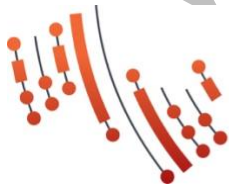


RESPONSIBLE AI IN ASIA

Working Draft



LIRNEasia
Pro-poor. Pro-market.

**Merl Chandana, Sukitha Bandaranayake, Anushka Jain, Aarushi Gupta,
Ana Florido**

Submitted to the Global Centre on AI Governance by LIRNEasia

About this report

This report is the final output of a Global Center on AI Governance (GCG)-funded project to develop a regional (Asia) report for the Global Index on Responsible AI (GIRAI) that focuses on responsible Artificial Intelligence in the Asia region.

About LIRNEasia

LIRNEasia is a pro-poor, pro-market think tank whose mission is catalyzing policy change and solutions through research to improve the lives of people in the Asia and Pacific using knowledge, information and technology. LIRNEasia has been active in the Asia Pacific since 2005, conducting both demand- and supply-side research as well as advocating for policy changes in the ICT sector on issues ranging from universal service policy to open data, gender, big data and more. For more information, visit lirneasia.net or follow @LIRNEasia.

About the Global Center on AI Governance (GCG)

GCG serves as a global hub for research and evidence-led action on inclusive and equitable approaches to the use and governance of AI technologies. Led from Africa, with partners from around the world, GCG advances local perspectives, interests, and expertise to level the playing field for global debates on AI governance. Filling major data gaps and equipping governments, policymakers, and researchers with local insights and tools, GCG is advancing new ways of addressing AI to ensure everyone benefits and no one is harmed through knowledge production and exchange. In parallel to GCG's multidisciplinary research activities, the organization conducts capacity strengthening in responsible AI, offering executive courses. GCG is also playing an active role in national, regional and global AI policy processes, and provide policy advisory services to AI decision-makers.

Authorship and Acknowledgements

This report was authored by Merl Chandana and Sukitha Bandaranayake, with the India case study written by Anushka Jain and Aarushi Gupta. Part II was co-authored by Merl Chandana, Sukitha Bandaranayake, and Ana Florido.

The authors thank the following individuals for reviewing and providing feedback on individual case studies: Grace Mirandilla-Santos, for the Philippines case study; Archana Atmakuri, for the Singapore case study; and Tom Barraclough, for the New Zealand case study.

The authors also wish to acknowledge Helani Galpaya for her valuable contributions during the conceptualization of the report's structure.

Contact: 15 2/1, Balcombe Place, Colombo 00800, Sri Lanka.

Telephone: +94 11 267 1160.

info@lirneasia.net

Table of Contents

<i>List of Acronyms</i>	3
<i>List of Tables</i>	6
<i>List of Figures</i>	6
<i>Executive Summary</i>	8
<i>Part I: Introduction</i>	10
<i>Part II: Analysis of Regional Data & Key Findings</i>	12
<i>Part III: Case Studies</i>	29
Responsible AI in Singapore	30
Responsible AI in The Philippines	42
Responsible AI in India	52
Responsible AI in Sri Lanka	62
Responsible AI in New Zealand	69
<i>Part IV: Outlook for Responsible AI in Asia</i>	79
RAI and Misinformation	79
AI & Data Protection in Asia	83
Regulating AI in Asia: Guiding Questions and Key Trends	87
Responsible AI in Government	92
<i>References</i>	96
<i>Appendix A: GIRAI Methodology</i>	121
<i>Appendix B: Heatmap of Adjusted Scores for Asia by Country by pillar sorted by score</i> .	122
<i>Appendix C: Description of the scope of thematic areas</i>	123
<i>Appendix D: Key Informant Interviews</i>	127

List of Acronyms

AAP	Analytics & Artificial Intelligence Association of the Philippines
ADI	Aboitiz Data Innovation
AI	Artificial Intelligence
AIAS	Artificial Intelligence Assessment Scale
AIDA	Generally accepted Principles for the use of artificial intelligence and data analytics
AIHGle	Artificial Intelligence in Healthcare Guidelines
AISG	Artificial Intelligence Singapore
AISI	AI Safety Institute
ARIDS	Automatic Road Incident Detection System
ASEAN	Association of Southeast Asian Nations
AWS	Amazon Web Services
BoK 2.0	Artificial Intelligence Ethics & Governance Body of Knowledge Version 2.0
BPO	Business Process Outsourcing
CAIR	Center for Artificial Intelligence Research
CEO	Chief Executive Officer
CeRAI	Centre for Responsible Artificial Intelligence, Madras
CFSAI	Committee on Formulating a Strategy for Artificial Intelligence
CoRE-AI	Coalition for Responsible Evolution of Artificial Intelligence
DPA	Data Protection Authority
DPDPA	Digital Personal Data Protection Act
DPI	Digital Public Infrastructure
DSA	Data Science and Analytics
ECCC	Emergency Command and Control Centre
EDCOM II	Second Congressional Commission on Education Act
EMT	Emergency Medical Technician
EU	European Union
FEAT	Fairness, ethics, accountability, and transparency
GCDO	Government Chief Digital Officer
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GenAI	Generative Artificial Intelligence
GIRAI	Global Index on Responsible Artificial Intelligence
IBM	International Business Machines Corporation
ICMR	Indian Council of Medical Research
ICT	Information Communication Technology
IRCTC	Indian Railways Catering and Tourism Corporation
IISc	Indian Institute of Science

IITs	Indian Institutes of Technology
IT	Information Technology
IMDA	Infocomm Media Development Authority
LLM	Large Language Model
MAS	Monetary Authority of Singapore
MCI	Ministry of Communications and Information
MeitY	Ministry for Electronics and Information Technology
MIT	Massachusetts Institute of Technology
MoC	Memorandum of Cooperation
MoU	Memorandum of Understanding
MSB	Multi Stakeholder Body
NAIC	National AI Center
NAIS 2.0	National Artificial Intelligence Strategy 2.0
NAISR 2.0	National Artificial Intelligence Strategy Roadmap 2.0
NAMFREL	National Citizens' Movement for Free Elections
NASSCOM	National Association of Software and Service Companies
NGO	Non-Governmental Organisation
NLP	Natural Language Processing
NITI	Aayog National Institution for Transforming India
NUS	National University of Singapore
NSA	Non-State Actor
NSAI	National Strategy for Artificial Intelligence
OECD	Organization for Economic Co-operation and Development
OFW	Overseas Foreign Worker
PDPA	Personal Data Protection Act
PDPC	Personal Data Protection Commission
RAI	Responsible AI
RAM	Readiness Assessment Methodology
R&D	Research and Development
SCAI	Singapore Conference on Artificial Intelligence
SEA-LION	South East Asian Languages in One Network
SKAIA	South Korean Artificial Intelligence Act
SLIIT	Sri Lanka Institute of Information Technology
SME	Small and medium-sized enterprises
SNG	Smart Nation Group
TRAI	Telecom Regulatory Authority of India
UK	United Kingdom
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization

UP	University of the Philippines
UPI	Unified Payments Interface
USD	United States Dollar
XAI	Explainable Artificial Intelligence

Working Draft

List of Tables

Table 1: Asian countries within the GIRAI	13
Table 2: Highest & lowest scoring thematic areas by region	19

List of Figures

Figure 1: GIRAI measurement structure (Adams et al., 2024, p. 10)	13
Figure 2: GIRAI scores by region	15
Figure 3: Distribution of GIRAI scores by region	15
Figure 4: Score by pillar by region	16
Figure 5: Boxplots of pillar-region scores.	17
Figure 6: Percentage of thematic areas addressed by pillar and region.	18
Figure 7: Average number of thematic areas addressed by pillar and region.....	18
Figure 8: GIRAI country overall scores, Asia	20
Figure 9: Regional scores, Asia sub-region scores	21
Figure 10: Asia sub-region pillar scores	21
Figure 11: Asia pillar by sub-region score boxplots	22
Figure 12: Percentage of thematic areas addressed by pillar and subregion in Asia	23
Figure 13: Average number of thematic areas addressed by pillar and subregion in Asia.....	24
Figure 14: Enforceability of government frameworks in Asia	25
Figure 15: Government actions in Asia - subregional breakdown	25
Figure 16: Distribution of Non-state actor initiatives in Asia.....	26
Figure 17: Thematic areas in Asia with more evidence of non-state actors initiatives than Governmental frameworks or actions	27
Figure 18: GIRAI country scores in Asia	32
Figure 19: Singapore RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 21)	32
Figure 20: Nine dimensions of Singapore’s proposed Model AI Governance Framework (Infocomm Media Development Authority, n.d.)	35
Figure 21: AI Verify Component 1: Testing Framework (AI Verify Foundation, 2024, p. 2).....	37
Figure 22: AI Verify Component 2: Toolkit (AI Verify Foundation 2023)	38
Figure 23: The Four Pillars For Ethical AI Adoption, Bok 2.0 (Lai, 2024)	40
Figure 24: GIRAI country scores in Asia	44
Figure 25: The Philippines RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 22)	44
Figure 26: National AI Strategy – Pillars, Dimensions, Imperatives, Tasks (Department of Trade and Industry, 2024a, p. 9)	45
Figure 27: GIRAI country scores in Asia	64
Figure 28: Sri Lanka RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 23)	64
Figure 29: Strategic Framework for Advancing AI in Sri Lanka (Committee on Formulating a Strategy for Artificial Intelligence (CFSAI), 2024, p. 5).....	65
Figure 30: New Zealand RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 22)	70
Figure 31: Artificial Intelligence and the Information Privacy Principles, (Office of the Privacy Commissioner, 2023, p. 1).....	72
Figure 32: Artificial Intelligence and the Information Privacy Principles (Office of the Privacy Commissioner, 2023, p. 6).....	73

Figure 33: Artificial Intelligence and the Information Privacy Principles (Office of the Privacy Commissioner, 2023, p. 9).....73
Figure 34: Algorithm Charter (New Zealand Government, 2020, p. 3)74
Figure 35: Algorithm Charter, Risk Matrix (New Zealand Government, 2020, p. 2)75
Figure 36: Decision tree for deciding to regulate (UNESCO, 2024, p. 43).....88
Figure 37: Heatmap of Adjusted Scores by Country by pillar sorted by score 122

Working Draft

Executive Summary

Artificial Intelligence is evolving rapidly, and Asia is at the center of this transformation—not just as a consumer of AI but as a driver of its innovation and governance. From semiconductors in Taiwan to digital public infrastructure in India to Singapore’s innovations in AI governance shaping responsible AI best practices, the region is defining AI’s global trajectory. However, AI’s governance remains highly uneven, raising concerns about whether AI’s benefits will be widely distributed or concentrated among a few. The Global Index on Responsible AI provides a data-driven benchmark to assess these disparities, offering insights into where Asia stands, how governance approaches differ, and what trends will shape the future of AI in the region.

An examination of the scores from the Global Index on Responsible AI reveals a stark gap between leading and lagging nations in Asia. Singapore and Japan, sit right outside of the global top 10, having made impressive strides on responsible AI, while countries in parts of South and Central Asia struggle with weak institutional capacity, limited regulatory oversight, and fragmented stakeholder engagement. One of the key findings is that government action often outpaces formal AI policy frameworks, highlighting a need for more structured governance mechanisms. While non-state actors—including civil society, academia, and the private sector—play a vital role in AI governance, their influence is strong in some countries but more constrained in others, highlighting the need to build capacity in this area.

Part III examines Responsible AI governance through case studies of India, Sri Lanka, Singapore, and the Philippines, representing diverse economic sizes, AI adoption levels, and regulatory approaches. Additionally, New Zealand is included for its strong regional ties and pioneering work in indigenous data sovereignty, offering insights relevant to marginalized communities across Asia. Singapore’s AI Verify initiative serves as a model for balancing regulatory oversight and industry flexibility, while India’s Digital Public Infrastructure (DPI) demonstrates how AI can drive inclusive digital transformation—albeit with concerns around data governance and algorithmic accountability. The Philippines’ National AI Strategy Roadmap aims to integrate AI into the economy, but implementation challenges persist due to regulatory gaps and uneven institutional capacity. In contrast, New Zealand’s human rights-centered approach—with a particular focus on Indigenous data sovereignty—offers a more ethical, inclusive model of AI governance. Meanwhile, Sri Lanka highlights the governance bottlenecks that limit Responsible AI adoption in smaller economies, where lack of coordination between government and industry on responsible AI and limited engagement by the academia poses a significant barrier.

Looking ahead, the Responsible AI landscape in Asia will increasingly intersect with several key emerging trends in the region that will shape governance and policy responses. AI will test the limits of existing and upcoming privacy laws, as countries like India and Sri Lanka refine their frameworks to balance data access for AI development with individual privacy rights. Misinformation is already a growing concern, particularly in election contexts across India and Indonesia, where AI-generated content is challenging efforts to curb disinformation. Meanwhile, governments are expanding AI use in public services, from Singapore’s smart city initiatives to Indonesia’s AI-assisted social protection programs. These applications will continue to keep posing questions about bias, transparency, and

human oversight. On the regulatory front, countries are taking varied approaches—from voluntary guidelines in Singapore to regulatory experimentation in ASEAN—as they navigate how to govern AI without stifling innovation.

AI governance cannot be effectively addressed at the national level alone - regional and global cooperation will be critical to ensuring Responsible AI is inclusive, fair, and aligned with human rights principles. Many countries in Asia still lack the institutional capacity to regulate AI effectively, making policy alignment, regulatory knowledge-sharing, and multi-stakeholder engagement even more essential. The Global Index on Responsible AI provides a practical tool to track progress, facilitate peer learning, and guide future policy efforts. Moving forward, regional platforms such as ASEAN could play a larger role in ensuring that all nations—not just the AI leaders—benefit from AI’s opportunities while mitigating its risks.

Working Draft

Part I: Introduction

Artificial Intelligence is evolving at an extraordinary pace, with new capabilities emerging almost weekly. Even without considering cutting-edge advancements, AI's current capabilities already have the potential to transform industries such as healthcare, education, and governance. However, like any powerful technology, AI brings both opportunities and risks. These range from privacy violations, algorithmic bias, and lack of accountability in individual applications to broader societal concerns such as labor displacement and disproportionate impacts on marginalized groups, including women and children. Addressing these risks requires a proactive and collective approach that ensures AI is deployed responsibly.

In response to these challenges, the concept of Responsible AI has gained prominence as a framework to align AI development with societal well-being. While terms like “trustworthy AI” and “AI ethics” are often used in similar discussions, Responsible AI emphasizes the obligations of AI actors in mitigating social and environmental harms. This approach complements existing AI governance frameworks and is rooted in international human rights law and the UN Principles on Business and Human Rights, reinforcing accountability and fairness in AI governance. The Global Index on Responsible AI represents a crucial step in this direction, providing an evidence-based tool to support governments, civil society, and stakeholders in advancing rights-based AI policies and practices.

While the foundational breakthroughs in Artificial Intelligence emerged from North America and Europe, AI's advancements and implications are increasingly global. Asia is no longer just an adopter of AI but a key driver of its development and deployment. Recent breakthroughs, such as DeepSeek in China, signal the region's growing role in frontier AI research. Taiwan's dominance in semiconductor manufacturing makes it a critical player in powering AI globally, while India's digital public infrastructure provides a model for scalable and inclusive AI applications. In Southeast Asia, Singapore is home to a thriving tech ecosystem, and the ASEAN region has made strides in AI governance, as seen in the ASEAN Guide on AI Governance and Ethics. Meanwhile, governments in Western Asia are investing heavily in AI-driven research institutes to advance AI for societal benefit.

However, AI progress across Asia is uneven, with several countries in South Asia and Central Asia lagging behind, risking further technological and economic isolation. Without collective action, disparities in AI access, governance, and benefits could widen, exacerbating regional inequalities. Examining the Global Index on Responsible AI through an Asian lens offers a critical opportunity to identify shared strengths and weaknesses, foster regional cooperation, and align governance practices with global standards while respecting local contexts. Understanding Asia's AI landscape is essential not only for advancing responsible AI in the region but also for shaping AI's global trajectory.

An examination of the Global Index on Responsible AI through the lens of Asia needs a data-driven assessment of the region's standing compared to the rest of the world. Understanding

intra-Asia variations—such as differences between East, South, Southeast, Central, and Western Asia—can help identify shared strengths and regional disparities. Beyond these comparisons, one could analyze progress across different dimensions of Responsible AI, considering the roles of governments, industry, and civil society in shaping AI governance. This foundational analysis ensures that the discussion is anchored in empirical evidence rather than abstract generalizations.

Beyond insights gleaned from the data, situating the findings within specific national contexts is crucial for a more nuanced understanding of Responsible AI in Asia. Country-level case studies can illustrate how AI policies and governance mechanisms interact with broader economic and digital landscapes, showing how different actors contribute—or fail to contribute—to Responsible AI. These case studies should span both leading and lagging nations, highlighting best practices, systemic challenges, and governance gaps. Finally, a forward-looking analysis should explore how Responsible AI will evolve in Asia, identifying emerging trends and priority issues that will shape the region’s AI governance landscape. Instead of focusing solely on country-level variations, this section should examine regional shifts, policy trajectories, and the broader forces influencing AI governance, ensuring that the report remains relevant for future developments.

As one of two major research hubs in Asia contributing to the first edition of the Global Index on Responsible AI, LIRNEasia has had a front-row seat to the evolving AI governance landscape in the region. Through this process, we have gathered extensive data, engaged with stakeholders, and analyzed policy developments, gaining insights into both the strengths and challenges of AI governance across different Asian contexts. In partnership with the Global Center on AI Governance, this report builds on that foundation to fill a critical gap—providing a structured analysis of Responsible AI in Asia to inform policymakers, researchers, and civil society.

This report is designed to offer a balanced and multidimensional perspective by integrating three key elements: quantitative data analysis, country case studies, and forward-looking insights. The first section examines where Asia stands in the Global Index, identifying key trends and regional patterns. The second section contextualizes these findings through in-depth national case studies, highlighting both best practices and governance gaps. The final section takes a forward-looking approach, identifying the key developments that will shape AI governance in the region. By synthesizing data, contextual insights, and future trends, this report aims to bridge governance gaps, support regional cooperation, and align AI policies with global best practices while respecting local contexts.

