative was published in the UK and has national scope.
Sector	Public sector

Key Success Factors	Some key success factors include a robust multi-stakeholder approach based largely on the perspectives of the UK public and actionable recommendations that are contextualized for the UK landscape.
Key Hurdles	A challenge in the multi-stakeholder process was building a basic level of understanding in the public in order to engage citizens and gather their perspectives on the matter.
Potential to contribute to GPAI objectives	Given concerns over social media growing worldwide, this initiative shows a specific area where AI’s development and deployment has significant impact as well as socio-ethical implications. Amongst its recommendations for the UK government, it calls for coherence and coordination across the current and future regulatory landscape. Likewise, GPAI seeks to establish mechanisms for cross-national and cross-sectorial international coordination and cooperation to solve global challenges, such as online harms.
Diversity and Inclusiveness	The report was produced through a multi-stakeholder process where diverse sets of individuals were consulted widely in the UK, inclusive of academia, civil society, regulators and the government. They also held interviews with and received evidence from a range of online platforms in addition to advertising companies and industry bodies. Furthermore, CDEI commissioned Ipsos MORI to deliver qualitative and quantitative analysis of public attitudes on online targeting. Ipsos MORI engaged 147 participants aged 16+ in two days of discussion across seven locations in Great Britain over June-July 2019.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The report has been effective in increasing awareness of the issue of online targeting and driving. The conversation and policy thinking around this. The initiative is closely aligned to SDG 16 (Peace, Justice, and Strong Institutions) as well as SDG 3 (Good Health and Wellbeing). In fact, in the public survey, mental health was cited as one of the key concerns related to online targeting.
Maturity / Potential for adoption	The review was published February 2020 in the UK and, according to the institutional design of the UK, the government has to respond to the recommendations made by the independent advisors. Some of the key issues identified in the report transcend national borders and, hence, CDEI is exploring international collaboration with other nations and/or global forums.

14. CEPEJ Ethical Charter on the Use of AI in Judicial Systems and their Environment

Initiative	CEPEJ European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment
Category	AI and Ethics
Brief Description	The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe has adopted the first European text setting out ethical principles relating to the use of artificial intelligence (AI) in judicial systems. The Charter provides a framework of principles that can guide policy makers, legislators and justice professionals when they grapple with the rapid development of AI in national judicial processes. The initiative's mission is: 'to ensure that AI remains a tool in the service of the general interest and that its use respects individual rights.'
Organization	Council of Europe: European Commission for the efficiency of Justice (CEPEJ)
Geography	The initiative is cross-regional, with primary scope the CoE member states most of which are countries in Eurasia but also includes observer states in North America and Asia.
Sector	International Organization
Key Success Factors	Key success factors include the fact that the charter was the first European text dealing with AI and ethics in the judiciary. Also, its targeted focus on human rights and the rule of law as well as deep domain expertise with input from AI experts across the CoE member states were critical to its success in designing principles that reflect both fundamental values and essential methodological requirements for the creation and development of algorithms.
Key Hurdles	Key challenge is that the set of principles are not binding and, at this stage, the mechanism to monitor how they are being implemented is not specified. CEPEJ is now conducting studies on operationalization of the Charter and potential certification mechanisms.
Potential to contribute to GPAI objectives	The charter highlights how the application of AI in the field of justice can contribute to improve efficiency and quality. Specifically, it calls for AI to be implemented in a responsible manner which complies with the fundamental rights guaranteed in particular in the European Convention on Human Rights (ECHR) and the Council of Europe Convention on the Protection of Personal Data. The clear link to fundamental rights closely aligns with the GPAI mandate to contribute to the promotion of human rights.

Diversity and Inclusiveness	The core team involved in the evidence base for the charter was composed of experts from all the 47 member states of the Council of Europe, both justice professionals called to use AI solutions in their daily practice and AI experts. The Charter is intended for multiple stakeholders, including public and private stakeholders responsible for the design and deployment of AI tools and services that involve the processing of judicial decisions and data (machine learning or any other methods deriving from data science).
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The charter has been effective in building consensus over the key principles that should be implemented for the development of AI in judicial systems. The initiative directly affects SDG 16 (Peace, Justice, and Strong Institutions) and also respects OECD AI principles.
Maturity / Potential for adoption	The charter was adopted in Strasbourg December 2018 and subsequently largely disseminated within CoE MS and beyond. The CoE is now working on possible operationalization of the Charter principles and potential certification mechanisms

15. Draft AI R&D Guidelines for International Discussions

Initiative	Draft AI R&D Guidelines for International Discussions
Category	AI and Ethics
Brief Description	The DAI R&D Guidelines for International Discussions and AI Utilization Guidelines were prepared to protect users' interests, prevent spread of risks, and realize a human-centered AI society by promoting the benefits of AI systems and controlling the risks through the sound progress of AI networking, and they are intended for AI developers and users, respectively. They collect the principles and explanations regarding the elements to which developers and users, respectively, are expected to pay attention. They were elaborated as proposed guiding principles to serve as draft non-regulatory and non-binding soft laws to be shared and discussed internationally.
Organization	Institute for Information and Communications Policy (IICP), The Conference toward AI Network Society
Geography	Draft was created in Japan, but the target is an international reach.
Sector	Public Sector
Key Success Factors	Success factors included: i) the fact that positive cooperation across government agencies including the Cabinet Office supported integration

	work in finalizing the draft guidelines; and ii) strong collaboration with OECD to leverage the guidelines towards consensus.
Key Hurdles	Challenges include concerns that the guidelines may be imposing an excessive burden on developers in R&D work by enforcing additional measures and costs to businesses and may hinder open and enabling environment for the innovation and progresses in AI development and utilization. However, such concerns were dispelled by thoroughly disseminating and enlightening the guidelines as non-regulatory and non-binding soft laws rather than hard laws. Furthermore, the initiative found gaps between developers and users in their understandings on various basic concepts (such as bias). With different cultures, socio-economic structures, legal systems, etc., it was extremely difficult to reach international consensus in the course of the international discussion.
Potential to contribute to GPAI objectives	This initiative serves as an example of a set of principles with international scope promoting and fostering the responsible development of AI. It is important to note that for the AI principles to be accepted by society, stakeholders should recognize various values within societies. It is also important to address people's anxieties about AI. To do so, this initiative focused on specific use cases, such as transfer (self-driving), health (medical/nursing care), and finance case examples.
Diversity and Inclusiveness	These guidelines were prepared by the Conference toward AI Network Society, with multiple stakeholders' participation. Specifically, the Conference consists of social and humanities researchers in law, economics, sociology, etc., technical researchers, including former presidents of the Japanese Society for Artificial Intelligence, and businesses developing and utilizing AI, including start-ups and foreign-affiliated companies, as well as representatives of consumer groups. Furthermore, many of these members have participated in international meetings, such as OECD and G7 expert meetings, and contributed to international discussions. In addition to the members, the Conference is conducting hearings from businesses and experts.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Over the last few years, Draft AI R&D Guidelines for International Discussions have been input and fully reflected in international discussion, including at G7, G20, and OECD. Specifically, the OECD recommendation was formulated around the core concept of human-centeredness and incorporates the principles of fairness, transparency, explainability, security, safety, and accountability. The initiative addresses SDG 5 (Gender equality), SDG 9 (Industry, Innovation & Infrastructure), SDG 10 (Reduced Inequalities), SDG 12 (Responsible consumption & production), SDG 16 (Peace, justice & strong institutions), and SDG 17 (Partnerships for the goals).