Part II: Analysis of Regional Data & Key Findings

In Part II, the report analyzes the GIRAI data, highlighting key trends, variations, and outliers with the aim of offering practical insights for different actors in the AI ecosystem. It starts with background information on the study and methodology. The analysis is divided into two parts: first, a comparison of Asia with the rest of the world; second, an intra-region comparison revealing similarities, variations, and notable granular trends within Asia.

1) Introduction/Methodology

What is the GIRAI?

The Global Index on Responsible AI GIRAI is the first tool to set globally relevant benchmarks for responsible AI and assess them in countries around the world. This study constitutes the largest global data collection on responsible AI to-date. In its first edition, the Global Index on Responsible AI covers 138 countries and jurisdictions. Within the Index, Asia consists of the following 5 subregions and 38 countries:

Central Asia	Eastern Asia	Southern Asia	South-Eastern Asia	Western Asia
Kazakhstan	China	Afghanistan	Cambodia	Armenia
Kyrgyz Republic	Hong Kong Special Administrative Region	Bhutan	Indonesia	Azerbaijan
Tajikistan	Japan	India	Lao People's Democratic Republic	Bahrain
Turkmenistan	Mongolia	Nepal	Malaysia	Georgia
Uzbekistan	Republic of Korea	Pakistan	Myanmar	Jordan
	Taiwan	Sri Lanka	Philippines	Kuwait
			Singapore	Lebanon
			Thailand	Oman
			Vietnam	Palestine
				Qatar
				Saudi Arabia
				United Arab Emirates

Table 1: Asian countries within the GIRAI

How does the GIRAI measure responsible AI?

The Index “measures [each country’s performance across] **19 thematic areas** of responsible AI, which are clustered into **3 dimensions**: Human Rights and AI, Responsible AI Governance and Responsible AI Capacities. Each thematic area assesses the performance of **3 different pillars** of the responsible AI ecosystem: Government frameworks, Government actions, and non-state actors’ initiatives” (Adams et al., 2024, p. 10). Figure 1 below illustrates how the thematic areas, dimensions, and pillars relate to one another.

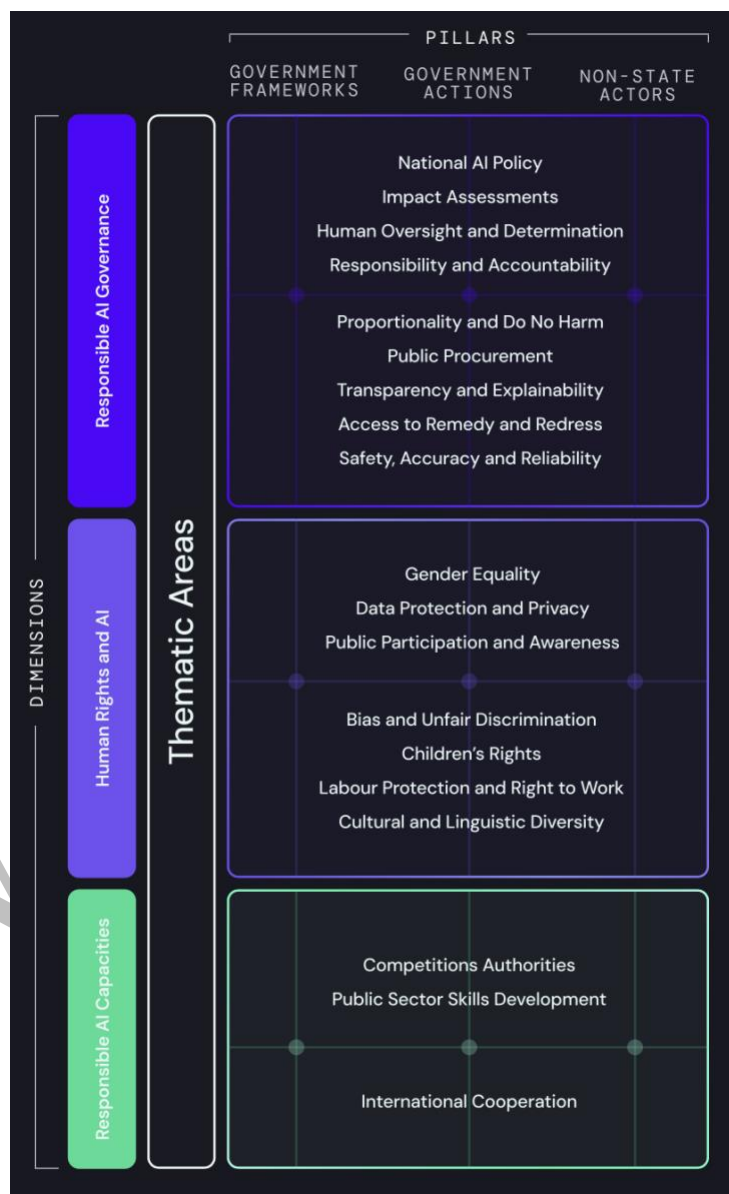


Figure 1: GIRAI measurement structure (Adams et al., 2024, p. 10)

Specific definitions are as follows (Adams et al., 2024, p. 9):

Thematic area = composite indicator measuring the performance of the responsible AI ecosystem in relation to a sub-component of responsible AI

Dimension = cluster of thematic areas

Pillar = separate category of evidence collected and assessed in relation to the responsible AI ecosystem

Government frameworks = national or federal laws, regulations, policies, strategies and/or guidelines that address the implications of AI with respect to a particular thematic area

Government actions = actions by national or federal government that involve the development or implementation of government frameworks, or government-led initiatives which advance action within the identified thematic area, even in the absence of a government framework

Non-state actors = actors outside government (universities, civil society organizations, and private sector entities) who are actively working on issues related to AI within the thematic area

Further information on data collection methodology and a discussion on potential limitations of the methodology can be found in Appendix 1 at the end of the report.

2) Asia vs the rest of world

While Asia is a diverse region comprising countries at varying stages of AI maturity. While Asia is not a monolith, comparing its scores on the Global Index of Responsible AI with those from other regions provides a window into the shared strengths and weaknesses, encouraging regional cooperation to address common challenges. Such insights will also support the alignment of governance practices with global standards, fostering policy coherence while respecting regional and country-level diversity.

In this report our comparator regions will be Asia (with 38 countries); Africa, 41; Europe, 31; the Americas, 26; Oceania, 2. For analytical purposes, countries are grouped according to the United Nations Standard Country or Area Codes for Statistical Use. It is important to note that Kosovo and Taiwan are not part of this classification; however, for the purposes of this analysis, they have been included in Europe and Asia, respectively.

Overall scores

Asia holds an average score of 19.3, on the index just below the global average of 19.8. For further context, 82 out of all 138 countries in the Index scored below the Asian average. The continent places third among the other regions – Oceania, Europe, the Americas, and Africa, as indicated by the graph below.

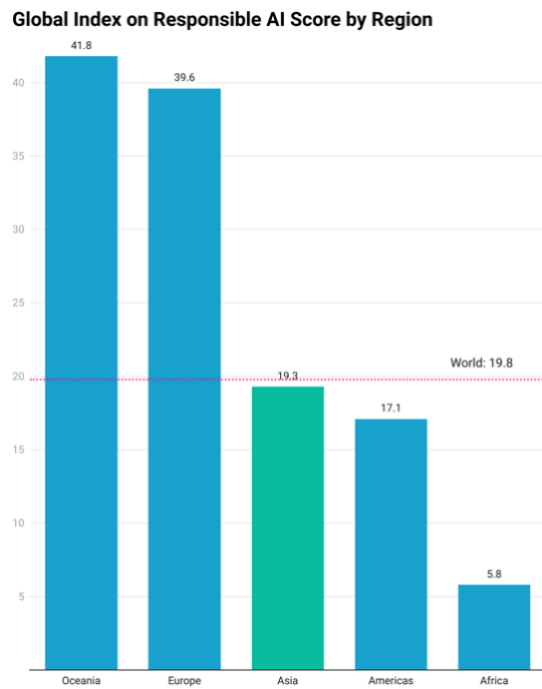


Figure 2: GIRAI scores by region

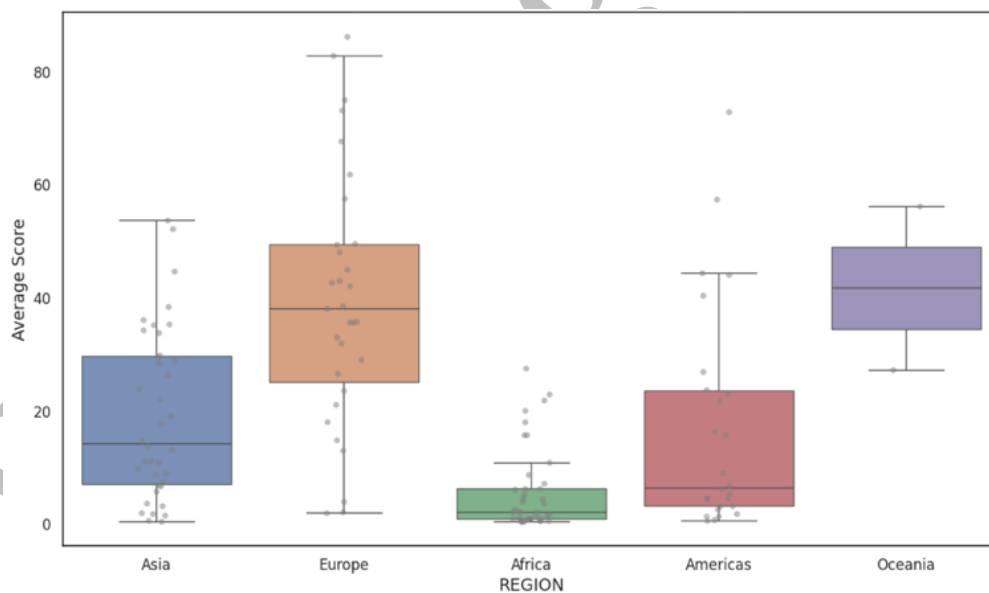


Figure 3: Distribution of GIRAI scores by region

It is important to note the considerable variation in scores within Asia (none of which ranked in the top 10), reflected in the box plots in Figure 3 above. Intra-Asian scores range from Singapore (ranked 11th overall with a score of 53.77) and Japan (12th; 52.21), to Myanmar (134th; 0.6) and Afghanistan (135th; 0.57). More on this in “Section 3) Intra-Asia analysis”

Overall pillar scores

Figure 4 shows the Asia scores across each pillar against the global average. As seen below, Asia ranked third across Government Frameworks and Government Actions pillars, while falling to fourth on the Non-State Actors pillar.

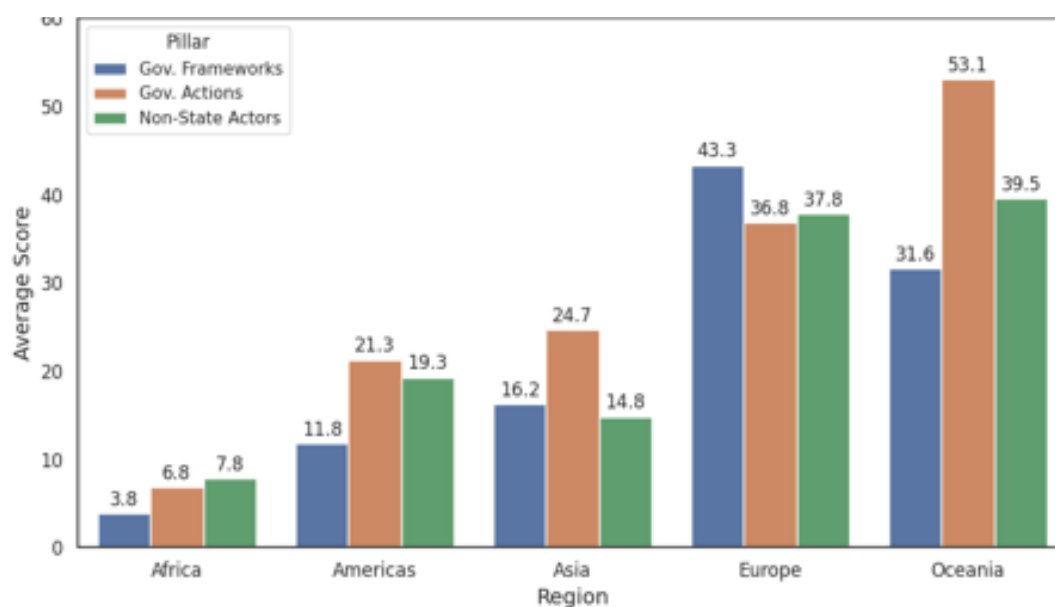


Figure 4: Score by pillar by region

The phenomenon of government actions outpacing the development of formal frameworks in Asia is common across all regions except Europe, widely regarded as the global leader in AI regulation. In Asia, this regulatory gap may result from deliberate decisions by governments to avoid stifling AI innovation with heavy, untested regulation. Alternatively, it may reflect the nascent stage of both AI adoption and Responsible AI (RAI) initiatives in these countries. Nonetheless, the significant disparity in government regulations between Asia and regions like Europe and Oceania highlights an area worth consideration for policymakers in the region.

Africa, like Asia, likely has fewer frameworks, due to the nascency of the field of Responsible AI. However, at least on relative terms, non-state actors in Africa have stepped up to fill this gap, as evidenced by their high scores, likely driven by collaborations with NGOs and international organizations. In Asia, by contrast, non-state-driven RAI efforts have lagged behind government actions. This may be because, in countries where RAI is strong, such as Singapore, top-down state-led initiatives dominate, leaving limited scope for non-state actors to contribute (e.g., private sector organizations may rely on government-issued guidelines rather than creating their own). Furthermore, the prevalence of autocratic or repressive governments in parts of Asia may restrict the role of non-state actors in advancing

RAI due to factors such as limited civil society engagement, constraints on freedom of expression, and more pressing human rights priorities taking precedence over AI.

Unlike Africa and the Americas, Asia doesn't find many outliers as found in the boxplots of pillar scores by region (Figure 5). However, there are two notable outliers in the non-state actors pillar which will resurface in the intra-regional analysis. Asia also exhibits significant variability across all three pillars (frameworks, actions, and non-state actors), second only to Europe, highlighting the region's diverse approaches and capacities.

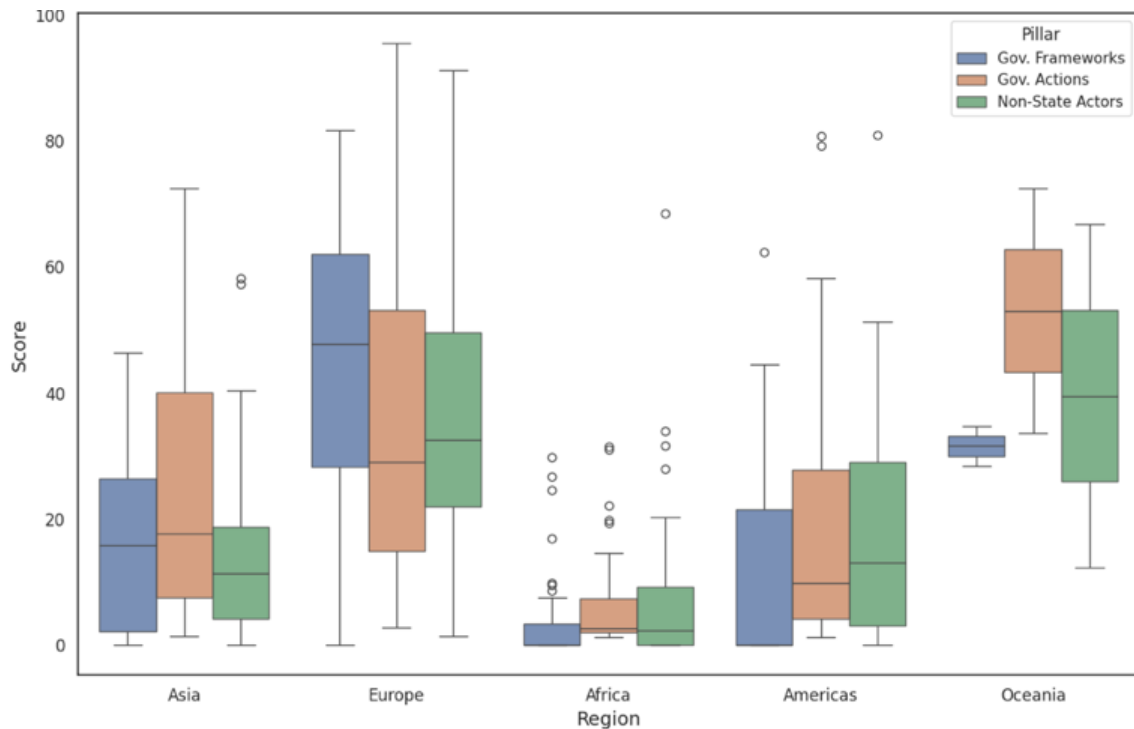


Figure 5: Boxplots of pillar-region scores.

Thematic area analysis

Asia ranked third in the percentage of different Responsible AI thematic areas addressed by different pillars (as shown in Figure 6). This percentage is calculated as follows: since all 19 thematic areas are considered for each country, the total number of thematic areas evaluated in a region is the number of countries in that region multiplied by 19. For example, if all 10 countries in a region show evidence for every thematic area, the region is assigned a score of 100%.

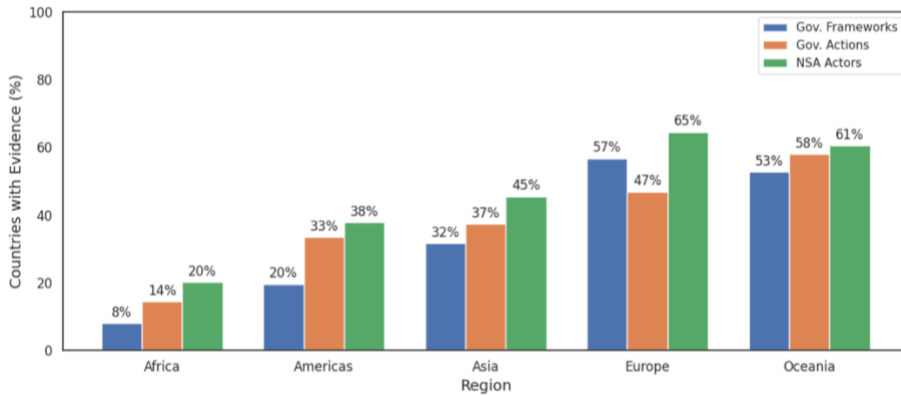


Figure 6: Percentage of thematic areas addressed by pillar and region.

Figure 7 shows the country-level average number of thematic areas covered by each pillar, across all five regions. Within Asia, Non-State Actor initiatives covered the highest number of thematic areas (9), followed by Government Actions (7), and Government Frameworks (6). This, naturally, does not tell us which thematic areas were satisfied in which countries, which is what this report clarifies in Section 3.

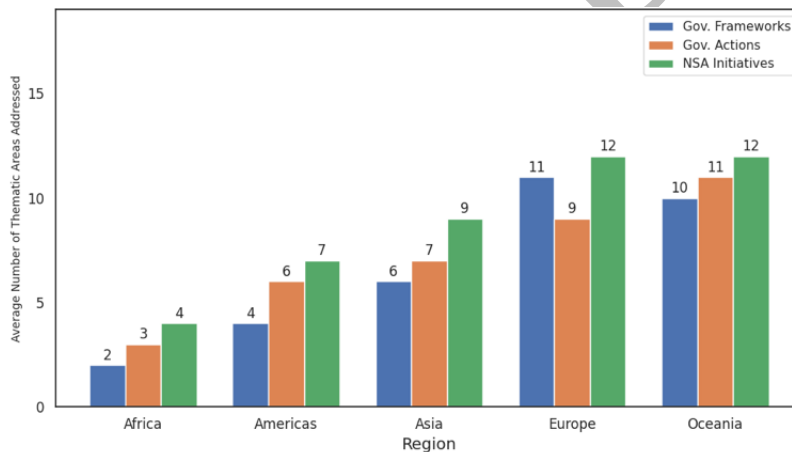


Figure 7: Average number of thematic areas addressed by pillar and region.

As seen in Table 2 below, Asia shows identical patterns in best-performing thematic areas to Africa and the Americas, suggesting shared priorities across these regions. For instance, despite varying levels of AI maturity, most countries in these regions have some form of data protection regulation and (at least on paper) a national AI strategy. Additionally, the diversity within these regions likely drives a stronger focus on international cooperation. Interestingly, Europe also ranks data protection and international cooperation among its top three thematic areas, though likely for different reasons.

Competition Authorities emerge as one of Asia's weakest-performing thematic areas, a pattern observed globally, suggesting this is an under-focused aspect of Responsible AI governance that needs global attention. Strengthening competition oversight could be critical as AI increasingly shapes markets and economic dynamics. Similarly, Public Procurement,

another weak area for Asia, warrants particular attention in the region, given governments' growing interest in adopting AI in Asia. Asia is not unique in this regard, as public procurement is also one of the lowest-performing areas in the Americas and Europe. This presents an opportunity for regional coordination among Asian countries to explore collective action, while ensuring that national contexts and priorities remain the primary focus.

Region	Top 3		Bottom 3	
	Thematic Area	Score	Thematic Area	Score
Africa <i>Average Score = 5.8</i>	International Cooperation	23.1	Children's Rights	0.8
	National AI Policy	13.0	Competitions Authorities	0.8
	Data Protection and Privacy	11.6	Access to Remedy and Redress	2.4
Americas <i>Average Score = 17.1</i>	International Cooperation	34.4	Access to Remedy and Redress	6.4
	National AI Policy	29.6	Competitions Authorities	7.0
	Transparency and Explainability	23.1	Public Procurement	10.8
Asia <i>Average Score = 19.3</i>	International Cooperation	41.7	Competitions Authorities	7.9
	National AI Policy	34.0	Access to Remedy and Redress	9.2
	Data Protection and Privacy	24.4	Public Procurement	9.3
Europe <i>Average Score = 39.6</i>	Data Protection and Privacy	58.6	Public Procurement	15.8
	International Cooperation	57.0	Competitions Authorities	21.9
	Responsibility and Accountability	52.9	Gender Equality	27.1
Oceania <i>Average Score = 41.8</i>	Bias and Unfair Discrimination	84.0	Competitions Authorities	6.5
	Human Oversight and Determination	77.3	Public Sector Skills Development	20.5
	Data Protection and Privacy	67.4	Access to Remedy and Redress	20.5

Table 2: Highest & lowest scoring thematic areas by region

3) Intra-Asia analysis

Next, we turn to an intra-Asia analysis, to explore the challenges faced by different parts of the region and highlight areas for action. Asia is home to some of the world's lowest-performing countries in the Global Index on Responsible AI, highlighting the importance of identifying regional best practices to foster collaboration and support underperforming areas. Additionally, such analysis can strengthen regional resilience by revealing shared challenges, such as improving public procurement processes and enhancing competition authorities, especially as the region increasingly turns to digitization to drive economic growth.

Overall

We see considerable disparity in scores across subregional levels and individual countries, overall and across the pillars. Figure 8 shows the disparity in scores across individual countries, from Singapore leading with an overall score of 53.8, to Afghanistan coming in last with at 0.6. Figure 9 further demonstrates this disparity by comparing Asia and its regions with the rest of the world and other sub-regions of Asia. East Asia is performing notably well, approaching scores comparable to Europe and Oceania, though not quite reaching their levels. In stark contrast, Central Asia lags significantly behind, underscoring the uneven development across the region and the need for interventions to bridge these gaps.

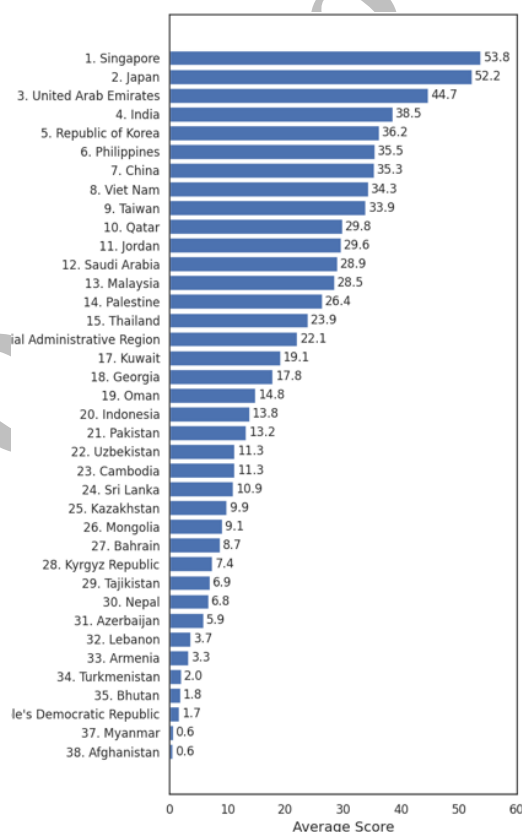


Figure 8: GIRAI country overall scores, Asia

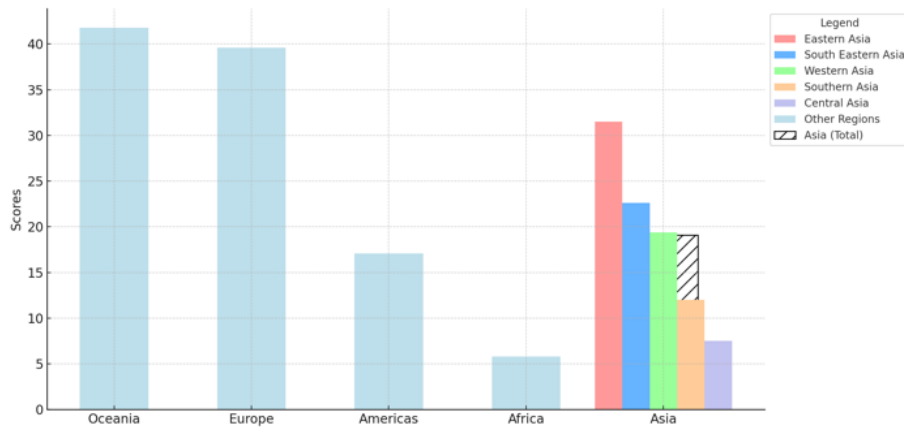


Figure 9: Regional scores, Asia sub-region scores

Overall pillar scores

The analysis of pillar scores across Asian subregions reveals varied progress within pillars and subregions:

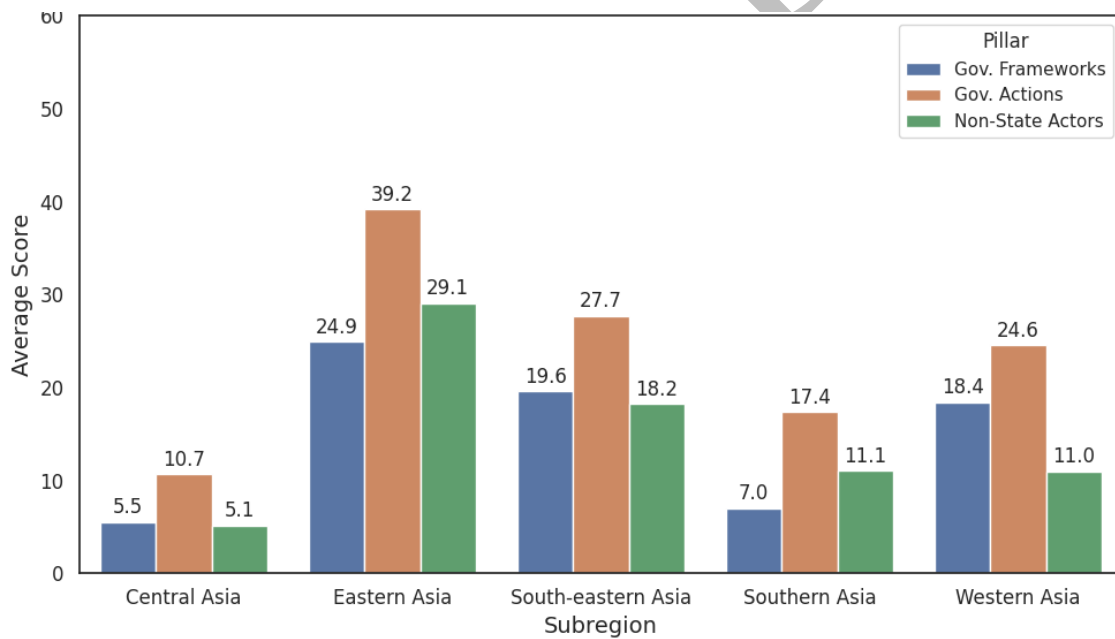


Figure 10: Asia sub-region pillar scores

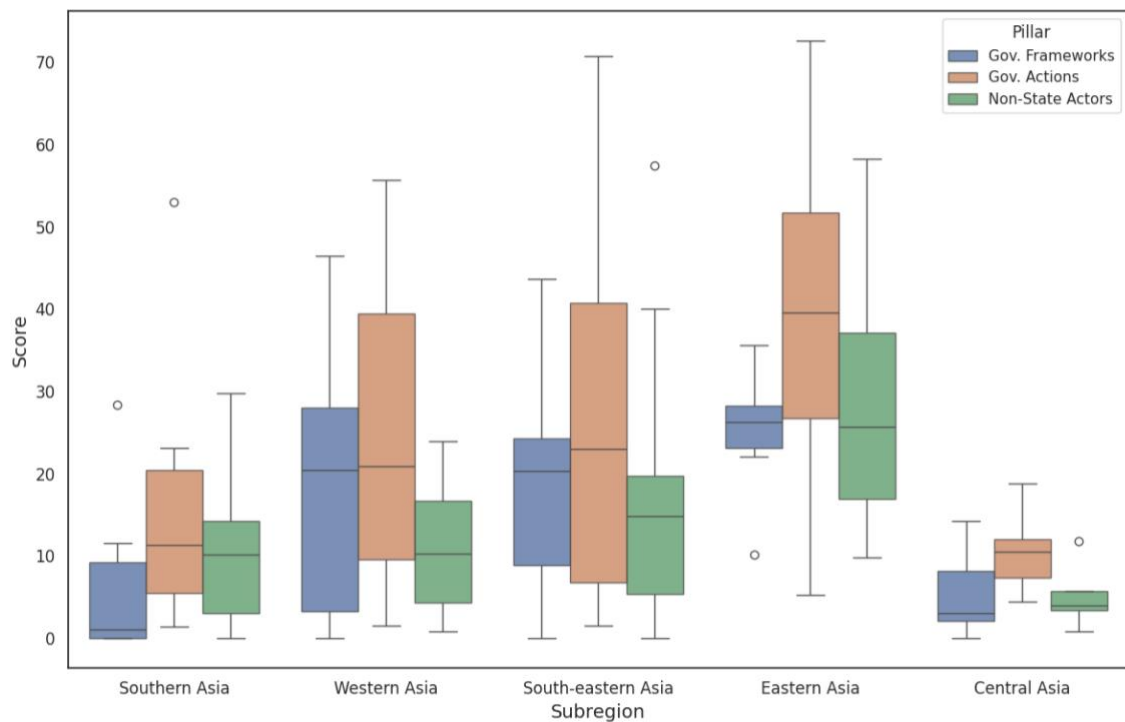


Figure 11: Asia pillar by sub-region score boxplots

Eastern Asia Outperforms Across All Pillars

Eastern Asia leads the region in Responsible AI, with strong performances across all pillars. Government actions score the highest at 39.2, followed by non-state actors at 29.1, and government frameworks at a relatively high 24.9, indicating a consistent and well-rounded approach to AI governance. Despite some variability reflected in the broad range of scores shown in the box plot, Eastern Asia maintains a generally high level of performance. Mongolia, however, stands out as an outlier on the lower end of the Government Frameworks pillar with a score of 10.19.

Balanced but Moderate Scores in South-Eastern and Western Asia

South-Eastern and Western Asia display balanced but relatively moderate scores across the pillars, with government actions leading in both subregions (27.7 and 24.6, respectively). While their scores for government frameworks and non-state actors are comparable, they fall short of the higher levels observed in Eastern Asia. The wide variability in scores within these subregions, as indicated by the box plots, highlights the uneven performance of individual countries. Notably, the Philippines emerges as an outlier in the Non-State Actors pillar, scoring an impressive 57.4.

Central and Southern Asia Lag Behind

Central and Southern Asia score significantly lower across all pillars, underscoring the challenges faced by these subregions. In Central Asia, government frameworks and non-state actors score around 5.1–5.5, while government actions reach 10.7. The narrower range in the

box plots suggests less variability within the subregion, with the Kyrgyz Republic standing out as an outlier in the NSA pillar at 11.8. Southern Asia also shows generally low scores across all pillars, but India emerges as a notable outlier, scoring 28.4 in Government Frameworks and an impressive 53.0 in Government Actions. These outlier scores underscore India’s leadership within the subregion, sharply contrasting with the lower performance of its neighbors.

Thematic area scores

For each subregion within Asia, Figure 12 shows the percentage of thematic areas covered by each pillar, while Figure 13 shows the average number of thematic areas covered by each pillar. Refer to explanation on page 6 of this section/report for how these percentages are calculated.

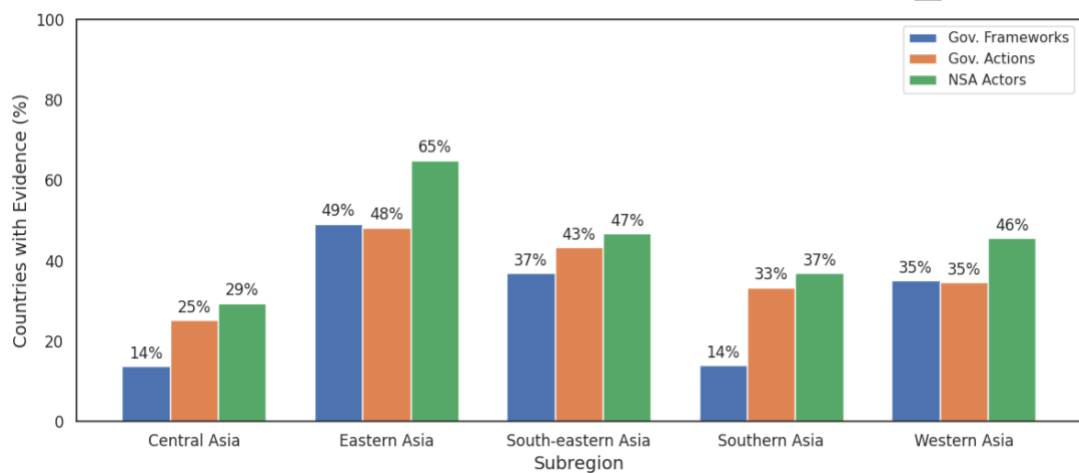


Figure 12: Percentage of thematic areas addressed by pillar and subregion in Asia

Eastern Asia is the best performing sub-region, with Government Frameworks, Government Actors and Non-State Initiatives all satisfying more thematic areas in Eastern Asia than any other region: 49%, 48%, and 65%, respectively. South-eastern and Western Asia, and Southern Asia follow, with Central Asia seeing the poorest performance across Government Actions (25%) and Non-State Actors (29%) and sharing the poorest performance on Government Frameworks with Southern Asia (14%).

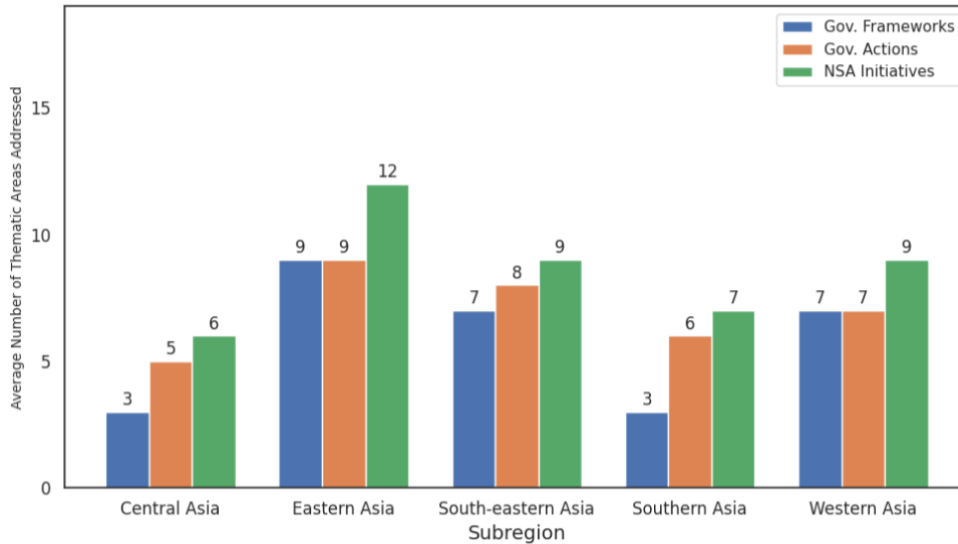


Figure 13: Average number of thematic areas addressed by pillar and subregion in Asia

Pillar Analysis: In depth

As described at the beginning, within GIRAI, a pillar refers to a distinct category of evidence assessed within the responsible AI ecosystem. Government efforts are classified into two pillars: regulations and policies (called frameworks), and other initiatives (called government actions). Non-state actions are scored under one pillar, but evidence is captured across three categories to reflect their diverse roles. This section of the intra-region analysis examines trends across these pillars and their variations across different sub-regions in Asia.