Maturity / Potential for adoption	The draft was published in July 2017. International fora including at G7 and OECD, positively accepted Japan's proposal to use the draft guidelines via the multi-stakeholder discussions as the basis of international discussion and the draft guidelines gained broad support. This led to making input to the G7 outcome document and serving as the basis of the OECD recommendation (May 2019). MIC continues to promote its AI governance initiative by collecting, accumulating, and disseminating case examples of developers and users formulating their guidelines based on these guidelines. The initiative will conduct hearings from developers and business users on their practices based on those guidelines and will compile collections of good practices on AI governance or AI utilization, and disseminate them for awareness raising. MIC will also lead AI developers and business users to voluntarily elaborate their own guidelines by making reference to the collected information and guidelines.

16. Elements of AI

Initiative	Elements of AI
Category	AI and Social Good
Brief Description	The Elements of AI is a series of free online courses created by Reaktor and the University of Helsinki. Their aim is to encourage as broad a group of people as possible to learn what AI is, what can (and can't) be done with AI, and how to start creating AI methods. The courses combine theory with practical exercises and can be completed at the user's own pace. It explains the implications of Artificial Intelligence (AI) in real everyday situations with interactive exercises, so that students can make informed decisions as workers, as voters, and as media and product consumers.
Organization	Reaktor and the University of Helsinki
Geography	The initiative originated in Finland and now is deployed globally.
Sector	Private sector and Academia
Key Success Factors	The initiative is an example of a successful public-private partnership between the University of Helsinki and Reaktor. The AI Challenge campaign with the original goal of training 1% of the Finnish population (which was later expanded to 1% of the world's population) has enabled hundreds of companies and other organizations to join the initiative by pledging to train their employees.

Key Hurdles	A key challenge for the initiative has been inventing ways to make education attractive and accessible by the general public while being able to explain the basics of a complex topic such as AI has required intensive work by a diverse team of academics, professionals, educators, and designers.
Potential to contribute to GPAI objectives	This initiative gives the public the ability to participate as informed citizens in global dialogue and decision-making around AI. It has elevated the level of AI-related discussions beyond myths towards more science-based, rational discussions. This is fundamental in order to support democratic policy dialogue where every voice is heard.
Diversity and Inclusiveness	The core team of some 20 to 30 people consists of academics and software engineers from the University of Helsinki, and designers, data scientists, marketing professionals, etc., from the software consultancy Reaktor. The team is highly diverse with roughly 50:50 gender-balance, education ranging from visual designers to lawyers and from professors to students. There are currently over 545,000 signups from over 170 countries. Nearly 40% of participants are women, and over 25% are over the age of 45.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The Elements of AI has been ranked the best online course in computer science by Class Central. It has been awarded the MIT Inclusive Innovation Challenge Grand Prize, the Nokia Foundation Recognition Award, and numerous other recognitions. The French President Emmanuel Macron, Google CEO Sundar Pichai, and others have praised the course. It directly affects SDG 4 (Quality education), SDG 5 (Gender equality), SDG 8 (Decent work & economic growth), SDG 9 (Industry, Innovation & Infrastructure, SDG 10 (Reduced Inequalities), SDG 12 (Responsible consumption & production), SDG 16 (Peace, justice & strong institutions) and SDG 17 (Partnerships for the goals).
Maturity / Potential for adoption	The initiative launched spring 2018 and with the support of the European Commission and the Finnish EU Presidency, the course is being translated in all EU languages and launched in all EU member states. The initiative relies on a model building partnership with local organizations in each country to reach as wide audiences as possible, including the hard-to-reach audiences that are typically left out of technology-related discussions.

17. Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)

Initiative	Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)
Category	AI and Governance
Brief Description	The initiative develops comprehensive suites of objective and verifiable criteria for ethical Transparency, Accountability, Reduction in Algorithmic Bias and Privacy in products, services and systems. So far, it has developed large suites of (roughly 200) criteria for each dimension cited with the exception of Ethical Privacy that's currently under development. The scope of work is generic and universal in that the criteria can be applied to any product/service/system to identify the strength and shortfalls in so far as ethicality is concerned. It also has provisions for customization towards specific priorities, idiosyncrasies and requirements of a given application, industry, discipline or sector.
Organization	IEEE SA
Geography	Initiative is cross-regional and has global scope.
Sector	Civil Society, Private Sector, and Academia
Key Success Factors	The initiative has adopted a model-based approach to knowledge elicitation, capture and representation that has assisted with much shorter development time scales as well as significant creative components driven by the adopted methodology. The development of three suites of ethical Transparency, Accountability and Algorithmic Bias has resulted in large and tiered criteria that are also qualified against risk of a product/service or system thus rendering a fair, efficient and value sensitive approach to conformity assessment and potential certification. In view of the model-based nature of the underlying concepts, the existing suites are being expediently redeployed for generating new variants and tailored sets for context specific applications that also underpin their success due to the agile and responsive adaptation for expedient deployment.
Key Hurdles	Key challenges include having access to a globally representative panel of diverse experts in the development of the criteria.

Potential to contribute to GPAI objectives	Developing metrics and processes towards the implementation of a certification methodology addressing transparency, accountability and algorithmic bias - the initiative is clearly aligned with the GPAI mandate. The development and readiness of the global ecosystem for ethical assurance of product/services or systems is a key success indicator. Organizations are ready to provide evidence for the due diligence they are doing to ensure the trustworthiness of the A/IS products and services they build. However, due to the novelty, the ecosystem is in its formative stages and the ECPAIS programme is going towards a collaborative and supportive model for its ecosystem comprising the Design Authority, Accredited Partners, Certification Bodies + the global marketplace for the developers and consumers of ethical products and services. Much effort is being put into the ecosystem development at the moment with a number of global partners. A comprehensive educational programme is also under development in support of the ecosystem and training of the partners and the support services. Furthermore, it is imperative that generic and tailored suites of ethical values and criteria are varied and enriched across cultural, historic, belief and value systems globally. The initiative's model-based approach facilitates rapid modification, adaptation and enrichment of the original criteria that are being kept under a systematic Change Control and Configuration Management regime.
Diversity and Inclusiveness	Each suite of ethical criteria is made up of a diverse team, largely gender balanced with complementary backgrounds and expertise such as AI technologies, research, law, engineering, manufacturing and public services. The current panel is more biased towards majority women participants.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has shown progress towards its objectives identifying suits of ethical criteria and pilot projects to implement them. To test the viability and workability, the initiative has conducted six use case deployments. In early 2020, the initiative has embarked on pilot studies and deployment cases for the criteria in a number of areas. One key initiative that spontaneously started in May 2020 was to adopt processes and utilize capabilities towards developing ethical assurance criteria for Contact Tracing Applications and Technologies. This was initiated as an altruistic response to the current global pandemic and the poor adoption by the citizens almost universally. Overall, the initiative is aligned with SDG3, SDG 9, SDG 10, SDG 11, SDG 12, SDG 16 and SDG 17. It also respects OECD AI Principles.
Maturity / Potential for adoption	The initiative was launched in October 2018. A number of pilot projects are underway during 2020 to implement the criteria in a few real-world contexts and verify the applicability and practicality of the criteria and the tailoring process.