Government frameworks pillar

The Global Index on Responsible AI analyzed 78 unique government frameworks relevant to responsible AI, that cover on average six thematic areas in Asian countries. The chart below illustrates the level of enforceability of these frameworks.

In Asia, 43.6% of government frameworks are binding, meaning they include policies, strategies, long-term plans, etc. that outline the government’s intended approach and expected outcomes, without imposing legal obligations. Another 35.9% are non-binding, while 20.5% are legally enforceable frameworks with specific legal provisions for compliance.

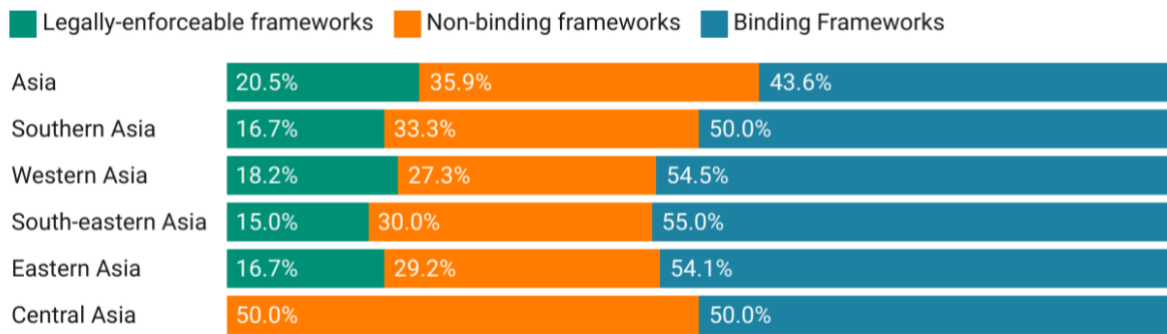


Figure 14: Enforceability of government frameworks in Asia

These figures indicate that a significant portion of frameworks (79.5%) are forms of 'soft regulation,' lacking legal weight. All subregions except Central Asia share a similar pattern, where:

- Binding frameworks are prioritized, ranging from 50% to 55%.
- Non-binding frameworks make up approximately 27-33%.
- Legally enforceable frameworks represent around 15-18%.

This pattern seems to suggest a general preference across these subregions for structured guidance that outlines government intentions and priorities without enforceable obligations. Central Asia stands out with an equal split between binding frameworks (50%) and non-binding frameworks (also 50%), pointing to a large gap in legally enforceable frameworks.

Government actions pillar

Across Asia, 189 unique government actions were identified, spread across 270 thematic areas. On average, government actions cover approximately 7 thematic areas per country, with 37.4% of Responsible AI thematic areas addressed at the regional level.

Subregional Breakdown:

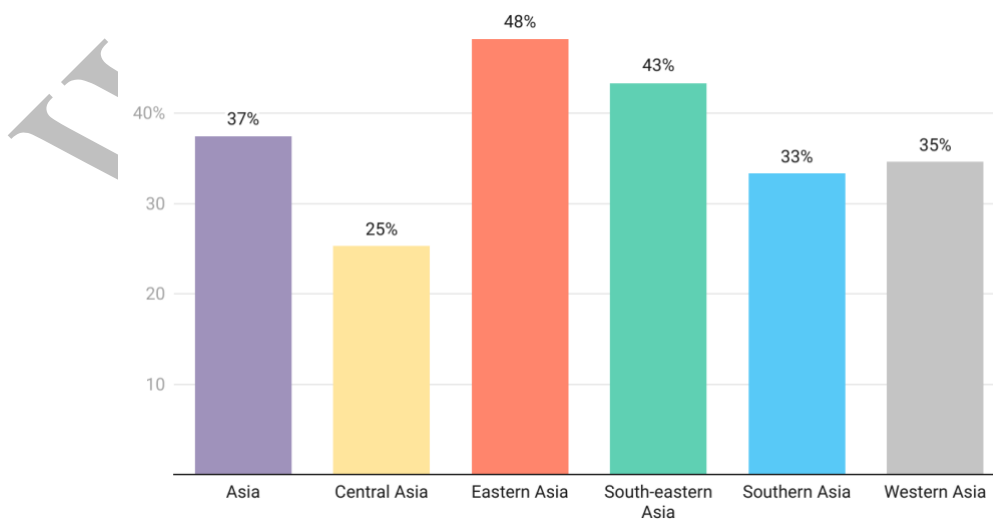


Figure 15: Government actions in Asia - subregional breakdown

Central Asia shows the lowest percentage of thematic areas covered by government actions at 25.3%, with each country averaging coverage of 5 thematic areas. This limited coverage suggests Central Asia as the least engaged subregion in Responsible AI government actions. Eastern Asia leads the region with the highest percentage of thematic area coverage at 48.2%. Countries in this subregion cover an average of 9 thematic areas, addressing a broader range of Responsible AI topics. South-eastern Asia also demonstrates significant engagement, with 43.3% of Responsible AI topics covered and an average of 8 thematic areas per country. Both Southern and Western Asia show moderate engagement in Responsible AI, covering roughly one-third of thematic areas (33.3% in Southern Asia and 34.6% in Western Asia). Each country in these subregions addresses around 6 thematic areas on average (6.3 in Southern Asia and 6.6 in Western Asia), reflecting a similar scope and depth of government action.

Overall, Eastern Asia and South-eastern Asia have the most comprehensive coverage, while Central Asia’s low percentage and average reveal a more constrained focus, positioning it as the subregion with the least engagement in Responsible AI government actions.

Non-state actors

Non-state actors, including academia, civil society, and the private sector, play a critical role in advancing Responsible AI by complementing government efforts and filling gaps in regulation and implementation. Academia drives research, innovation and evidence-based policymaking, civil society ensures accountability and advocates for ethical practices, and the private sector integrates responsible practices into operations while setting industry standards.

In Asia, the Global Index analyzed evidence of 469 initiatives from non-state actors—including academia, civil society, and the private sector. The charts below illustrate the distribution of these initiatives across sectors, with a breakdown by subregion.

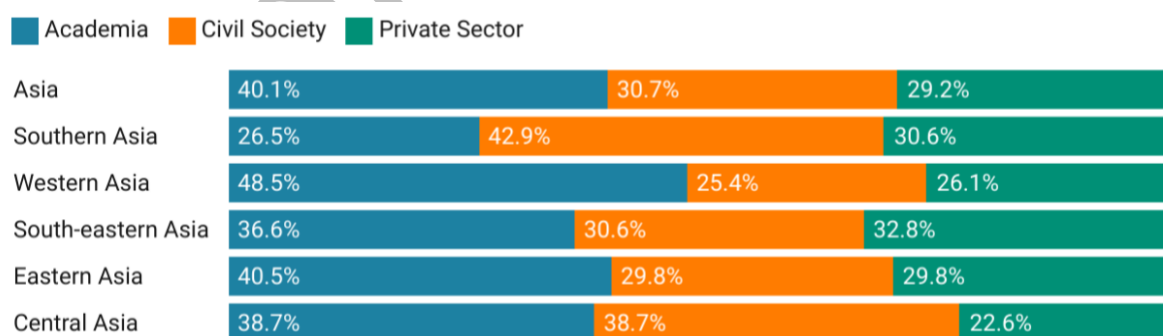


Figure 16: Distribution of Non-state actor initiatives in Asia

The overall distribution of non-state actor (NSA) initiatives in Asia reveals that 40.1% are led by academia, underscoring the outsized role of universities and research institutions in advancing Responsible AI in the region. Civil society accounts for 30.7% of initiatives, while

the private sector contributes 29.2%, reflecting a relatively balanced involvement between these two sectors.

Subregional trends paint a slightly more nuanced picture. In Western Asia almost half of the non-state actor initiatives are led by academics (at 48.5%), while Southern Asia stands out for its strong civil society involvement, with 42.9% of initiatives driven by this sector, surpassing both academia and the private sector. Central Asia shows an equal distribution between academia and civil society at 38.7% each, with the private sector contributing a smaller share (22.6%). South-eastern Asia sees the most balanced efforts from all non-state actors, with academia (36.6%), private sector (32.8%), and civil society (30.6%) contributing equally. In fact, it is the only region where all 3 types of actors have a contribution over 30%.

Private sector engagement varies significantly across subregions of Asia, ranging from 22.6% in Central Asia to 32.8% in South-eastern Asia. This variation reflects differing levels of industry commitment to Responsible AI across the region, with stronger private sector involvement in South-eastern Asia and the lowest in Western and Central Asia. Overall, academia's strong leadership in advancing Responsible AI is evident across Asia, except in Southern Asia where civil society playing a particularly strong role.

Government frameworks/actions vs non-state actors' initiatives

The chart highlights thematic areas in Asia where non-state actor (NSA) initiatives outnumber governmental frameworks or actions. This underscores the vital role of civil society, academia, and the private sector in advancing Responsible AI and contributing valuable expertise in key areas.

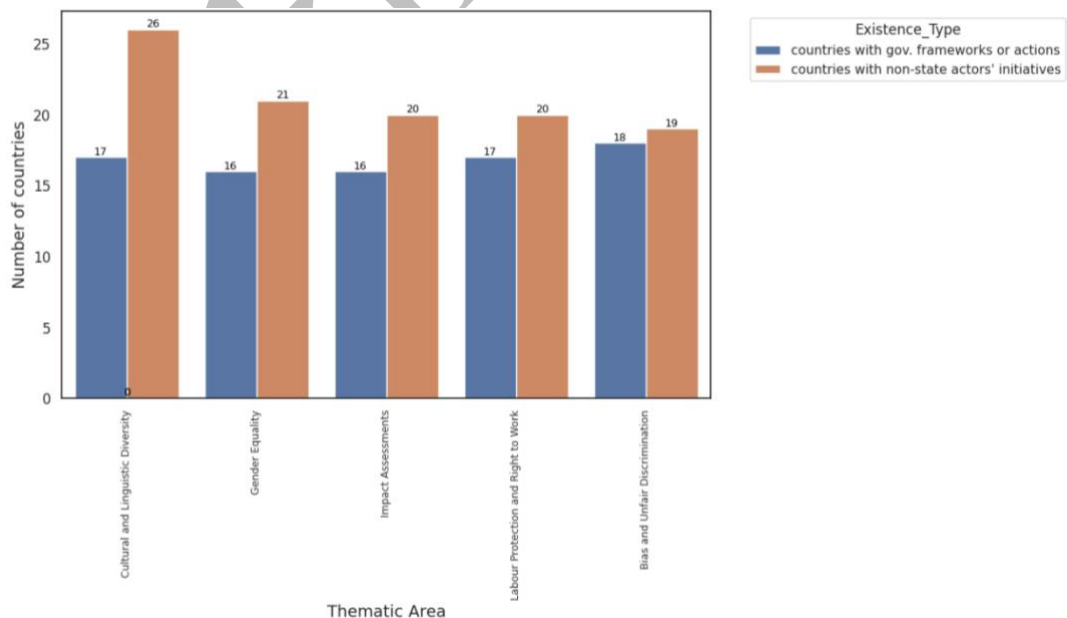


Figure 17: Thematic areas in Asia with more evidence of non-state actors initiatives than Governmental frameworks or actions

Cultural and Linguistic Diversity displays the largest gap, with 26 countries showing NSA initiatives compared to only 17 with government-led efforts. Gender Equality and Labour Protection and Right to Work follow similar trends, with 21 and 20 countries, respectively, having NSA involvement versus 16 and 17 countries with governmental initiatives.

Thematic areas like Impact Assessments and Bias and Unfair Discrimination show narrower gaps. For Impact Assessments, NSA initiatives are present in 20 countries compared to 16 with governmental frameworks. Similarly, in Bias and Unfair Discrimination, 19 countries have NSA involvement, slightly surpassing the 18 countries with government-led actions.

This evidence suggests that NSA initiatives are particularly prominent in areas related to social inclusivity and human rights, emphasizing the critical role of civil society, academia, and private sector actors in addressing gaps where government actions may fall short.

4) Conclusion

In conclusion, the analysis highlights both the strengths and gaps in Asia's Responsible AI landscape. While government actions in the region have been proactive, the development of formal regulatory frameworks lags behind, with significant disparities between subregions. Non-state actors, particularly academia and civil society, play a crucial role in advancing Responsible AI, especially in areas related to social inclusivity and human rights, and their influence varies across subregions, with some relying more heavily on state-led initiatives. However, different sub-regions would benefit from more engagement from different types of non-state actors. Moving forward, filling regulatory gaps, forging multi-stakeholder collaboration, engaging regional coordination and aligning governance practices with global standards—while respecting local contexts—will be essential in shaping a more inclusive, effective, and responsible AI ecosystem across Asia.

Part III: Case Studies

To capture the diversity of Responsible AI governance in Asia, this section examines case studies from Singapore, the Philippines, India, and Sri Lanka, reflecting a range of economic sizes (Sri Lanka – small, the Philippines – mid-sized, India – large), AI adoption levels (Sri Lanka – low, the Philippines – moderate, Singapore – high), and maturity of Responsible AI practices (Sri Lanka – emerging, the Philippines/India – developing, Singapore – advanced). These cases highlight the varied challenges and governance approaches, from resource-constrained settings to advanced responsible AI ecosystems, offering a comparative view of how different countries balance innovation, ethics, and regulation in AI development.

Beyond these four Asian case studies, we also include New Zealand, which, while not geographically part of Asia, maintains strong economic and strategic ties to the region through trade agreements, digital economy partnerships, and participation in forums like APEC and RCEP. Despite its high human rights standards and strong democratic institutions, it is not a leader in AI adoption or Responsible AI governance, making it a relevant case for mid-sized economies with relative advantages but emerging AI policies. However, New Zealand is a pioneer in indigenous data sovereignty, with Māori-led frameworks offering valuable insights into how AI governance can address marginalized communities and data rights—a challenge many Asian nations also face.

Responsible AI in Singapore

LIRNEasia, Sri Lanka

1) Introduction

Country Overview

Singapore is a Southeast Asian state holding a population of around 6 million that mostly consists of Han Chinese, with Malay, Indian, and other ethnic minorities (The World Factbook, 2025). The country is a parliamentary republic, dominated by the People's Action Party and incumbent prime Minister Lee Hsien Loong's family from 1959 (Freedom House, 2021). Singapore is well known for its rapid post-independence, industrialization-driven growth from a low income country to its current status as a high income economy, built upon a "business-friendly regulatory environment and strong investments in infrastructure, education, healthcare, and public services" (The World Bank, 2024).

Some key points on human rights and democracy in Singapore. The country sees a modest performance of 6.2 on the Democracy Index in 2023 (on a scale of 1-10) (Our World in Data, 2023b), with an under par score of 0.37 on the Freedom of Expression Index in 2023 (on a scale of 0-1) (Our World in Data, 2023c) (the index "captures the extent to which people can voice their views and the media can present different political perspectives" (Our World in Data, 2023c)). On the Civil Society Participation Index, which indicates the "extent to which citizens are active in diverse organizations which choose and influence policy-makers...[ranging] from 0 to 1 (most active)" (Our World in Data, 2023a), Singapore sees a weak performance of 0.40, which helps explain to some degree our finding in this report of a civil society that lags behind government action vis-à-vis RAI.

A noteworthy area to set up our discussion of AI, is digital integration in Singapore. Digital is said to have contributed much to improving the quality of life of citizens across healthcare, education, and infrastructure (Ministry of Digital Development and Information, 2024). In 2022, the country ranked number 1 in digital inclusion (Economist Impact, 2022); Singapore's digital economy "contributed 17.7% to our [its] Gross Domestic Product in 2023, up from 13.8% in 2018" (Ministry of Digital Development and Information, 2024); and nearly "99% of our [its] government services can be completed online".

AI Innovation and Adoption

Driven by the vision "AI for the Public Good, for Singapore and the World" (Smart Nation and Digital Government Office, 2025a), Singapore is said to lead AI adoption in the region, with active integration of AI technology across state and non-state actors: into government functioning and services, business operations, and universities (Mobiloitte, 2024). Singapore is considered an "ideal hub for AI innovation" (The Financial Times, 2025) on account of its "stability and geographical location". The country has seen considerable investment in AI, committing "\$767mn) over the next five years to AI development" (The Financial Times, 2025) as part of its National AI Strategy 2.0, and "more than S\$500 million through AI Singapore (AISG) under the Research, Innovation and Enterprise (RIE) 2020 and 2025 plans" (Smart Nation and Digital Government Office, 2023, p. 7), which is predicted to benefit the country handsomely if leveraged well – "responsible adoption of generative AI could create an additional US\$10.3 trillion (S\$14t) in economic value by 2038" (Economic Development Board, 2024).

AI is “both commonplace and an uncommon force” (Smart Nation and Digital Government Office, 2023, p. 5) in the lives of Singaporeans. The main impetus for Singapore’s AI adoption drive was arguably the 2014 Smart Nation vision (Smart Nation and Digital Government Office, 2025b), which sought to spur broad digital adoption across Singapore – in government and a range of domains, from education to the environment, and the 2019 National AI Strategy, which saw the “establishment of about 150 teams working on research and development, and 900 startups exploring new ideas with AI... [and] significant breakthroughs in AI, resulting in the development of new products, enhanced capabilities, and transformative interactions” (Smart Nation and Digital Government Office, 2025a).

Today, Smart Nation 2.0, and the 2023 National AI Strategy (NAIS 2.0) continue to drive the leveraging of AI in numerous aspects of Singaporeans’ lives. More on NAIS 2.0 and its implications for RAI in the following section. Singapore has done much in the way of boosting AI literacy and skills among public servants. For example, the “Data & AI Literacy ePrimer” – initiated in 2019 (Singapore Government Developer Portal, 2023) and covering topics such as “data quality to visual analytics, machine learning, project scoping” – has been completed by over 90,000 public sector officers. The fruits of such efforts can be seen in how AI successfully drives a number of public services in Singapore, like “adaptive learning systems in our schools, and chronic health management systems in our hospitals... [and] immigration and customs clearance, and to detect and deter online scams” (Smart Nation and Digital Government Office, 2023, p. 8). We also see the harnessing of AI within Singapore’s public service to aid officers in various tasks. For example, Pair, an AI-driven chatbot designed to assist public officers across more than 100 agencies in tasks such as “emails, research, and idea generation” (Government Technology Agency of Singapore, 2025b); AIBots, a platform that allows agencies to “create customized Generative AI chatbots...add internal knowledge bases, configure the bots for their use cases, and share their bots with their teams” (Government Technology Agency of Singapore, 2025a).

Non-state actors also see significant AI adoption over a range of sectors (MobiLoitte, 2024), with government emphasis on AI adoption and development in sectors like finance and healthcare (for example, the S\$100 million committed by the Monetary Authority of Singapore to “support financial institutions in building capabilities in quantum and artificial intelligence (AI) technologies, as well as enable the advancement of quantum and AI related innovation and adoption in financial services” (Monetary Authority of Singapore, 2024). Adoption among businesses has seen a rise in 37% from 2021 to 2024. In finance, almost 80% Singapore’s financial institutions incorporate AI in efforts to detect fraudulent activity, and the banking sector offers AI chatbot-integrated customer services. In the healthcare sector, we see nearly 50% of hospitals using AI in detecting and diagnosing, along with an increase in use of 30% of “AI-powered telemedicine services”. The retail/e-commerce industry uses AI to offer more personalized experiences to customers and aid in managing inventories. Universities such as the National University of Singapore (2025b) and the Singapore Institute of Technology (2025) offers skilling in AI technology that would feed back into AI adoption across country.

Responsible AI

As stated previously, Singapore leads the region in Responsible AI, with an overall score of 53.77, well ahead of the *global* average 19.8 of (refer to Section 2 of this broader report for scoring methodology).

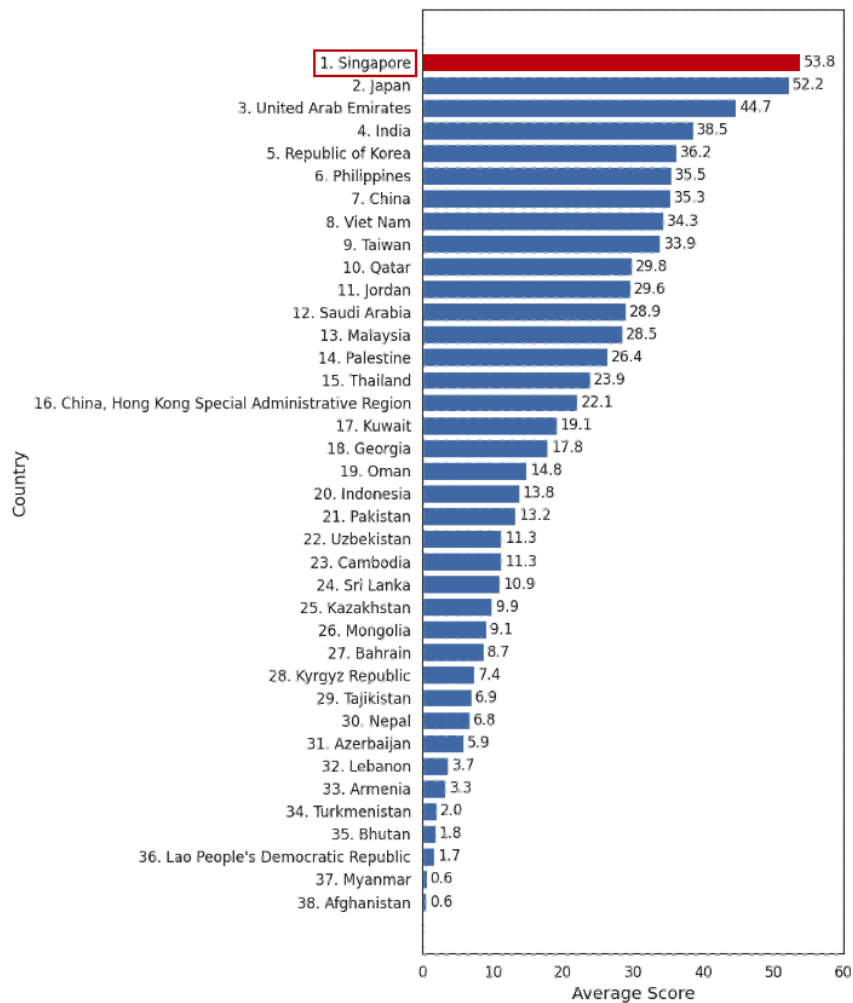


Figure 18: GIRAI country scores in Asia

PILLAR SCORE			DIMENSION SCORE		
Government frameworks	Government actions	Non-state actors	Human rights and AI	Responsible AI capacities	Responsible AI governance
43.70	70.68	40.09	48.35	52.10	58.54

Figure 19: Singapore RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 21)

Figure 18 indicates that RAI in Singapore sees by far the greatest push from the government; non-state actor score of 40.09 well above the global average shows that non-state actors also see substantial involvement. In terms of dimensions, Singapore sees the greatest emphasis on Responsible AI and Governance, with scores for the other two dimensions (Responsible AI Capacities and Human Rights and AI) that are comfortably ahead of their respective global averages.

2) Government Policy and Initiatives

National Strategy

Within the very wording of Smart Nation 2.0’s mandate – to drive the use of “technology more effectively to transform our [Singapore’s] future and shape our [Singapore’s] nation together” (Ministry of Digital Development and Information, 2024, p. 2) is implied an emphasis on RAI. It is recognized that if Singapore is to “continue harnessing technology as a force for good and maximize opportunities for our [its] people” (p. 4) “must also address risks that have arisen, such as challenges to our [its] individual wellbeing and social harmony”. Two outputs under the National Strategy worth highlighting are the National AI Strategy and the Model AI Governance Framework for Generative AI.

National AI Strategy 2.0

Launched in 2023, Singapore’s National AI Strategy 2.0 (NAIS 2.0) builds on its 2019 National AI Strategy and envisions “Singapore to be a place where AI serves as a force for good, and where we harness AI to uplift and empower our people and businesses” (Smart Nation and Digital Government Office, 2023, p. 3). This section discusses facets of NAIS 2.0 that concern RAI.

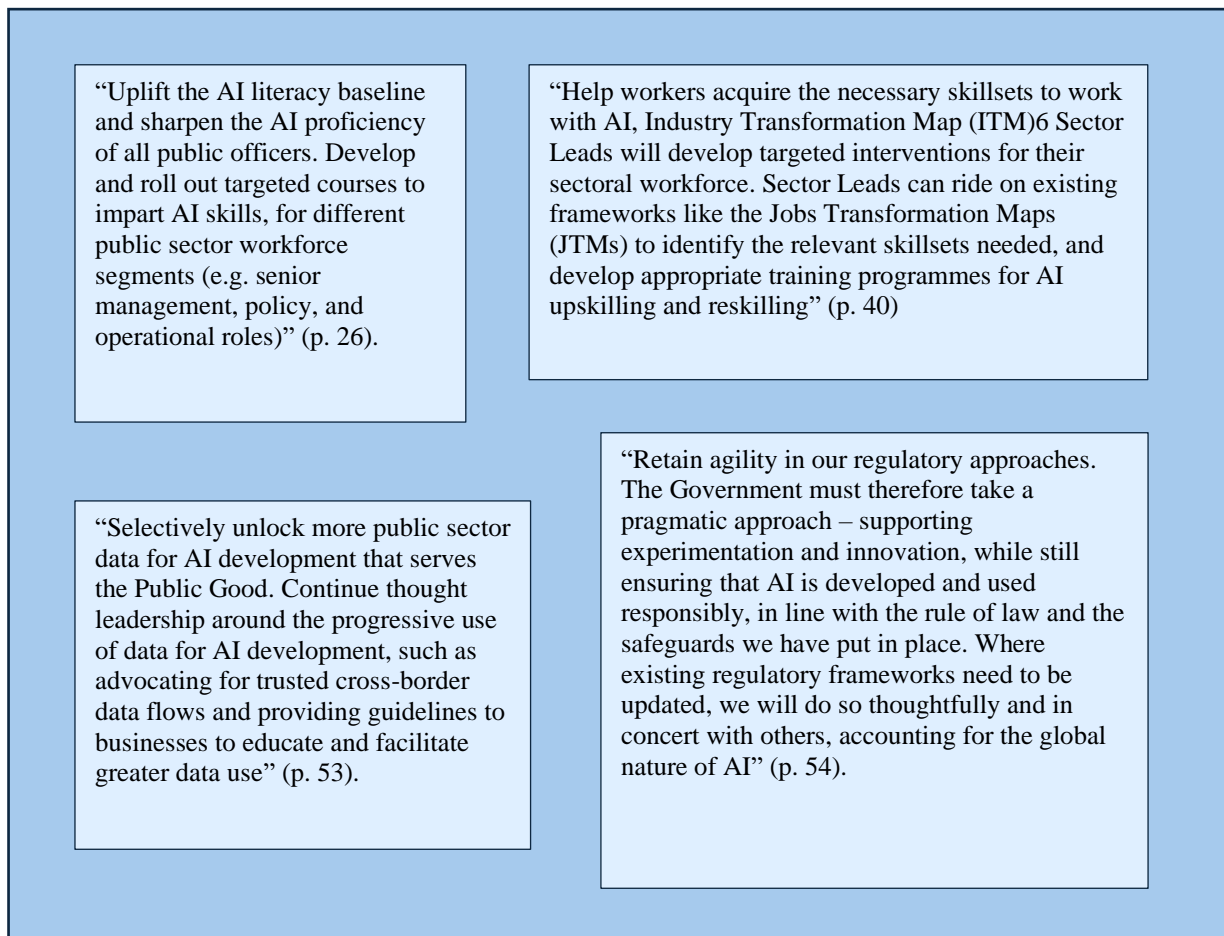
From the onset, the Strategy emphasizes its conception of AI as a “potent force for good, to uplift human potential” (p. 6) that must be leveraged in a “sustainable way to create positive impact – for new opportunities, better jobs, and safer, more meaningful connections” and manner that tackles “risks from the potential abuse and mismanagement of AI”. The Strategy recognizes the need for action in response to increasing AI capabilities and accessibility and worries regarding risks pertaining to AI, such as “malicious attacks on AI models, to the current inscrutability of large language models” (Smart Nation and Digital Government Office, 2023, p. 9).

The Strategy consists of three Systems, that perform the following functions (Smart Nation and Digital Government Office, 2023, p. 14):

System 1: Have industry, government, and public research work in concert around “meaningful use cases and problem statements to transform our economy and society”	System 2: Attract talent to work “with and from Singapore”. Scale “novel AI solutions” that will form part of a “toolkit” that organizations and individuals can use.	System 3: Make Singapore a “preferred site for AI development, deployment, and adoption” by establishing both the infrastructure and conducive environment for innovation.
--	---	--

Across these three systems, the Strategy proposes 15 Actions to be undertaken, e.g., Action 3: “Improve Public Service productivity, with new value propositions for our citizens” (p. 26); Action 8: “Upskill workforce through sector-specific AI training programmes” (p. 40); Action 11: “Build capabilities in data services and Privacy Enhancing Technologies” (p. 52). More concrete initiatives can be found under each Action.

Some of these initiatives are listed in the figure below (Smart Nation and Digital Government Office, 2023).



Model AI Governance Framework for Generative AI

The proposed 2024 framework builds on the Infocomm Media Development Authority’s (IMDA) existing 2019 first edition of the Model AI Governance Framework for Traditional AI framework, which is a detailed and readily-implementable guide for private sector organisations to address key ethical and governance issues when deploying AI solutions (Infocomm Media Development Authority, n.d.). That the Model AI Governance Framework for Generative AI is closely aligned with Singapore’s broader “AI for Public Good” vision mentioned previously reflects what is a very deliberate attempt to strategically situate RAI within AI initiatives across the board (i.e., AI innovation and integration does not occur divorced from RAI). The Framework consists of nine dimensions detailed in the graphic below.

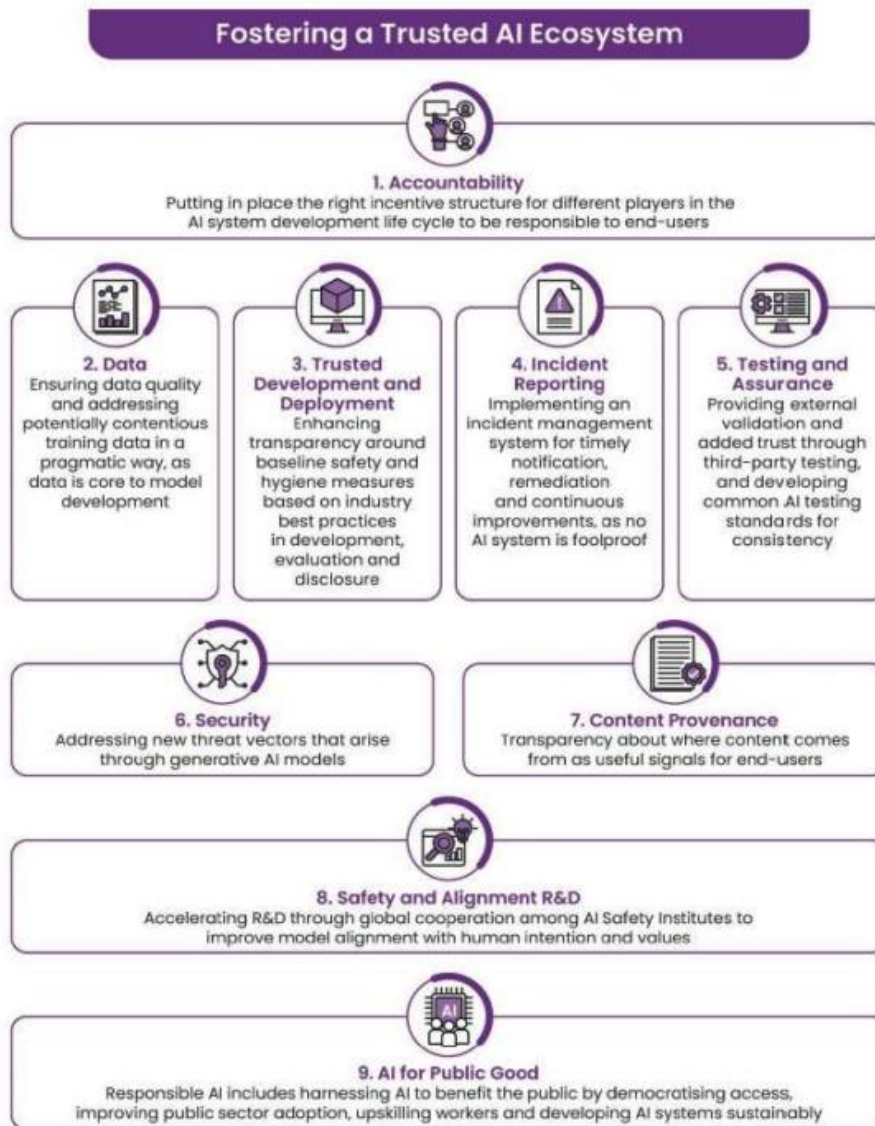


Figure 20: Nine dimensions of Singapore’s proposed Model AI Governance Framework (Infocomm Media Development Authority, n.d.)

AI Agency

AI Singapore (AISG) was launched in 2017 with the intention of “building a vibrant and sustainable AI ecosystem”. AISG connects research entities with AI-based start-ups and organizations to “share knowledge and develop products, talents and tools” (AI Singapore, 2024), to build capacities that will address issues in domains ranging from healthcare to the environment. AISG consists of five programs (AI Singapore, 2025): AI research, which seeks to “sustain investments in fundamental AI research to ensure Singapore’s relevance in the global AI race”; AI Governance, which seeks “research in governance, ethics and accountability of AI systems at the national level”; AI Technology, which supports applied AI research projects which are aligned to the National AI Strategy; AI Innovation, which builds “catalytic AI products/solutions to enrich Singapore AI Ecosystem”; and LearnAI, which seeks to “develop Generational AI Capability Programs to build AI Aware and AI Ready talents”.

RAI in AISG can be found in three broad areas (AI Singapore, 2025):

1. The technological developments and, research to support those developments, that produces AI solutions that advance RAI.

One example would be South East Asian Languages in One Network (SEA-LION), a “family of open-source Large Language Models (LLMs) that better understands Southeast Asia’s (SEA) diverse contexts, languages, and cultures”. SEA-LION aims to overcome biases in existing LLMs across “cultural values, political beliefs and social attitudes” due to training data being sourced from Western contexts, by providing models steeped in Southeast Asian languages and local contexts.

2. Direct research of principles of “fairness, transparency, and accountability” in AI solutions, “privacy and ethics”, and “risk and mitigations”.

Some examples of research grant awardees under the AI Governance program include a project that “examines the social impact of AI algorithmic management systems to motivate individual performance and drive productivity”, and an “online experiment involving respondents from Singapore to examine the institutions and regulatory arrangements that would enhance their trust in AI”.

3. Promotion of AI awareness and capacity building to make Singaporean citizens, policymakers and businesses AI-ready.

Two examples would be the AI Readiness Index, which helps organizations assess AI readiness across five pillars: organizational, infrastructure, data, business value, and ethics and governance readiness, and the AI Student Outreach Program to promote AI literacy among students above the age of 12

Laws and Regulations

Singapore takes a “sectoral approach” (Chng & Jones, 2024) to AI regulation, with agencies adopting “soft-law” approaches – i.e., “non-binding guidelines and regulations”. This subsection looks at these approaches across regulatory agencies in three sectors.