18. Global Governance of AI Roundtable

Initiative	Global Governance of AI Roundtable
Category	AI and Governance
Brief Description	Held yearly in Dubai on the occasion of the World Government Summit (WGS) under the aegis of the UAE State Minister for AI, the Global Governance of AI Roundtable (GGAR) is a revolving international multi-stakeholder governance process that brings together a diverse community of 250 global experts and practitioners from government, business, academia, international organizations, and civil society. GGAR has been envisioned and designed as a unique collective intelligence exercise to help shape and deploy global, but culturally adaptable, norms for the governance of AI. It has no panels, no keynotes; only curated breakout sessions to maximize productivity and outcome. The insights and recommendations have been captured into a comprehensive report, which includes an action-oriented summary for policymakers. The Global Governance of AI Roundtable has three chief objectives: i) Gathering information about the state of AI technologies, their socioeconomic impact, and the state of AI governance practices and policies around the world; ii)Synthesizing that information into a governance framework, actionable public policy options, and implementation-level guidelines that can be implemented by the UAE and other governments around the world; and iii) Serving as the world's authoritative forum for AI governance.
Organization	World Government Summit
Geography	The initiative was launched by UAE but gathers global stakeholders.
Sector	Public sector, private sector, academia, and international organizations
Key Success Factors	Success factors included an International multi-stakeholder governance process that brings together a diverse community of 250 global experts and practitioners from government, business, academia, international organizations, and civil society. Some key institutes include OECD, UNESCO, IEEE, CXI and TFS. Success is measured by the synergies developed across the landscape of AI governance globally and the new partnerships and the greater inclusiveness of the conversation (in terms of gender balance and geographic representation).
Key Hurdles	Challenges include difficulty getting developers interested to contribute to the policy debate, difficulty to get stakeholders from different backgrounds to start from common terminology and agreeing upon definition, and the high budgetary requirements to sustain the initiative. Furthermore, travel restrictions due to Covid-19 have shifted plans for a 2020 event.
Potential to contribute to GPAI objectives	The initiative is an example of multistakeholder cross-national dialogue necessary to set the pathways for the responsible development and deployment of AI. As a result of such fora, synergies are built, and the ecosystem can work together to address shared challenges.
Diversity and Inclusiveness	The team which co-produced the event was made up of 4 people. 3 women, 1 man; from France, India, UK/Singapore and Belgium; with background in civil society, think tank, private sector and public sector (international affairs). Beyond the co-producing team, there were ~20 volunteers from all over the

	world (North America, Middle East, Latin America, Asia, Europe, etc.) as well as two directors (from Middle East and Europe). Initiative partners with a host of prestigious international organizations including the OECD, UNESCO, IEEE, the Council on Extended Intelligence, and the Global Data Commons Task Force. After providing each partner-organization with a platform to meet and advance its own goals and initiatives on AI policy during two days ahead of the World Government Summit (WGS), the Global Governance of AI Roundtable culminated into a one-day big Roundtable Collective intelligence Workshop held on the first day of Summit.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Thus far, the initiative successfully produced 2 conferences, bringing over 200 experts in total to discuss 14 different topics, from the design of GPAI to the potential of AI for developing countries. The 2019 gathering led to the publication of 14 background research papers on different topics ranging from agile governance, cybersecurity, geopolitics, explainability, international development, sustainability, and more. The initiative addresses SDG 17 directly building partnerships amongst stakeholders in the ecosystem. There is also a dedicated track exploring AI's impact on all seventeen SDGs.
Maturity / Potential for adoption	Building upon the first edition held in February 2018, the 2019 edition began in August with an intensive six-months preparation and curation period. The initiative is planned to take place annually in the UAE, bringing together a diverse set of stakeholders from around the world. This year's annual event has been postponed due to Covid-19.

19. HumanE AI Net

Initiative	HumanE AI Net
Category	AI and Ethics; AI and Social Good
Brief Description	An inter-disciplinary EU research initiative specifically aimed at technical/methodological breakthroughs to operationalize the full spectrum of OECD and European AI principles. It leverages the synergies between the involved centers of excellence to develop the scientific foundations and technological breakthroughs needed to shape the AI revolution in a direction that is beneficial to humans both individually and societally, and that adheres to European ethical values and social, cultural, legal, and political norms. The aim is to facilitate AI systems that enhance human capabilities and empower individuals and society as a whole while respecting human autonomy and self-determination.
Organization	European Commission with a network of 53 academic and industrial partners across Europe
Geography	Initiative originated in Germany and has a European scope. The impact of the research should have global reach.
Sector	International organization, academia, and private sector
Key Success Factors	Key success factors include high seed resources, broad geographic participation within the EU, clear mandate & objectives for the project, clear scope and drive to cooperate across nations and sectors and focus on interdisciplinarity.
Key Hurdles	Challenges are not specified.

Potential to contribute to GPAI objectives	The initiative exemplifies an interdisciplinary research project aiming to operationalize AI principles, including autonomy and self-determination, to ultimately enhance human capabilities. It is cross-sectoral, illustrating strong collaboration between industry, academia and support from public sector/government. The HumanE AI Net project will engender the mobilization of a research landscape far beyond direct project funding, involve and engage European industry, reach out to relevant social stakeholders, and create a unique innovation ecosystem that provides a manyfold return on investment for the European economy and society.
Diversity and Inclusiveness	The network includes more than 50 research centers in Europe and industry. Key deliverables of the projects (Ethics framework, research roadmap, policy guidelines, community, funding mechanisms) can be reused, promoted and shared to a broader group of stakeholders.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Current work packages include: WP1 - Learning, Reasoning and Planning with Human in the Loop; WP2 - Multi Modal Perception and Modeling; WP3 - Human AI Interaction and Collaboration; WP4 - Societal AI; WP5 - AI Ethics and Responsible AI; WP6 - Applied research with industrial and societal use cases; WP7 - Innovation Ecosystem and Socio-Economic Impact; WP8 - Virtual Center of Excellence, Capacity building and Dissemination; and WP9 - Synergies with AI on demand platform(s) and the Broader European AI Community. The initiative aims to operationalize the full spectrum of OECD and European AI principles.
Maturity / Potential for adoption	HumaneAI was launched in 2018 and, recently, the HumaneAI project has been successfully extended into the HumanE-AI-Net under the H2020 call topic ICT-48-2020 – Towards a vibrant European network of AI excellence centres which now gives the consortium 3 years to continue its work.

20. iGamma

Initiative	iGamma
Category	AI and Social Good
Brief Description	An AI system to assess the condition of an ecosystem and its benefits. The initiative applies the Ecosystem Integrity Concept which, like human health diagnosis, informs a latent variable through measurable attributes. It has successfully processed data under a unified computational framework based on Bayesian networks, to estimate the condition of terrestrial ecosystems for multiple timesteps, and the crisscross relations of variables that deliver ecosystem services. It is also producing information services (dashboards, reports, and infographics) and disseminates them.
Organization	Instituto Nacional de Ecología
Geography	Initiative originated in Mexico and has national scope.
Sector	Academia
Key Success Factors	Ecosystems are highly valuable sources of goods and services and a heritage for future generations. Assessing their condition is important for all management and conservation activities and to inform public policies. The initiative has found a revealing indicator of success while providing ecosystem

	condition cartography for Mexican scientists and decision makers. Indeed, there are many open problems in ecology that prevent the utilization of concepts like ecosystem integrity in decision making, even though they are already adopted in legislation. The use of AI is useful to overcome this, like the use of Bayesian network models as operational implementation of ecosystem integrity to enable quantification of ecosystem conditions, and its use to inform and evaluate the impact of public policy decisions on nature assets. The initiative uses AI to support functional adoption of the concept in formal environmental decision making.
Key Hurdles	Challenges include keeping government officials interested in the initiative and the potential uses of ecosystem integrity assessment and monitoring, despite administrative turn-over. The initiative also noted the need to further promote the use of AI to address environmental problems to increase the investment in economic and human resources for further exploring the use of this approach.
Potential to contribute to GPAI objectives	iGamma is an applied project using AI to address issues around biodiversity, a key challenge that international collaboration and cooperation can yield economic and environmental benefits for.
Diversity and Inclusiveness	The team is multidisciplinary, and gender balanced. It includes biologists, ecologists, geographers, data scientists, physicists, political scientists, and anthropologists in the team. It also includes the perspectives of government officials, experts on implementing environmental regulations and members of the public interested in nature conservation (including children). Partners include the Government authorities of three States, and three at the Federal level: INEGI, CONAFOR, and CONANP.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Initiative has been effective in achieving key milestones, including analyzing current operational biodiversity data acquisition systems in Mexico, fostering the development of new monitoring programs, producing maps on the integrity of terrestrial ecosystems in Mexico, contributing to the UN-SEEA pilot on new environmental accounting, and developing new dashboards on environmental data for the government of the State of Guanajuato. The initiative achieves progress towards SDG 3 (Good health & well-being), SDG 11 (Sustainable cities & communities), SDG 13 (Climate action), and SDG 15 (Life on land). It aligns with the following OECD AI principles: Inclusive growth, sustainable development and well-being, Human-centered values and fairness, Transparency and explainability, and Accountability.
Maturity / Potential for adoption	The initiative was launched in 2018 and is currently in the pilot phase. Partnering with authorities is fundamental to ensure the project is relevant and will be used for public policy making and dialoguing with NGO ensures various perspectives are considered.

21. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Initiative	IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Category	AI and Ethics; AI and Governance
Brief Description	The mission of the IEEE Global Initiative on Ethics of A/IS mission is to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity. It includes the Ethically Aligned Design, First Edition - a comprehensive report that combines a conceptual framework addressing universal human values, data agency, and technical dependability with a set of principles to guide A/IS creators and users through a comprehensive set of recommendations. EAD inspired the IEEE P7000 series: a series of standards projects that address specific issues at the intersection of technological and ethical considerations.
Organization	IEEE SA
Geography	Cross-regional initiative with a global scope
Sector	Mixed, including stakeholders from civil society, academia, the private sector and the public sector
Key Success Factors	During the open consultation period for the first two versions of Ethically Aligned Design, the initiative received over 500 pages of feedback. This garnered significant interest in participating in the IEEE work from various stakeholders from around the globe, resulting in over 500 additional members engaging in various activities within the IEEE Global Initiative community. The second iteration of EAD was utilized by the OECD to create the OECD Principles on Artificial Intelligence. It also informed the work of the Future of Life Institute, the EU High Level Experts Group, and multiple companies, including IBM. In aggregate, all three versions of EAD have been mentioned in more than three dozen global policy documents, highly cited academic journals and articles, and in the media.
Key Hurdles	A key challenge for any AIS-related work is in recognizing that it is primarily human data that drives algorithmic systems. A foundational aspect to the initiative is in creating a mental model for sovereign data honoring the need for all people to be able to access and share their data in parity with how it is already tracked.
Potential to contribute to GPAI objectives	The initiative promotes the responsible development and deployment of AI, shrinking the gap from principles to practice. Involving several stakeholders from the international arena, the initiative has global reach and is multidisciplinary. The work of the Global Initiative and the IEEE 7000 series of standards inspired the IEEE SA to lead the formation of the Open

	Community for Ethics in Autonomous and Intelligent Systems (OCEANIS), a global forum that brings together organizations interested in the development and use of standards as a means to address ethical matters in autonomous and intelligent systems. IEEE, as an organization, is uniquely positioned to help advance industry and the community forward.
Diversity and Inclusiveness	The team is about 50% male / female and includes experts from academia, engineering and business. They are working to increase global diversity as currently most members are from the US, EU and UK. EAD was prepared using an open, collaborative, and consensus building approach. Outputs are open-source and available online to the public. Currently, the output is available in English, Chinese, and Arabic.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	IEEE was the first in the larger socio-technical arena and as an organization of our size to boldly declare the need to prioritize applied ethical decisions at the outset of design. Organizations like the OECD, UNESCO, UNICEF, the EU High Level Experts Group, and policy makers from the European Commission, UAE, India, Australia, United States and Canada, among others, seek guidance from the IEEE Global Initiative and its members on how to instantiate the ethical principles and standards for AI they wish to build and implement and cite and use the outputs of the Global Initiative. Specifically, the second iteration of EAD was utilized by the OECD to create the OECD Principles on Artificial Intelligence. It advances SDG 3 (Good health & well-being), SDG 4 (Quality education), SDG 5 (Gender equality), SDG 7 (Affordable & clean energy), SDG 8 (Decent work & economic growth), SDG 9 (Industry, Innovation & Infrastructure), SDG 10 (Reduced Inequalities), SDG 11 (Sustainable cities & communities), SDG 13 (Climate action), and SDG 16 (Peace, justice, and strong institutions).
Maturity / Potential for adoption	The first version of EAD was launched in April 2016 and open for feedback. and the second version in 2018 for further feedback. The first edition was published in March 2019. The IEEEP7000 series is currently under development, with over thirteen standards approved.

22. ISO/IEC JTC 1/SC 42

Initiative	ISO/IEC JTC 1/SC 42
Category	AI and Governance
Brief Description	A standardization program made up of eight project working groups aiming to standardize technologies in the area of AI. It also provides guidance to JTC 1, IEC, and ISO committees developing AI applications. One committee, ISO/IEC TR 24028, focuses on improving trustworthiness in AI systems as well as identifying standardization gaps in AI. Another committee, ISO/IEC WD TS 4213, is working on an assessment of machine learning classification performance.