Financial Services

The Monetary Authority of Singapore (MAS) is said to be the first regulator to move on AI-based regulation (Chng & Jones, 2024). MAS released a set of “generally accepted Principles for the use of artificial intelligence and data analytics (“AIDA”) in decision-making in the provision of financial products and services” (Monetary Authority of Singapore, 2018, p. 3). The principles – “fairness, ethics, accountability, and transparency” (p. 6) (FEAT) – emphasize, among many things, the importance of minimizing bias in “ADIA-driven decisions” and ensuring accountability, ensuring AIDA use is compatible with an

organization’s own ethics codes, a means to appeal AIDA-related decisions, and “increase public confidence” (Monetary Authority of Singapore, 2018, p. 6) in AIDA. MAS also worked with partners in the financial industry to formulate Veritas, a “framework for financial institutions to promote the responsible adoption” (Monetary Authority of Singapore, 2019) of AIDA. The initiative will “provide financial institutions with a verifiable way to incorporate the FEAT principles into their AIDA solutions”.

Info-Communication and Media

Since 2019, Singapore’s Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission have released AI governance-related guides every year (Chng & Jones, 2024). Some examples include the Model AI Governance Framework at the World Economic Forum Annual Meeting in Davos (2019), which provides private sector organizations using AI tools with guidance on salient “ethical and governance issues” that they can readily apply.

A very notable initiative that is ubiquitous with Singapore’s RAI story is AI Verify, launched by the IMDA in 2022. AI Verify is an “AI governance testing framework and software toolkit” (AI Verify Foundation, 2024) that allows users to perform “standardized tests” to understand how their AI systems stack up against a “set of internationally recognized principles” and demonstrate that these systems are products of a development journey intended to “achieve the desired outcomes of these principles” (AI Verify Foundation, 2023). The output received from AI Verify tests can be used to prove that the entity has implemented RAI and “build trust with their stakeholders”, as well as to reveal possible gaps. AI Verify consists of three components. First, a testing framework that contains 11 principles aligned with existing international AI governance “principles, guidelines, and frameworks” (AI Verify Foundation, 2024) with “corresponding testable criteria and testing processes”. Figure 20 below contains the 11 principles (in purple text).

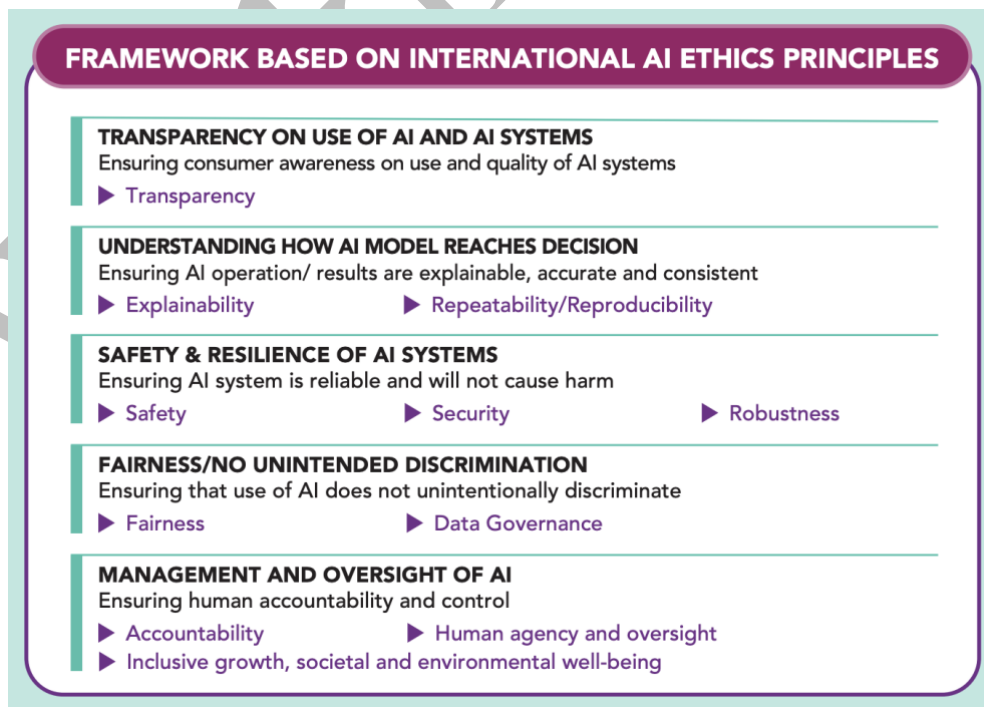


Figure 21: AI Verify Component 1: Testing Framework (AI Verify Foundation, 2024, p. 2)

Second, users access a toolkit, which serves two parties: users, and developers and researchers. Component 3 is a testing report that contains, among other things, outputs of the tests (AI Verify Foundation, 2023). The functions available to each party are as follows:

For Users

Use AI Verify

Complete a self-assessment or engage independent testing services to test AI model(s) against AI Verify framework

- Address 11 AI governance principles
- Self-assessment through process checks and technical tests
- Supports supervised learning AI models on tabular and image for:
 - Binary classification
 - Multiclass classification
 - Regression

For Developers & Researchers

Build on top of the AI Verify toolkit

Unlock the world of possibilities and make a meaningful impact in AI testing by contributing to AI Verify

- Report oriented design. Customisable canvas to design your own reports. AI Verify report template cannot be modified. User-customised reports will not be considered as AI Verify reports
- Build plug-ins. Third-parties can develop algorithms and components on top of AI Verify codebase
- Contribute to the project. Found a bug or want to add a feature? Contribute to the core codebase on GitHub

Figure 22: AI Verify Component 2: Toolkit (AI Verify Foundation 2023)

The very model of AI Verify as a soft-approach regulatory tool is laudable. As opposed to simply providing a set of non-binding principles, AI Verify goes a step further by also providing a clear incentive to entities to adhere to those principles through the self-assessment – the report serves as a badge or certificate of sorts which may help lend credibility to and increase trust in their AI systems.

Health

In light of the increasing integration of AI in Singapore’s health sector, Singapore’s Ministry of Health released AI in Healthcare Guidelines (AIHGIE) (Ministry of Health, Health Sciences Authority, & Integrated Health Information Systems, 2021). in partnership with the Health Services Authority and Integrated Health Information Systems to increase confidence in applications of AI in the sector and improve the safety of patients (Chng & Jones, 2024). AIHGIE delivers guiding principles that must dictate AI tools in healthcare – “fairness, responsibility, transparency, explainability, and patient-centricity” (Ministry of Health, Health Sciences Authority, & Integrated Health Information Systems, 2021, p. 6), lays out who the guidelines apply to (“developers” (p. 8) of AI-based healthcare tools; and implementers, who use those tools), how to use them. The document provides recommendations for developers at the design, build, and test stages of AI healthcare tools (p. 14); for implementers, recommendations when using, monitoring tools and responding to reports of issues with the tool, and conducting reviews of the tool (p. 28).

Other Government Actions

There is also interest in government in AI-based public-private partnerships (The Financial Times, 2025), with some attention paid to RAI. The director for startup ecosystem and community of government agency Enterprise Singapore was quoted as saying that “in collaboration with the private sector, we [they] want to help small and medium-sized enterprises (SMEs) develop AI skillsets”. Additionally, Singapore’s partnerships with major tech firms to boost AI adoption in businesses and equip workers with the necessary skills speaks to the kind of capacity building emphasized in its National Strategy. One example of a government-non-government entity partnership is the 2023 Singapore Conference on AI, a collaborative initiative between the Ministry of Communications and Information (MCI),

Smart Nation Group (SNG), and Topos Institute (a non-government entity) (SCAI, 2023). The conference brought together experts across government, industry, and academia, to “[explore] and [articulate] critical questions of AI that, if answered, will enable the development and deployment of AI for societies to flourish” (SCAI, 2023).

Singapore is also an active force in RAI at an international level, both multi and unilaterally. One example is Singapore’s leading of the development of ASEAN’s Guide on AI Governance and Ethics (The Financial Times, 2025). The guide aids regional organizations in the “design, develop, and deploy traditional AI technologies in commercial and non-military or dual-use applications” (The Association of Southeast Asian Nations, 2024, p. 3), laying out guiding principles (“transparency and explainability, fairness and equity, security and safety, human-centricity, privacy and data governance, accountability and integrity, robustness and reliability” (p. 2)). The guide dedicates an entire section to how organizations may operate in a manner supportive of RAI, which advises organizations on how to set up internal governance structures to “incorporate values, risks, and responsibilities relating to algorithmic decision-making” (p. 18), deciding to what degree human’s play a role in “AI-augmented decision-making” (p. 23), important considerations at each point of the “AI Lifecycle” (p. 28) and measures to develop “trust with stakeholders throughout the design, development, and deployment of AI” (p. 41)

Some unilateral engagements are as follows. In November 2024, Singapore and the United Kingdom established a Memorandum of Cooperation (MoC) to “enhance the safety, reliability, and responsible development of artificial intelligence (AI) technologies” (Sharon, 2024). The MoC consists of four facets: “AI Safety Research...Global Norms and Standards...Information Sharing...[and] Comprehensive AI Testing”. Singapore also holds a Memorandum of Understanding with Australia to “help progress the safe and responsible uptake of AI” (Department of Industry, Science and Resources, 2024). The MoU serves as a guide on applying AI to civil affairs: “encouraging information sharing...facilitating increased access to AI technologies, market and talent...support for the commercialisation of AI applications...prioritising a human-centric approach to AI” (Department of Industry, Science and Resources, 2024).

3) Non-government Initiatives (Industry, Civil Society, Academia)

While not as active as the Singaporean government – as indicated by the relative differences in government vs. non-state actor scores recorded by the GIRAI – non-state actors in Singapore certainly do play an active role in Singapore’s RAI efforts.

The Artificial Intelligence Institute in at the National University of Singapore (NUS) is a central hub for AI researchers at NUS (National University of Singapore, 2025b). The institute has baked into its mandate an emphasis on RAI, conducting “basic and applied research” as well as research on “societal implications of AI”, with the aim to – beyond AI innovation alone – “explore the use of AI to improve lives, empower communities, and create a better future for all” by studying the “the policy and societal implications of AI in order to mitigate its risks and ethical concerns” to help “harness AI for public good”. To this end, the institute has AI Governance and Policy as a dedicated domain, that asks the following broad questions: “How can we reap the benefits of AI while minimising or mitigating the risks? How can we ensure that those benefits are equitably distributed? How can we manage the inevitable disruptions that AI will cause to society, to the economy, to our political systems?”

(National University of Singapore, 2025a). While the institute is yet to see substantial research done under this domain, some RAI-related projects include: development of “an XAI Evaluation toolkit for automatic, scalable evaluation of XAI [explainable AI] techniques under simulations” (National University of Singapore, 2025c) (the rationale being that the ability to explain how AI systems work is a key part of helping users “understand and trust AI decisions” and use AI responsibly); a project that explores “digital footprint removal and its consequences without affecting existing AI systems while providing ways to check for compliance” (in support of the data privacy principle that people have the “right to be forgotten” and have their data removed from organization records).

Another example of an RAI-based academic initiative is an international, multi-university research project in which Dr Jasper Roe FHEA, Head of Department Language School at James Cook University in Singapore, was involved (James Cook University, 2024). The research paper outlines a “practical, simple, and sufficiently comprehensive tool to allow for the integration of GenAI tools into educational assessment: the AI Assessment Scale (AIAS)” (Perkins et al., 2024). The tool “empowers educators to select the appropriate level of GenAI usage in assessments based on the learning outcomes they seek to address”.

It appears that civil society lags behind the extensive RAI work done by government, industry, and universities in Singapore. Civil society typically plugs gaps in matters of governance and socioeconomic affairs where formal institutions lack or fail in their roles, and since those other entities have made substantial advancements in RAI, there is less of an onus on civil society to perform here, particularly given the existence of an environment that is somewhat stifling for civil society (CIVICUS, 2024). This isn’t to say that civil society should not play a role in RAI; indeed, it is recognized that civil society ought to become a “critical stakeholder in global efforts to expedite the benefits of AI and mitigate against its risks” (Sanchez, 2021). The Computer Society is Singapore’s “leading infocomm and digital media society for industry professionals, leaders, students, and tech enthusiasts” (Singapore Computer Society, 2025). The body collaborated with the IMDA in 2024 to release the AI Ethics & Governance Body of Knowledge Version 2.0 (BoK 2.0) (Lai, 2024). Building on 2020’s Bok 1.0, the framework tackles “practical issues related to human safety, fairness, privacy, data governance and general ethical values in AI deployment”. The framework has as its foundation four pillars that concern “ethical challenges” vis-à-vis AI adoption:

1. **Internal Governance** – BoK 2.0 underscores the importance of adapting existing governance structures or establishing new ones to incorporate values, risks and responsibilities related to algorithmic decision-making.
2. **Operations Management** – This pillar focuses on the operational aspects of AI development, including data management, model selection and maintenance.
3. **Human-centricity** – Recognising the critical role of human judgment in AI-augmented decision-making, BoK 2.0 introduces a methodology to help organisations determine their risk appetite for the use of AI and the appropriate level of human involvement.
4. **Stakeholder Communications** – Effective communication with stakeholders is crucial for responsible AI governance and this pillar provides strategies for managing relationships with an organisation’s stakeholders to foster a collaborative approach to AI governance.

Figure 23: The Four Pillars For Ethical AI Adoption, Bok 2.0 (Lai, 2024)

The Law Reform Committee at the Singapore Academy of Law released a report that discusses “issues that law and policy makers may face in promoting ethical principles when reforming laws and regulations to adapt to AI, and offering examples of human-centred approaches that could be taken to address these” (Singapore Academy of Law, 2020). The report considers 8 principles: respecting fundamental interests, considering effects, wellbeing and safety, managing risks to human wellbeing, respect for values and culture, transparency, accountability, and the ethical use of data, and finds that, on account of how varied AI systems are, there is no “one-size-fits-all regulatory solution” (Singapore Academy of Law, 2020, p. 1).

4) Key Trends & Outlook of Responsible AI in the country

Singapore’s impressive rank on the GIRAI speaks for itself. We’ve discussed how much of Singapore’s RAI regime is driven by strong government intent and action to ensure 1) strong AI integration in (specifically also the public sector, industry & academia) institutions, organizations, and individual citizens’ lives; 2) design, development and implementation of AI systems in a manner that socioeconomically profits Singapore and its citizens. This section offers a few final thoughts on RAI in Singapore.

First, civil society RAI participation. We have established that civil society plays a relatively minimal role in RAI compared to government, which is understandable, given 1) the strong involvement by government in RAI and fewer obvious gaps in governance for civil society actors to plug themselves; 2) a relatively stifling operating environment. However, there is still a strong need for active civil society involvement in RIA (beyond any participation in the many public consultations offered by the government) so that actors have the capacity to be ready to address/dissent in the event gaps do appear. While the Singaporean government’s RIA action is laudable, in any authoritarian-leaning state, civil society may be the only bulwark against potential harmful regime policy. Despite emphasizing the importance of strong civil society involvement in RAI, this report does not explore how civil society actors can navigate the current operating environment to achieve this, as it lies beyond the scope of the report.

Second, in terms of international cooperation, Singapore could expand on its existing practice of engaging with other countries on RAI. For example, the Digital Economy Partnership Agreement (New Zealand Ministry of Foreign Affairs and Trade, 2025) between Singapore, Chile, and New Zealand, which contains a module on AI to “recognize the importance of developing frameworks for the trusted, safe, and progressive use of AI” (New Zealand Ministry of Foreign Affairs and Trade, 2019), which involves knowledge sharing (including “regular dialogues to share Singapore’s expertise in developing model AI governance framework”), aligning principles and practices, and “research collaborations on AI governance and ethics”. Given the relatively lower regional scores recorded in the GIRAI, it would be worth considering how Singapore might use existing organizations such as ASEAN to promote RAI.

Responsible AI in The Philippines

LIRNEasia, Sri Lanka

1) Introduction

Country Overview

The Philippines is a presidential republic located in Southeast Asia, home to a culturally diverse population of over 110 million (The World Factbook, 2025). Of the over 120 languages spoken in the country, Filipino and English are the key main languages governing socioeconomic activity. Catholicism is the most subscribed to religion.

In the realms of human rights and democracy, the Philippines sees a modest performance, scoring a 6.7 on the Democracy Index in 2023 (on a scale of 1-10) (Our World in Data, 2023b), with a 0.66 on the Freedom of Expression Index in 2023 (on a scale of 0-1) (which “captures the extent to which people can voice their views and the media can present different political perspectives”) (Our World in Data, 2023c). Civil society in the Philippines sees some agency, with a score of 0.65 in 2023 in the Civil Society Participation Index (Our World in Data, 2023a) indicating that the country’s civil society operates somewhat actively “in diverse organizations which choose and influence policy-makers”.

The country’s economy – considered one of the “most dynamic economies in the East Asia and Pacific region” (The World Bank, 2024) – is driven by a strong BPO industry (Remulla, M. M., & Medina, G. M., 2012; Ramotowski, 2025), consumer demand, substantial remittances from migrant workers (Philippine Institute for Development Studies, 2024), a fairly young population, and a strong labor market and buoyant private sector. A strong network of universities and a robust ICT industry (International Trade Administration, 2024) make the country well positioned to participate meaningfully in AI development and see broad AI integration in products and services across the country.

AI Innovation and Adoption

The Philippines has taken solid steps in the realms of AI adoption and development, with substantial discourse among multiple stakeholders extolling the virtues of AI for Filipinos. Before proceeding to RAI, it is worth noting the relative nascency of the Philippines’ AI innovation and adoption enterprises in comparison to global leaders, and the reasons behind this. The Philippines has not developed or integrated advanced AI systems into government, industry, and services to the same extent as higher-income nations with greater economic and technological resources, due to the following barriers: (Department of Trade and Industry, 2024a): limited understanding of data science and AI across organizations; limited knowledge on possible use cases of AI; limited resources; the absence of a data strategy that considers “how acquiring technologies and technical capabilities can help enterprises achieve their strategic goals” (p. 6); the lack of a “clear legal and regulatory framework for AI” (p. 7) that will “both protect the rights of Filipinos and spur responsible adoption of and innovations in AI”; and underdeveloped data quality and infrastructure (which was found during Index research on the Philippines). There is, however, growing interest in AI across government, industry, and civil society.