Organization	International Standards Organization
Geography	Cross-regional initiative with global scope.
Sector	International organization, civil society, and private sector
Key Success Factors	In addition to providing clearer guidance on trustworthiness and how it is being embedded in IT systems, ISO/IEC TR 24028 will help the standards community to better understand and identify the specific standardization gaps in AI and, importantly, how to address these through future standards work.
Key Hurdles	Some of the outputs are available only to ISO members or must be purchased. Process for joining working groups is complex.
Potential to contribute to GPAI objectives	Aligned with GPAI objectives, the initiative aims to foster the responsible development of AI through international standards. Successful elevation of national standards to the international level benefits national firms that have already built compliant systems. Successful inclusion of corporate patents into international standards can mean lucrative windfalls for both the firm and its home country.
Diversity and Inclusiveness	Standards are created through a multi-stakeholder process of international working groups, made up of technical and industry experts. Researchers can join the groups through their respective national standards body. Multinational organizations can join via liaison status. The initiative has high level partnerships with international organizations like European Commission, ITU, and OECD - exemplifying the importance of collaboration and coordination in the Responsible AI ecosystem.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	Till now, the initiative includes six published ISO standards, twenty-one standards under development, thirty-one participating members, and 16 observers. The committee contributes with eleven standards to the following SDGs: SDG 3, SDG 4, SDG 5, SDG 7, SDG 8, SDG 9, SDG 10, SDG 12, and SDG 14.
Maturity / Potential for adoption	The program was launched in 2017 and there are several initiatives that are currently under development. National actors, including the US and China, have agreed that international standards in AI are a priority.

23. Machine Learning Quality Management Guidelines

Initiative	Machine Learning Quality Management Guidelines
Category	AI and Governance
Brief Description	The Machine Learning Quality Management Guidelines provides a method to enable consistent quality management for AI-based product developments. Its mission is to “manage the quality of products and services using AI safely and securely”. The primary output will be a guideline document that provides guidance for goal-definitions and methods for AI developers. Specifically, it builds a quality assessment framework (such as setting levels of quality) associated with some technical guidance (similar to checklists) that allows developers to objectively evaluate quality with aims for international standardization. The initiative also develops tools, publishes reference documents and undertakes academic research on AI quality.

Organization	National Institute of Advanced Industrial Science and Technology, Cyber Physical Research Center, Software Quality Assurance Research Team, Artificial Intelligence Research Center
Geography	Initiative originated in Japan and has global scope.
Sector	Academia
Key Success Factors	Success factors include development of an (inter)active dialogue between academia and the private sector. As a result, experience from developing AI for products and services from the 'real world' informs the overall guideline document and vice-versa, making it highly relevant and useful industry players.
Key Hurdles	In the absence of any consensus of how quality of AI software can be assured, the initiative needs to start from zero and start disentangling how AI software quality can be understood logically.
Potential to contribute to GPAI objectives	By working closely with private sector representatives, the initiative provides a nexus to encourage and influence the safe development of AI in products and services. It is cross-sectoral, involving academia and the private sector.
Diversity and Inclusiveness	The project is hosted by the National Institute of Advanced Industrial Science and Technology (AIST), a national institute composed of researchers from both academia and industry. Members are active in the AI and software space. The project has a regular council meeting composed of engineers from industry partners. As such, the project targets primarily the Japanese private sector (e.g. online information is primarily in Japanese). However, the initiative seeks to broaden its scope beyond Japanese academics and industry partners and has published an English summary.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative aligns with SDG 4 (Quality education), SDG 9 (Industry, Innovation & Infrastructure), and SDG 10 (Reduced Inequalities). It relates to the OECD AI Principles on Human-centered values and fairness, Transparency and explainability, Robustness, security and safety, and Accountability.
Maturity / Potential for adoption	The programme was launched in June 2020 and is currently collecting feedback. It is scalable across specific academic communities. The project is funded by NEDO (one of Japan's national research funding agencies) and works closely with several industry partners.

24. Montreal Declaration: Responsible AI

Initiative	Montréal Declaration: Responsible AI
Category	AI and Ethics
Brief Description	The Montréal Declaration is a collective endeavor that aims to steer the development of AI to support the common good and guide social change by making recommendations with strong democratic legitimacy. The Declaration's first objective consists of identifying general ethical principles and values, applied to the digital and AI field, that promote the fundamental interests of people and groups. Its mission is to spark public debate and

	encourage a progressive and inclusive orientation to the development of AI. More specifically, the initiative aims to: (i) Develop an ethical framework for the development and deployment of AI; (ii) Guide the digital transition so everyone benefits from this technological revolution; and (iii) Open a national and international forum for discussion to collectively achieve equitable, inclusive, and ecologically sustainable AI development.
Organization	Université de Montréal
Geography	Initiative originated in Canada and has global scope.
Sector	Academia
Key Success Factors	The Montreal Declaration's success is notably based on its unique methodology and inclusiveness, around seven core values. These values, suggested by a group of ethics, law, public policy and artificial intelligence experts, have then been informed by a thorough deliberation process. This deliberation occurred through consultations held over three months in 2018, in 15 different public spaces, and sparked exchanges between over 500 citizens, experts and stakeholders from every horizon. Following this process, ten principles were put forward in the current Declaration. Although these principles reflect the moral and political culture of the society in which they were developed, they provide the basis for an intercultural and international dialogue.
Key Hurdles	Challenges included the complexity and resource-intensity of facilitating a multistakeholder deliberation process (including the general public) in a sustainable way. Coordinating and ensuring the smooth development of the project, in a limited timeframe, has been a major challenge. Going forward, the initiative will have to keep the declaration relevant and updated and extend signatories beyond Canada.
Potential to contribute to GPAI objectives	The initiative draws on both cross-sectoral academic expertise and informed exchanges with stakeholders and the general public. It is unique because the principles were a product of a deliberation process involving several stakeholders, including citizens. Although these principles reflect the moral and political culture of the society in which they were developed, they provide the basis for an intercultural and international dialogue. The Declaration is addressed to any person, organization, company or political representatives that wishes to take part in the responsible development of AI, whether it is to contribute scientifically or technologically, to develop social projects, to elaborate rules (regulations, codes) that apply to it, to be able to contest bad or unwise approaches, or to be able to alert public opinion when necessary.
Diversity and Inclusiveness	The project was led by a Steering committee, with representatives from Université de Montréal, Polytechnique Montréal, Université Laval and CIFAR. The Declaration development committee included a scientific team of more than multidisciplinary 30 experts. The aim of the methodology is to ensure that g stakeholders outside AI were extensively consulted. The deliberation process occurred through consultations held over three months, in 15 different public spaces, and sparked exchanges between over 500 citizens, experts and stakeholders from every horizon. Stakeholders also had the possibility to submit comments and reports online. The Declaration is easy to access and read (in seven languages), simple and short with multiple signatories.

Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has contributed to SDG 8 (Decent work & economic growth), SDG 9 (Industry, Innovation & Infrastructure), SDG 10 (Reduced Inequalities), SDG 12 (Responsible consumption & production), SDG 13 (Climate action), and SDG 16 (Peace, justice & strong institutions). The ethical values and principles put forward in the Montreal Declaration for a Responsible Development of AI, as well as the methodology used to draft it, are perfectly aligned with the OECD's principles.
Maturity / Potential for adoption	The Declaration was officially launched in December 2018. As of October 2020, 1,932 citizens and 108 organizations have signed. Although it is so far focused on Canada, organizers are working to promote the declaration nationally and internationally to expand its impact; researchers involved in its drafting have been directly involved in various international representations (Quebec Government; Government of Canada; U7; UNESCO; etc.). Locally, the initiative also works directly with companies (including a set of toolboxes to facilitate its appropriation by companies and help them turn ethical principles into operations / actions.)

25. Observatory from the fAIr LAC Initiative

Initiative	Observatory from the FAIR LAC initiative
Category	AI and Social Good
Brief Description	The Inter-American Development Bank (IDB) is leading fAIr LAC with the aim of promoting the responsible development and application of AI to improve the delivery of social services and eventually reducing growing inequalities in Latin America and the Caribbean. fAIr LAC initiative has three main objectives: i) Promote the dialogue around the responsible use of AI focused on citizens from a perspective of diversity and inclusion, through the promotion of a diverse ecosystem of experts, discussion tables, and conferences; ii)) Develop tools to guide the ethical and reliable use of AI in Latin America and the Caribbean through manuals, algorithmic audits, and specific guides; and iii) Encourage responsible AI adoption through pilot projects and the creation of regional hubs. fAIr LAC includes a map of beneficial AI applications in the region that is easily searchable for initiatives by country, sector, or case study. It also runs pilot AI projects to systematize the lessons learned from applications where AI helps create greater social impact and to create a cooperative environment so that projects may be scaled and emulated in the region.
Organization	Inter-American Development Bank
Geography	The initiative’s scope is Latin America & the Caribbean.
Sector	International organization
Key Success Factors	The motto of success for fAIr LAC is “from principles to practice” by producing and offering tools and services that provide added value, for both governments and entrepreneurs, in their paths to implement and adopt ethical and responsible AI. Key success factors include its multidisciplinary and multisectoral approach. It is critical to understand the use of AI as a tool to

	solve real problems; the technology should not be an end per se. Also, to listen to entrepreneurs' and governments' actual needs so that the projects/solutions have a real impact on the ecosystems they intend to help. The initiative's efforts directed towards the public sector measure success by the increasing demand for use cases or pilot projects using AI from governments. The development of knowledge products and technical tools have better informed governments and improved the understanding of opportunities and potential risks of using AI for Social Good.
Key Hurdles	The main challenge in trying to promote the ethical use of AI is associated with the lack of technical capacity, especially in the public sector. To increase the AI adoption rate in the region it is necessary to create capacity and make efforts to update the available expertise related to AI, not only to train enough specialists in this field of knowledge but also to enable a large number of persons to live and work with AI systems. Another main challenge is associated with the lack of incentives and interest that investors and markets put on the development of trustworthy AI solutions. Despite the growing concerns on data privacy and the impact of algorithms in society, this has not been translated into clear decisions from investors and consumers. On the contrary, developers continue to produce ethically questionable solutions and big players like corporations and governments continue to violate general principles of ethical AI.
Potential to contribute to GPAI objectives	This initiative serves as an example of cross-national collaboration to use AI for social good. fAIr LAC is part of the OECD network of experts on AI (ONE AI) and working on different international initiatives such as the OECD AI Policy Observatory, the AIxSDGs initiative of Oxford University, the Center for the Fourth Industrial Revolution (WEF), and IFC.
Diversity and Inclusiveness	The initiative has a diverse team of professionals in the fields of innovation, technology, education, health, entrepreneurship, and business all over Latin America and the Caribbean with HQ in Washington DC and collaborators in Europe. The initiative targets policymakers and entrepreneurs alike by disseminating lessons learned across AI for Social Good initiatives across the region. The fAIr LAC efforts directed towards the entrepreneurial ecosystems that develop AI and data solutions just started in May 2020, yet in this brief period of time have already obtained great support from the advisory board and partners from 12 countries and across 17 sectors. The demand for the services that fAIr LAC has created for entrepreneurs and investors, is another measure of success.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has published several different knowledge products and tools such as the report of Artificial Intelligence for Social Good in Latin America and the Caribbean, a guide on interoperability for governments, the guide of ethical data management, and the technical manual of AI life cycle (forthcoming), among others, and is testing the ethical self-assessment tool for AI projects with real initiatives within the IDB group to deploy the tool with governments later (the one for entrepreneurs, investors and business accelerators is also in progress). The initiative contributes to SDG 3 (Good health & well-being), SDG 4 (Quality education), SDG 5 (Gender equality), SDG 9 (Industry, Innovation & Infrastructure), SDG 10 (Reduced Inequalities), SDG 11 (Sustainable cities & communities), and SDG 17

	(Partnerships for the goals). It aligns with the following OECD AI Principles: Inclusive growth, sustainable development and well-being, Human-centered values and fairness, Transparency and explainability, Robustness, security and safety, Accountability (practically done through ethical self-assessment by participating governments and entrepreneurs).
Maturity / Potential for adoption	The initiative originated in 2020. fAIr LAC is currently designing and implementing nine pilot projects using AI as a tool to solve pressing issues and improve social services with governments in the four regional hubs of fAIr LAC: Jalisco (Mexico), Uruguay, Costa Rica, and Medellín (Colombia).

26. OECD Recommendation of the Council on Artificial Intelligence

Initiative	OECD Recommendation of the Council on Artificial Intelligence
Category	AI and Ethics; AI and Governance
Brief Description	The initiative provides a set of internationally agreed principles to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values. The Principles focus on AI-specific issues and set a standard that is implementable and sufficiently flexible to stand the test of time in this rapidly evolving field. The principles identify five complementary values-based principles for the responsible stewardship of trustworthy AI and call on AI actors to promote and implement them, these are: inclusive growth, sustainable development and well-being; human-centered values and fairness; transparency and explainability; robustness, security and safety; and accountability.
Organization	Organization for Economic Co-operation and Development (OECD) - within the Committee on Digital Economy Policy.
Geography	Global
Sector	International organization
Key Success Factors	There have been three key success factors. First, OECD's convening power and nature helped rally support by its member states and beyond, which has provided important stimulus for uptake and implementation. As such, the initiative has been able to inform the agenda setting of the past two G20 presidencies of Japan and Saudi Arabia. Second, its multi-stakeholder and inclusive decision process has allowed for input across sectors and disciplines. To inform the principles, the *AI Group of experts at the OECD* (AIGO) was established in 2018. AIGO comprised over 50 experts from different disciplines and different sectors. To move from principles to practice (i.e. development of practical guidance on implementation for policy makers), the *OECD Network of experts on AI* (or ONE AI) was established. Third, knowledge sharing and dissemination: the *OECD.AI Policy Observatory* (OECD.AI) was launched in February 2020 to help countries encourage, nurture and monitor the responsible development of trustworthy AI systems for the benefit of society.