There are a number of instances of AI adoption across government. The Department of Trade and Industry launched the National AI Strategy Roadmap 2.0 (NAISR 2.0), and Center for AI

Research (CAIR), the country's first AI hub the country's first AI hub, recently in July 2024. The roadmap proposes multiple objectives to accelerate AI integration in the Philippines, such as “new technologies like generative AI to harness new opportunity for the Philippine economy and enhance the lives of its citizens” (Department of Trade and Industry, 2024c), while CAIR seeks to create “AI solutions for regional concerns notably sustainable agriculture, urban planning, and disaster resilience” (Department of Trade and Industry, 2024b). The Department of Science and Technology launched 9 AI-powered projects in 2021 (Arayata, 2021) – ranging from Chatbots to monitor public school student health, to a system to aid “traffic control and management” – recently begun integrating AI in its weather forecasting process to improve accuracy (Dharmaraj, 2024), and the Philippines' Commission on Audit now incorporates AI to flag anomalous government transactions (Trinchera, 2024). While it was found during Index research on the Philippines that the government previously created a ‘hands-off’ enabling policy environment in that it did not through regulation heavily impede AI innovation by non-state actors, bills pertaining to AI have been filed in Congress (Bañez, 2024), such as Bill 10567 (the “Deepfake Accountability and Transparency Act” (Data Guidance, 2024), which “mandates disclosures for deepfakes distributed over the internet”. The government has also signaled willingness to work with non-state actors to help shape AI's direction in the country. For example, government agencies partner with non-state actors to host dialogues on AI in the Philippines, and the government signed a memorandum of understanding with a California startup on AI weather forecasting, specifically to “use machine learning model of historical weather data to produce...[weather] forecast[s]” (Trinchera, 2024).

The private sector has also made some progress in AI adoption; for example, firms have begun using basic AI technologies like customer service chatbots online that leverage basic generative AI. However, while there appears to be broad interest in AI among the private sector – for example, in a study among Filipino businesses, “74% of respondents believe AI tools can increase work efficiency” (Yondu Team, 2024) – adoption among businesses has been slow, with a little over half of surveyed Filipino CEOs (54%) “report[ing] that their organizations have yet to integrate AI tools”, despite around 80% agreeing that, within the next three years, AI technologies will necessitate re-skilling among their workforce (that around half of surveyed CEOs have already integrated AI tools is certainly commendable). Efforts in academia, such as the University of the Philippines Diliman's College of Engineering degree programs that build capacity in AI technology (UPD College of Engineering, 2025), may further assist in the adoption of AI systems if increased AI-based capacities – and thus a more AI-familiar society – in turn increase the demand for AI-integrated processes and services in government and private sector.

Responsible AI

The graph below shows the Philippine's overall score in Responsible AI within Asia: leading the pack with an overall score of 35.2, and comfortably ahead of the *global* average 19.8 of (refer to Section 2 of this broader report for scoring methodology).

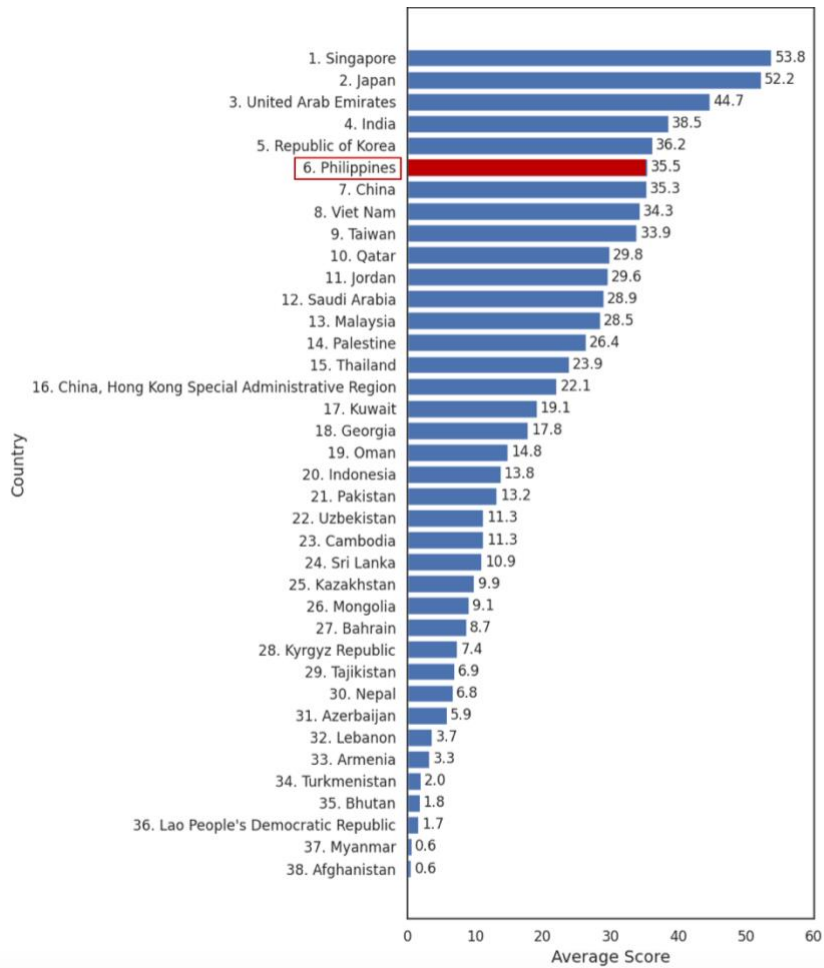


Figure 24: GIRAI country scores in Asia

What does this mean, substantively? The nascency of AI adoption and development in the Philippines naturally extends to the Philippines’ progress in Responsible AI: there are ample discourse and modest steps towards concrete action among state and non-state actors (which we discuss in detail in the next two sections).

PILLAR SCORE			DIMENSION SCORE		
Government frameworks	Government actions	Non-state actors	Human rights and AI	Responsible AI capacities	Responsible AI governance
19.24	40.72	57.36	34.87	29.67	37.84

Figure 25: The Philippines RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 22)

As Figure 24 indicates, RAI efforts in the Philippines are largely non-state actor driven, with a greater focus on the realms of human rights and AI and responsible AI governance, than responsible AI capacities (for further clarity on scoring, refer to Section 2 of the broader report). From the government, there exists a National AI framework (NAISR 2.0) and legislature that indirectly apply to RAI (i.e., while not explicitly formulated with RAI as the primary purpose, does cover some facet of RAI) like the Data Privacy Act of 2012. In terms of concrete, binding documents, it appears Responsible AI is currently conceived of from a digital privacy lens, with no regulation targeting AI technology or technical standards. Non-state actors from academia to the private sector have published guidelines on the use of AI and organize events on AI governance and engage in conferences and panels. The next two sections discuss some of these initiatives in depth.

2) Government Policy and Initiatives

National Strategy

The most noteworthy RAI-related government initiative is the Philippines’ National AI Strategy Roadmap 2.0, (NAISR 2.0), released in July 2024 and, building off version 1.0 released in 2021. Granted, it is only a roadmap; nonetheless, it is the most concrete articulation of the government’s RAI vision to date.

At the time of writing, however, the only publicly available roadmap was that of NAISR 1.0 (an 18-page document released by the Department of Trade and Industry). Because NAISR 2.0 builds upon its predecessor, there is utility in discussing what NAISR 1.0 looks like, and what it has to say both directly and indirectly about RAI. NAISR 1.0 is structured as indicated in Figure 25, with two broad **pillars**, implementation and innovation, branching out into 4 **strategic dimensions**, 7 strategic imperatives, and 42 **strategic tasks**.

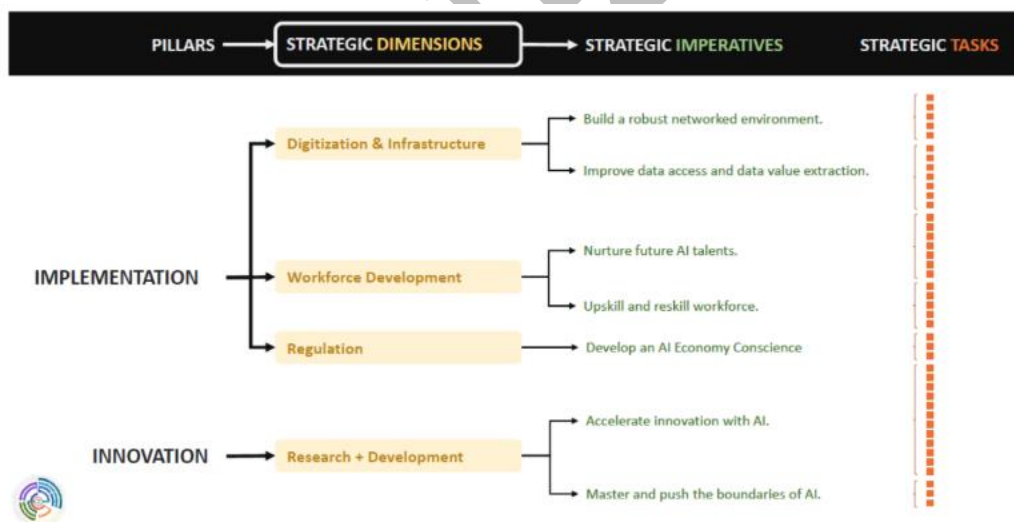


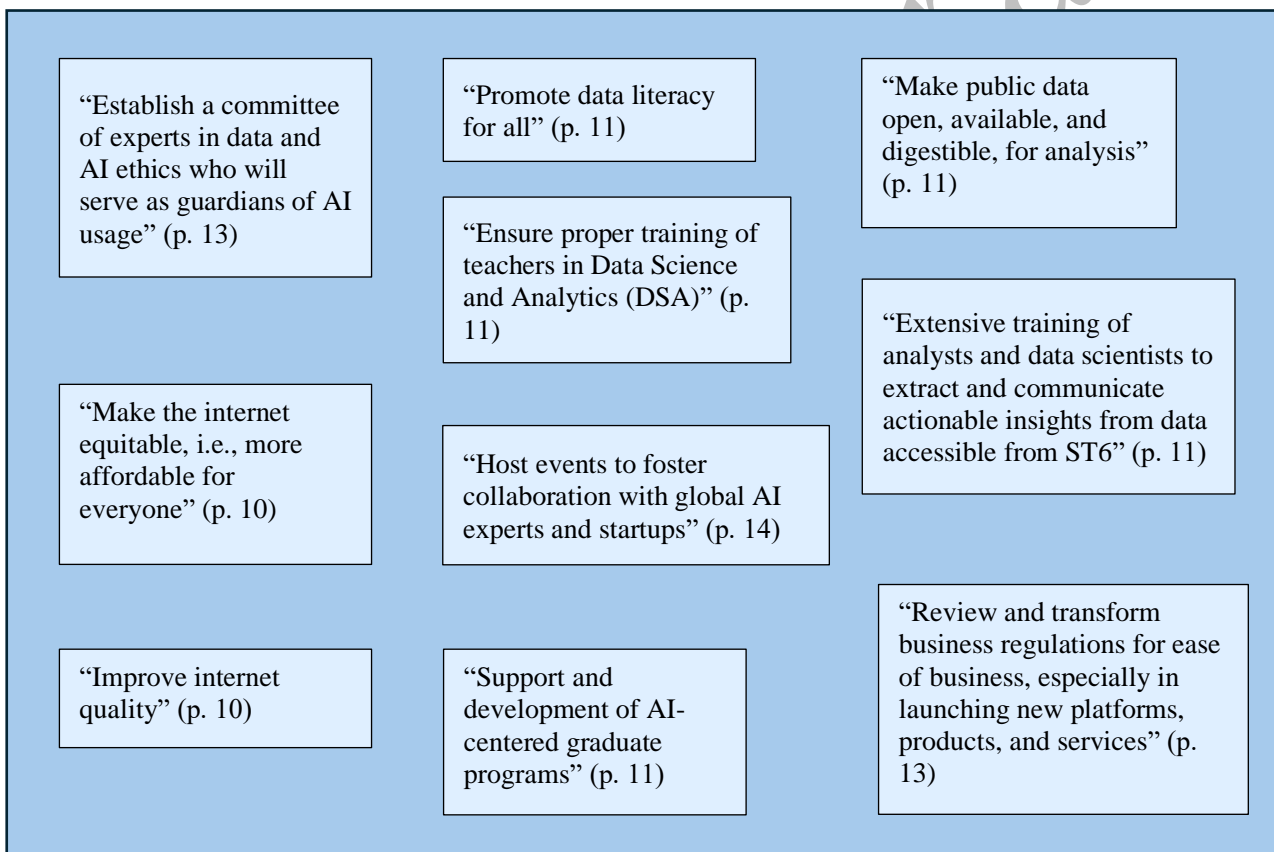
Figure 26: National AI Strategy – Pillars, Dimensions, Imperatives, Tasks (Department of Trade and Industry, 2024a, p. 9)

The roadmap’s main objectives are as follows:

1. “To maintain the regional and global competitiveness of Filipino industries, noting that AI is one of the biggest drivers of innovation for enterprises; (p. 8)

2. To identify key areas (in both R&D and technology application) for investing both time and resources of government, industry, and society, which includes harnessing existing global knowledge and new technologies that will aid in the development of new processes, products, and services for increasing productivity and for promoting overall public welfare; (Department of Trade and Industry, 2024a, p. 8)
3. To recommend ways for effectively fostering a triple-helix (R&D) collaboration among government, industry, and the academe, which is essential to national development;
4. To suggest approaches for preparing the future workforce for the jobs of the future; and
5. To attract the biggest industries to set shop in the country, and thus generating more jobs for the Filipino people” (8)

While NAISR 1.0 does not refer explicitly to RAI, it contains within it, numerous imperatives that pertain to RAI. A read through the 42 strategic tasks tells us what RAI-specific provisions the roadmap contains. Some of these provisions are as follows:



Broadly, these strategic tasks cover ethics and equality, capacity building among students, teachers, and practitioners, dialogues and partnership, and regulation conducive to doing business. How they will be actualized remains to be seen.

While official sources report that NAISR 2.0 “addresses emerging themes such as ethics and governance” (Department of Trade and Industry, 2024b), it is unclear whether this means a reiteration of or expansion upon existing related points in NAISR 1.0.

AI Hub

The key role of the recently launched Center for AI Research (CAIR; the AI hub mentioned previously that was launched alongside the roadmap) is to “harness AI’s transformative power to address pressing societal and industrial challenges, fuel economic growth, and promote inclusive development” (Department of Trade and Industry, 2024c). The primary focus of the hub would be research and development to advance AI technological innovation, exploring “AI, Machine Learning, and GenAI solutions from Amazon Web Services” with the goal of augmenting services available to citizens.

It is also intended that efforts of the Center for AI Research in “technological innovation, multidisciplinary and cross-disciplinary research, and the development of full-time research scientists, engineers, and R&D personnel” (Department of Trade and Industry, 2024b) will see in parallel the balancing/ensuring of the responsible adoption of AI technologies developed through those initiatives. As with the roadmap, concrete RAI work will become apparent as CAIR commences projects and rolls out AI-based R&D.

Laws and Regulations

As stated previously, the Philippines lacks legislation that directly pertains to AI, and RAI is largely viewed through a data privacy lens, with no regulation targeting AI technology or technical standards. This report discusses the four laws in the Philippines with provisions that hold reasonable links to RAI.

First, the Republic Act No. 10173, or the Data Privacy Act of 2012, aims to “protect the fundamental human right of privacy, of communication while ensuring free flow of information to promote innovation and growth”. Among the rules and regulations promulgated to support the implementation of the Act’s provisions are the following (quoted verbatim) (The Lawphil Project, 2012):

Section 34. Rights of the Data Subject. *The data subject is entitled to the following rights:*

a. Right to be informed.

- 1. The data subject has a right to be informed whether personal data pertaining to him or her shall be, are being, or have been processed, including the existence of automated decision-making and profiling.*
- 2. The data subject shall be notified and furnished with information indicated hereunder before the entry of his or her personal data into the processing system of the personal information controller, or at the next practical opportunity:*

(f) Methods utilized for automated access, if the same is allowed by the data subject, and the extent to which such access is authorized, including meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject;

c. Right to Access. *The data subject has the right to reasonable access to, upon demand, the following:*

- 6. Information on automated processes where the data will, or is likely to, be made as the sole basis for any decision that significantly affects or will affect the data subject;*

Section 48. Notification of Automated Processing Operations. *The personal information controller carrying out any wholly or partly automated processing operations or set of such*

operations intended to serve a single purpose or several related purposes shall notify the Commission when the automated processing becomes the sole basis for making decisions about a data subject, and when the decision would significantly affect the data subject.

b. No decision with legal effects concerning a data subject shall be made solely on the basis of automated processing without the consent of the data subject

What does all this mean? Simply put, data subjects 1) have the “right to access information on automated processes where (such automatically processed) data “will, or is likely, to be made as the sole basis for any decision significantly affecting, or will affect, the data subject”” (Ipac, 2024); 2) may “object to “automated processing where the personal data will, or is likely, to be made as the sole basis for any decision that significantly affects, or will affect, him or her””.

Second, the Republic Act No. 11930 – the Anti-Online Sexual Abuse or Exploitation of Children and Anti-Child Sexual Abuse or Exploitation Material Act (The Lawphil Project, 2022), which prohibits AI-generated deepfake videos that are pornographic in nature (quoted verbatim):

Section 3. Definition of Terms. — *As used in this Act:*

(j) Image-based sexual abuse (ISA) refers to a form of technology-facilitated sexual violence. The term describes a pattern of behavior involving the nonconsensual creation, distribution, or threats to distribute nude or sexual images. It includes a diversity of behaviors including, but not limited to, "sextortion scams," the use of artificial intelligence to construct "deepfake" pornographic videos, threats to distribute photographs and videos; and the taking or sharing of sexual assault imagery;

Finally, two laws – the Republic Act No. 11927 – or the Philippine Digital Workforce Competitiveness Act (The Lawphil Project, 2022) – and the Republic Act No. 11899 – or the Second Congressional Commission on Education Act (EDCOM II) – were enacted in response to the threat of AI-based automation-induced labor market disruptions to prioritize digital transformation in education and upskill the workforce (Ipac, 2024). This concerns RAI, because it is intended to curb the kind of socioeconomic inequalities that layoffs and widening skill gaps may create with increasing AI automation.

The National Privacy Commission released guidelines on the use of AI to ensure compliance with the 2012 Data Privacy Act (National Privacy Commission, 2024). The guidelines, among many things, require that organizations make data subjects aware of the purposes, risks and impacts of AI systems, and stresses the need for regular compliance checks.