Key Hurdles	A key challenge is the implementation of the OECD AI Principles ('bringing them to life') as well as coordination with numerous other complementary or parallel initiatives.
Potential to contribute to GPAI objectives	By design, the principles are linked to the Global Partnership on AI (GPAI), which was conceived as an international and multi-stakeholder initiative that advances cutting-edge research and pilot projects on AI priorities to advance the responsible development and use of AI that respects human rights and shared democratic values, as elaborated in the OECD AI Principles. Multi-stakeholdership and interdisciplinarity underpins the OECD AI Principles. The principles aim at adoption by governments.
Diversity and Inclusiveness	The principles are coordinated by the Digital Economy Policy Division within the OECD directorate for Science, Technology and Innovation. Through the *OECD Network of Experts on AI* (ONE AI), the principles and their road to implementation are informed by an outward perspective on AI, bringing in different voices from across sectors and disciplinaries. ONE AI is currently composed of 88 members and 76 observers, again from across sectors and disciplines.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative effectively fosters Responsible AI by providing principles and five recommendations to policy-makers pertaining to national policies and international co-operation for trustworthy AI, with special attention to small and medium-sized enterprises (SMEs) namely: investing in AI research and development; fostering a digital ecosystem for AI; shaping an enabling policy environment for AI; building human capacity and preparing for labor market transformation; and international co-operation for trustworthy AI. The initiative helps to address all SDGs. In their preamble, the OECD AI Principles articulate their regard to the Sustainable Development Goals set out in the 2030 Agenda for Sustainable Development adopted by the United Nations General Assembly (A/RES/70/1), which is further reflected in Principle 1.1.
Maturity / Potential for adoption	The OECD AI Principles were adopted by the OECD Council at Ministerial level on 22 May 2019. As the first intergovernmental standard on AI policies, they were adopted by OECD member countries and by countries beyond OECD members including Argentina, Brazil, Costa Rica, Malta, Peru, Romania and Ukraine. They have also informed the G20 agenda under the presidencies of Japan and most recently Saudi Arabia. Going forward, the *ONE AI* expert group has designed three working groups to help move the principles from theory to practice: : i) the ONE AI working group on Classifying of AI systems is developing a user-friendly framework to classify and help policy makers navigate AI systems and understand the different policy considerations associated with different types of AI systems; ii) the ONE AI working group on Trustworthy AI is identifying practical guidance and shared procedural approaches to help AI actors and decision-makers implement effective, efficient and fair policies for trustworthy AI, and; iii) the ONE AI working group on AI Policies is developing practical guidance for policy makers on investing in AI R&D; data, infrastructure, software & knowledge; regulation, testbeds and documentation; skills and labor markets; and international cooperation. In addition, ONE AI is creating a task force on AI computing to create a framework for understanding and measuring the key components of domestic AI computing capacity.

27. Open Kinyarwanda

Initiative	Open Kinyarwanda
Category	AI and Social Good
Brief Description	Open Kinyarwanda voice dataset is an initiative to build a freely publicly available speech to text data in Kinyarwanda (Rwanda's official language spoken by over 12 million people in Rwanda & 40 million in the region). Digital Umuganda in collaboration with the German development agency (GiZ), Mozilla & Government institutions is building a dataset of over 1,200 hours and 1,200,000 sentences through crowd-building. The objective is to give innovators, researchers & developers access to a key infrastructure to develop voice technology in Kinyarwanda. The end goal is to take away barriers to access information & services and build inclusive digital solutions that can be accessed by marginalized communities including areas with low literacy levels as well as people living with disabilities.
Organization	Digital Umuganda, GIZ, Mozilla Foundation
Geography	The initiative originated in Rwanda and has national scope.
Sector	A coalition of civil society and private sector organizations.
Key Success Factors	The initiative is crowd-build based. The network of dedicated contributors who also became mobilizers (through a program called Commoneers) successfully mobilized new contributors during the pandemic that halted physical data collection but also ensured that different anti-bias metrics were put in place. These included gender and age. By leveraging a voluntary contribution mechanism named "Umuganda" that was targeting physical infrastructure, the initiative adapted it to the digital age to be used to build digital infrastructure (open datasets).
Key Hurdles	Some of the key challenges were that freely publicly available texts in Kinyarwanda were inexistent, internet connection especially in remote areas was unstable, and the pandemic meant physical data collection events had to be changed to virtual events, slowing down mobilization efforts.
Potential to contribute to GPAI objectives	Open Kinyarwanda is an example of an applied AI project directly impacting three SDGs and using AI for social good. It exemplifies how AI can be used to make societies that are more inclusive, breaking down physical and digital barriers that are currently affecting marginalized groups.
Diversity and Inclusiveness	The core team is made mostly of Rwandan technologists. The initiative worked with voluntary contributors from local universities as well as the general public donating sentences and voices under CC-0 license.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has reached its objectives, turning Kinyarwanda, a heavily underresouced language, to one of the fastest growing open voice datasets globally with over 1200 hours in just 11 months. It directly has advanced SDG 9 (Industry, Innovation & Infrastructure), SDG 10 (Reduced Inequalities) and SDG 11 (Sustainable cities & communities).
Maturity / Potential for adoption	The initiative was launched in June 2017. By building a network of voluntary contributors, the Commoneers program ensured that the initiative would not only be sustainable but can easily be duplicated. Currently, the initiative is sharing best practices and lessons learnt with the Luganda data collection

	efforts, the initiative is also looking to expand data collection & knowledge transfer to other communities in the region and on the continent in general.

28. Partnership on AI Issue Area on Safety-Critical AI (SCAI)

Initiative	Partnership on AI Issue Area on Safety-Critical AI (SCAI)
Category	AI and Ethics; AI and Governance
Brief Description	Safety-Critical AI is an initiative within the PAI multistakeholder organization. PAI's goal is to develop the norms, institutions, and technical best practices necessary to ensure the safe research and deployment of AI technologies - particularly in high-stakes, dual-use, and/or safety-critical domains. It does so through a mix of whitepapers, academic research, workshops and convenings, and institutions and services such as expert committees. Thus far, domains that have been identified as high priority and safety critical are healthcare, finance, and autonomous vehicles.
Organization	Partnership on AI
Geography	The organization is based in the US and has global scope, including member organizations from around the world.
Sector	Civil society, academia, private sector
Key Success Factors	Key success factors include: i) a strong partner network that helps tap into diverse perspectives and interact with SCAI's target audience; ii) a team that combines both research (technical and non-technical) and project and programme management skills; and iii) working under autonomy in designing the initiative and prioritizing. In addition, three main initiative-specific success factors are that initiatives need to be grounded in technical realities (rather than lofty goals), evidence-based and pragmatic, and focused on specific interventions rather than being overly ideological. Multistakeholder projects need a project driver empowered with time, resources, and authority.
Key Hurdles	Challenges include: i) difficulty of measuring the success of projects that seek to create slow, long-term, institutional change (note some proxy indicators can be used to measure, e.g., level of engagement by partners) and slow feedback loops; ii) reconciling different perspectives across stakeholders and reaching consensus; iii) incentive structures both within the academic research ecosystem, and within commercial development; iv) field silos and blind spots as it can otherwise be hard to identify relevant experts in high-stake/ dual-use fields (as networks are primarily based within the AI community; and v) internal factions within the AI community (near-term versus long-term concerns and views).
Potential to contribute to GPAI objectives	PAI's mission to promote safe research and deployment of AI aligns with that of GPAI. Leveraging the PAI network of partners, which contains organizations across different regions and sectors (civil society, industry, academia, etc.) can be useful towards reaching GPAI objectives.
Diversity and Inclusiveness	SCAI has an interdisciplinary, multicultural and gender-balanced team (albeit rather small). Furthermore, PAI as an organization is currently working on a 'Diverse Voices' program to ensure diversity is built into its work and plans to

	include all relevant stakeholders, including those from marginalized communities and those most affected by the technologies going forward.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative is aligned with the UN SDG framework and, in particular, SDG 16 (Peace, Justice, and Strong Institutions) and SDG 17 (Partnerships for the Goals). In regard to alignment with OEDD principles, robustness, security, and safety are key aspects of safety-critical AI with current focus on the first and the latter.
Maturity / Potential for adoption	The initiative started in September 2016. Seeking long-term institutional change, it has been difficult to measure success (although some output indicators, such as number and frequency of partners for advice / help or their frequency or intensity of engagement provide a reference). So far, the initiative has hosted a series of repeat engagement exercises with the AI research community and advised companies on responsible AI development, particularly on advanced language models and deepfake detection.

29. UNESCO Recommendation on the Ethics of Artificial Intelligence & AI Decision Makers' Toolkit

Initiative	UNESCO Recommendation on the Ethics of Artificial Intelligence & AI Decision Makers' Toolkit
Category	AI and Ethics; AI and Governance
Brief Description	The UNESCO Recommendation expects to define shared values and principles and identifies concrete policy measures on the ethics of AI. Its role will be to help UNESCO Member States and other stakeholders ensure that they uphold the fundamental rights of the UN Charter and of the Universal Declaration of Human Rights and that research, design, development, and deployment of AI systems take into account the well-being of humanity, the environment and sustainable development. The recommendation will have a strong focus on moving from principles to practice, including through UNESCO's AI Decision Makers' Toolkit - a collection of knowledge products and tools from across UNESCO's fields of competence to help decision makers address some of the practical questions they face with respect to the development, use and governance of AI.
Organization	UNESCO
Geography	The initiative has global scope. Its target audience includes UNESCO member states and other AI actors, including the private sector.
Sector	International Organization
Key Success Factors	Key success factors included a rigorous multi-stakeholder process involving the perspectives of UNESCO member states, AI experts, civil society, and other stakeholder groups. Furthermore, the toolkit
Key Hurdles	Challenges included Covid-19 causing a disruption to physical meetings planned under the consultation process.
Potential to contribute to	The recommendation's mission closely aligns with that of GPAI. If the recommendation is adopted, it will serve as an ethical guiding compass and a global normative bedrock to build strong respect for the rule of law in the

GPAI objectives	digital world. The Decision Maker's toolkit will enable decision makers to respond to the challenges and opportunities of AI through guidance on policy development and provision of capacity building resources in UNESCO's fields of competence.
Diversity and Inclusiveness	The first draft of the recommendation was produced by 24 renowned specialists with multidisciplinary and pluralistic expertise on the ethics of AI. This initiative possesses an almost all female Bureau (4 women 2 men) and multidisciplinary Secretariat with inputs from all UNESCO sectors. The process for producing the draft leveraged multi-stakeholder consultations, including public online consultations generating more than 50000 comments; 11 regional and sub-regional virtual consultations in all regions of UNESCO involving more than 500 participants; and open multi-stakeholder citizen deliberation workshops in 25 countries with approximately 500 participants. It also includes inputs received from United Nations entities, major stakeholders from the private sector such as Google, Facebook and Microsoft, and the world of academia with the University of Stanford and the Chinese Academy of Sciences. The toolkit is informed by UNESCO's AI Needs Assessment Survey in Africa that received responses from 32 governments in Africa.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative is closely aligned to SDG 16 (Peace, justice, and strong institutions), as well as SDG 4 (Quality education), SDG 5 (Gender equality), SDG 9 (Industry, Innovation & Infrastructure), SDG 10 (Reduced Inequalities), SDG 11 (Sustainable cities & communities), SDG 13 (Climate action), and SDG 17 (Partnerships for the goals).
Maturity / Potential for adoption	The initiative launched November 2019. The final draft text will be presented for adoption by Member States during the 41st session of UNESCO's General Conference in November 2021.

30. UNICEF AI for Children

Initiative	UNICEF AI for Children
Category	AI and Governance; AI and Social Good
Brief Description	To explore how to embed child rights in the governing policies of AI, UNICEF's Office of Global Insight and Policy is leading a two-year project to explore approaches to protecting and upholding the rights of children in an evolving AI world. As part of the AI and Children policy project, UNICEF hosted a series of workshops around the world to gain regional perspectives on AI systems and children. These conversations helped UNICEF develop a draft policy guidance on how to promote child development in AI strategies and practices. UNICEF offers this draft policy guidance as a complement to efforts to promote human-centric AI, by introducing a child rights lens. The ultimate purpose of the guidance is to aid the protection and empowerment of children in interactions with AI systems and enable access to its benefits.
Organization	UNICEF
Geography	The initiative is cross-regional and has global scope.
Sector	International organization

Key Success Factors	Key success factors include specific focus on a targeted group: children. Also, the draft guidelines were produced through a multi-stakeholder process involving input by advisory groups and experts.
Key Hurdles	It is not clear if/how the guidance has been implemented at this stage.
Potential to contribute to GPAI objectives	Initiative aims to promote guidelines for the responsible development and deployment of AI systems influencing children. It builds understanding of what the unique issues to children are and also sets the principles developers, policymakers, and other relevant stakeholders should consider. Issues around children have already received wide international consensus, hence, a focus on this group can be a steppingstone for Responsible AI.
Diversity and Inclusiveness	The guidelines are produced by a diverse team, supported by a multistakeholder gender balanced advisory board. The guidelines were open to public consultation. Furthermore, the initiative is supported by and partnering with the Government of Finland, and collaborating with the IEEE Standards Association, the Berkman Klein Centre for Internet & Society, the World Economic Forum, the 5Rights Foundation and other organizations that form part of Generation AI.
Effectiveness / Alignment with UN SDG(s) and OECD AI Principles	The initiative has raised awareness of the impact of AI on children, and also provides practical guidelines addressing key issues. It advances the following UN SDGs: SDG 4, SDG 5, SDG 8, SDG 10, SDG 16, and SDG 17.
Maturity / Potential for adoption	The UNICEF AI for Children program was initiated in 2019 and the draft guidelines were published in September 2020. High potential for adoption as the initiative invites governments and the business sector to pilot this guidance in their field and openly share their findings about how it was used, and what worked and what did not, so that their real experiences can improve the document

Photo Credit: Christopher Burns

THE FUTURE SOCIETY

The Future Society is a nonprofit think-and-do tank with the mission to advance the responsible adoption of AI and other emerging technologies for the benefit of humanity.

GLOBAL PARTNERSHIP ON AI (GPAI)

GPAI is an international and multistakeholder initiative to guide the responsible development and use of artificial intelligence consistent with human rights, fundamental freedoms, and shared democratic values, as reflected in the OECD Recommendation on AI.

INTERNATIONAL CENTRE OF EXPERTISE IN MONTREAL FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE (CEIMIA)

CEIMIA is a Montreal-based Centre of Expertise, established to support GPAI's working groups.

The Future Society
867 Boylston Street, 5th Floor, Boston MA 02116, United States
www.thefuturesociety.org