Other Government Actions

The government is also engaged in partnerships with non-state actors to advance RAI. For example, the Department of Information and Communications Technology partnered with the International Telecommunications Union to co-organized an event called “AI Dialogue: Gender-Based AI Policy in the Philippines” (Department of Information and Communications Technology, 2023). The event covered, many topics, a Gender and AI Guideline to ensure the development of “gender-safe products”, the need to “invest in women in AI networks, training, company development, and policy participation”, and the importance of “advocating for open data and open science, increasing compliance with sex-disaggregated data in national databases”.

3) Non-government Initiatives (industry, Civil Society, Academia)

As highlighted in Part II of this report, South-eastern Asia has the most balanced involvement of non-government actors, and the Philippines lends credence to that claim.

Academia

In addition to its work in AI innovation, the University of the Philippines has taken multiple steps in support of RAI. The university released 15 Principles for Responsible and Trustworthy Artificial Intelligence (University of the Philippines, 2025) to “provide guardrails and indicate the way forward on the development and use of AI in the University and the country”. Broadly, the list emphasizes that AI developments must serve the Filipino people, foster inclusive economic growth, democracy and the environment; empower the marginalized and vulnerable; ensure human control over AI systems; be fair, transparent, and safe; aid education efforts and be incorporated into university curriculums. The University will also create two bodies 1) an AI Advancement Committee whose RAI actions include drafting an AI code of conduct, implementing AI capacity-building among faculty and staff, and ensure equitable access to AI among its members; and 2) an AI Advisory Board that will “advise the Board of Regents, through the UP President, on matters related to responsible and trustworthy AI”. In August 2023, Ateneo De Manila University released a memo on the creation of a GenAI Task Force to “chart the University’s stance and aspirations on Generative AI tools in connection with [its] mission in Education” (Ateneo de Manila University, 2023). The responsibilities of the Task Force include creating a Course Policy on the use of GenAI, AI capacity building, and training students to be “ethical and humane users of AI”.

Private Sector & Industry

The private sector and industry are also active in this space. Aboitiz Data Innovation is a firm working in the AI adoption space – specifically, at the intersection of AI and data in a business context– that produces and implements generative AI-driven platforms with applications in many industries (Aboitiz Data Innovation, 2024). ADI was a signatory to a joint publication that provided comments on the University of the Philippines AI Principles discussed above (Chua et al., 2023). The Analytics & Artificial Intelligence Association of the Philippines (AAP), the Philippines' National Industry Board for Analytics and AI, holds events (Analytics & Artificial Intelligence Association of the Philippines, 2025) and publishes reports (Ng, M., Khoo, M., Haridas, G., & Toh, J. T., 2023) on matters pertaining to RAI. A recent AAP report advocated for governments and business adoption of “guardrails needed to minimize risks and prevent the harmful use of [AI]” (p. 23); specifically, RAI frameworks to prevent “discrimination, bias and malicious use”, and business risk assessments ahead of implementing generative AI. Connected Women is a private enterprise that provides AI-based skilling for women. For example, their flagship program, Elevate AIDA, “upskills women for work as Data Annotation Specialists in the AI industry” (Dunuan, 2023), prioritizing “women from low-income and marginalized sectors, such as single mothers, returning OFWs (overseas foreign workers), unemployed housewives, disadvantaged youth, and persons with disabilities”.

Civil Society

The Ambit, a “global civic and social organization [founded in the Philippines] converging the Global Majority/Global South nations at the center of ethical AI governance” (Chua et al.,

2023) conducts research impacts of AI on the marginalized and disenfranchised, networking with stakeholders, policymaker and research capacity-building, and advocacy for RAI. The coalition released a document in March 2024, titled “A Position on the Initial Consideration of Artificial Intelligence Development and Regulation Act of the Philippines”. Two noteworthy propositions relevant to RAI are as follows: 1) advocacy for the “establishment of a High-Level Advisory Body on AI Governance” to guide AI regulation and policies” (Chua & Castillo, 2024, p. 4) and develop an AI governance framework that “addresses the challenges posed by deepfakes and other AI systems, and align the Philippine AI ethics framework with international standards”; 2) the assertion that “Policies should not only protect tenured employees but also extend those in the gig economy and in the informal sector who are increasingly vulnerable to job displacement caused by AI” (Chua & Castillo, 2024, p. 5). The Ambit also published a petition in 2023 (The Ambit Coalition - Philippines Council, 2023) that put forth several propositions vis-à-vis RAI, three of which were 1) the establishment of a National AI Ethics, Safety, and Governance Committee; 2) the recommendation for provisions that protect the rights of vulnerable groups and children against risks posed by algorithms and generative AI; 3) the importance of independent oversight over the adoption and procurement of AI-based systems for public services.

4) Key Trends & Outlook of Responsible AI in the country

In summation, RAI is a burgeoning enterprise in the Philippines, with vibrant discourse among state and non-state actors signaling willingness to ensure the country’s developments in AI are accompanied by measures to guide those developments in a manner that is conducive to the wellbeing of all Filipinos and the country’s democracy. The next step would be the implementation of the various proposed strategies, frameworks, and best-practices, and the permeation of RAI principles into the design, development, and everyday use of AI among the public service, businesses, and individual citizens. This report ends with three key priorities pertaining to RAI.

First, the need for dedicated RAI policy. The problem with legislation that only indirectly relates to RAI forming the basis of a country’s RAI regime is that RAI efforts risk being framed based on the priorities of those indirect laws, which may in the long run see a less wholistic RAI regime. For example, it is said that, because the Data Privacy Act is the only substantial piece of legislation that advances RAI, Filipino law only really considers/seeks to guard against the harms of AI as it pertains to the processing of personal data. Much has been said and written about AI bias – when AI systems “produce biased results that reflect and perpetuate human biases within a society, including historical and current social inequality” (IBM Data and AI Team, 2023); the absence of legislation targeting the production of AI technology in the Philippines risks leaving the design and development stage open to the manifestation of such biases.

Second, it is imperative that discourse on capacity building and upskilling concern not only public sector and academia, but all citizens, so as to not further exacerbate existing socioeconomic inequalities in the Philippines (The World Bank, 2022). Citizens left without at least a basic understanding of the uses and hazards of AI, risk being left behind and unable to reap the benefits and circumnavigate the risks of AI in further technological developments and integration of AI into public services in the future.

Third, low levels of digital literacy and a widespread inability to determine the veracity of information leaves many Filipinos, and the country’s democracy, vulnerable to increasingly

sophisticated AI-based misinformation (Enriquez, 2024). There are calls from legislators to combat deepfakes, such as a bill filed to regulate the used of AI ahead of the 2025 elections (Porcalla, 2024). However, anti-disinformation legislature may have a chilling effect and facilitate political persecution by opponents of a governing party. Alternative solutions could include steps to increase media and digital literacy, and raise awareness on the principles of transparency, accountability, and truthfulness among election stakeholders (National Citizens' Movement for Free Elections (NAMFREL), 2024).

Working Draft

Responsible AI in India

Digital Futures Lab, India

1) Introduction

Country Overview

India is a federal republic governed by a democratic parliamentary system. With a population exceeding 1.4 billion, it has a multi-cultural, multi-linguistic, and multi-religious society with deep historical divides across religion, class, caste, and gender. India is the fifth-largest economy in the world by nominal Gross Domestic Product (GDP) and boasts the second-largest workforce globally. Despite this, the country grapples with persistent challenges, including unemployment, income inequality, and a significant reliance on informal labour markets (Inamdar, 2024).

The Information Technology (IT) sector has emerged as a cornerstone of India's economy, contributing substantially to GDP and employment (Jayswal, 2021). Over recent years, India has developed a thriving tech startup ecosystem, cementing its reputation as a global technology hub. This growth has been supported by government-led initiatives such as the National E-Governance Plan and the Digital India programme, which aim to promote e-governance and accelerate digitalisation across sectors. A critical aspect of India's digital transformation is the development of Digital Public Infrastructure (DPI). Initiatives such as Aadhaar (biometric-based identification), Unified Payments Interface (UPI), and the CoWIN platform for COVID-19 vaccination management exemplify India's approach to building population-scale digital technologies (Alonso et al., 2023). However, the benefits of digitalisation are unevenly distributed, with a significant digital divide persisting along gender, rural-urban, and socio-economic lines (Chaudhuri et al., 2024).

AI Development & Adoption in India

Artificial Intelligence (AI) in India is rapidly evolving, driven by a thriving technology ecosystem, government-led initiatives, and private-sector innovation. The adoption of AI in India has steadily grown in recent years, with adoption rates surpassing the global average (Hindustan Times, 2024). Numerous service-oriented sectors such as education, healthcare, and finance have rapidly integrated AI within their offerings. For example, over 60% of educators in India reported using AI for teaching, preparation, and student engagement (Mishra, 2024). In the banking and fintech sectors, there has been a marked proliferation of AI-powered conversation agents integrated into customer-facing services (Pal, 2024). The healthcare industry in India is also embracing AI technologies. Projections suggest that by 2025, AI could contribute an additional USD 25-30 billion to India's GDP through improvements in diagnostics, operational efficiencies, and patient care (Natarajan et al., 2021).

In the last few years, the Indian government has also shown a deep commitment towards catalysing AI development and adoption, with a key focus on integrating it within public services. A prominent example is the DigiYatra initiative of the Ministry of Civil Aviation. The initiative aims to make air travel hassle-free and paperless by using facial recognition-based biometric assessment to validate passenger identity. Even as privacy and efficiency concerns persist (Jain, 2022), DigiYatra has been implemented at least 24 airports in India, with plans for expansion in 2025 (Vaid, 2024). The Indian Railways Catering and Tourism

Corporation (IRCTC) — the nodal agency responsible for the administration of the state-owned railway system — has also integrated an AI-based virtual assistant named *AskDisha* as part of its ticketing platform (Times of India, 2024). State governments are also taking significant strides towards AI adoption, with the Andhra Pradesh government signing a Memorandum of Understanding (MoU) with Google for AI adoption in various sectors including healthcare (Business Standard, 2024b) and the Telangana government introducing a strategy document identifying the six critical pillars to enable better AI-driven governance in the state (Business Standard, 2024a).

India's growing enthusiasm for AI reflects its ambition to position itself as a global AI hub, driven by the strengths of its IT sector and large-scale digitalisation efforts. At the same time, however, India faces significant resource constraints, particularly in areas such as advanced computing infrastructure and access to large-scale, high-quality datasets.

One of the key bottlenecks for AI development is the lack of availability of data in the form of rich, standardised, well-annotated datasets that can be used to train AI models (Nandi & Yadav, 2024). In terms of computational power, India remains reliant on foreign companies — particularly NVIDIA — for access to specialised computer chips required for advanced AI computing (Bajwa, 2024).

To bridge many such gaps, the government has launched the IndiaAI Mission. With a substantial budget allocation of USD 1.25 billion, the mission aims to democratise computing access, improve dataset quality, and develop indigenous AI capabilities (Press Insider, 2024). It has seven key pillars, each dedicated to a specific component of the Indian AI ecosystem. These include computing capacity, application development, skilling, safety & trust, innovation, datasets, and lastly, startup financing (IndiaAI, 2025).

In 2021, the Government also announced the setting up of the National Language Translation Mission, which eventually led to the development of the *Bhashini* Initiative, an AI-led language translation platform dedicated to numerous Indian languages and dialects (Ministry of Electronics and Information Technology, Government of India, 2022b). Recently, the private sector has also focused on developing large language models (LLMs) that are pre-trained using data from Indian languages. Between 2023-24, indigenous companies such as Ola and Sarvam AI, have released pre-trained models — *Krutrim* and *OpenHathi* respectively — built using large-scale datasets in Indian languages (Jindal, 2023).

We highlight some of the key players in the Indian AI ecosystem below.

- **AI Companies:** Companies such as Tata Consultancy Services, Infosys, Wipro, HCL Technologies, Ola, Tech Mahindra, Reliance Industries, Karya, and Sarvam AI are at the forefront of AI development and adoption in India.
- **Government Institutions:** Government institutes such as the NITI Aayog, the Ministry for Electronics and Information Technology (MeitY), and the Digital India Initiative play a key role in the promotion and regulation of AI in India.
- **Research centres:** Institutes such as the Wadhvani Institute, Microsoft Research India, Centre for Development of Advanced Computing, and Indian Institute of Science (IISc), Bengaluru are some of the research institutes working to advance AI research in India. Additionally, public-funded institutions such as the Indian Institutes of Technology (IITs) have dedicated AI research centres.
- **Other Startups:** Numerous organisations including Yellow.ai, Haptik, SigTuple, Fractal.ai, Digital Green, KissanAI, Myna Mahila Foundation, and Armman are involved in building AI-based solutions for specific sectors.

Responsible AI in India: A Brief Overview

The Indian government has recognised the importance of responsible AI (RAI) and has taken steps to promote ethical AI practices through frameworks like the National Strategy for Artificial Intelligence (NSAI). However, beyond NSAI, the government has been reticent to institutionalise any binding AI-specific regulation that directly governs the development and use of AI technologies, fearing that such measures could potentially stifle innovation within the industry (Liu, 2023).

Even within the existing RAI discourse and practice in India, the focus has mostly revolved around data privacy and security. The recently enacted Digital Personal Data Protection Act (DPDPA), 2023, which outlines rules for digital personal data processing, is one of the very few regulations that apply to the development and use of AI in India. While the DPDPA addresses data privacy concerns, it does not specifically regulate AI development or deployment. Concerns related to algorithmic transparency and accountability or issues arising from AI-induced bias and discrimination largely remain outside the purview of any enforceable legal framework in India. Notably, the use of AI for socio-economic development is also another vital aspect of RAI in India. The country is leveraging AI to address pressing challenges in healthcare, education, agriculture, and governance (Aneja, Gupta, et al., 2024). This is closely linked with its efforts to introduce indigenous AI models that are much more tailored to the Indian context. Indian AI developers and users have also expressed concern over the potential for cultural stereotypes and bias against India to be built into AI systems developed abroad (Dhillon, 2024). This has strengthened calls for AI for India to be built in India, i.e., “sovereign” AI (Singh & Rekhi, 2024; Letzing, 2024). However, such calls also raise concerns about “AI nationalism” (Aaronson, 2024) and hegemonic understandings of “India” being built into systems deployed for Indians (Atari et al., 2023).

Further, at the beginning of 2024, several warnings were issued about the misuse of AI for election misinformation and its potential to affect political outcomes in India (Sebastian, 2024) and globally (Aggarwal & Frayer, 2024). However, emerging research has highlighted that AI-based political misinformation did not significantly impact elections (Kapoor & Narayanan, 2024). Nonetheless, experts agree that AI, specifically Generative AI, has the potential to affect our information ecosystem significantly (Mitra et al., 2024). Subject matter experts have increasingly highlighted the misuse of generative AI to perpetuate technology-facilitated gender-based violence against women, girls, and gender-diverse individuals. This emerged as a significant concern when an Indian actress was targeted with a generative AI video that morphed her face onto another person’s body (Business Today, 2023). While existing provisions of the Information Technology Act, 2000 and the Indian Penal Code, 1862 were utilised to charge the culprit, incidents such as this raise the question of whether India needs specific legislation to regulate AI and its harms.

2) Responsible AI in Action

Part A: Government Policy and Initiatives

Policy & Regulation

India has a National Strategy for Artificial Intelligence, which was introduced in 2018 (NITI Aayog, 2018). The strategy aims to provide ‘*AI for All*’ by enabling an ecosystem where “the full potential of AI is realised in pursuance of the country’s unique needs and aspirations”

(Ministry of Electronics and Information Technology, Government of India, 2022a). NSAI acknowledges that the full potential of AI to drive inclusive economic and social development has not been achieved due to a focus on commercial adoption. It thus emphasises the need for a national strategy that can balance commercial benefit with the greater good. As a result, the strategy identifies key sectors that will benefit the most from AI. These include healthcare, agriculture, education, smart cities and infrastructure, and smart mobility and transportation.

India's NSAI includes the implementation of RAI principles as one of its key recommendations. It emphasises that any AI strategy should recognise that these technologies have the potential to undermine ethical conduct, violate privacy, and weaken security protocols. It stresses the mitigation of these risks as an essential part of AI governance frameworks. To deal with issues such as bias, NSAI espouses the principle of fairness and recommends adopting a use-case approach to “identify in-built biases and assess their impact, and in turn, find ways to reduce the bias”.

It also acknowledges the black box phenomenon that AI systems suffer from and suggests increasing the transparency of these systems by aiming for the explainability of their decision-making processes. To tackle privacy concerns, NSAI recommends the establishment of data protection frameworks (instituted through DPDPA in 2023) and sectoral regulations. To bolster security, it recommends moving towards an accountability framework which would facilitate easy identification of failures in the system in order to prevent them from recurring in the future. NSAI suggests that AI developers self-regulate by conducting damage impact assessments at each stage of the AI development process. It further suggests limiting liability for AI developers through safe harbour provisions, given they take every measure to ensure that harm does not occur through their systems. Here, the strategy also suggests creating a framework which will place proportional liability on all actors involved in the development process and implementing a regulatory regime that attaches liability only when actual harm occurs, preventing speculative damage claims.

In terms of binding legal frameworks, India does not currently have any regulation that directly and/or exclusively focuses on the development and deployment of AI. The government has adopted a cautious approach to introducing regulations, believing that such measures could stifle innovation and risk discouraging private actors from developing and testing AI models and applications. MeitY has further justified its non-interventionist stance by emphasising that existing legislation can be utilised to address harms and assign liability arising from AI-related risks (Grover, 2025). These primarily include:

- The DPDPA, 2023, which outlines regulations for personal data processing by AI platforms,
- The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, which lays down content moderation and platform regulation guidelines for intermediaries, which also includes AI-powered systems, and
- The Information Technology (IT) Act, 2000 and the *Bharatiya Nyaya Sanhita*, 2023 indirectly cover AI harms such as cybersecurity threats and online violence.

In addition to the above, a few state governments have also developed more localised governance policies for AI development and use within their jurisdiction. For example, the Tamil Nadu state government released its ‘Safe & Ethical AI Policy’ in 2020. It had 3 key objectives, including promoting ethical use of AI, attracting investments in AI development,

and lastly, providing access to open data, models, and computational resources (Government of Tamil Nadu, 2020).

It is worth noting that while no codified governance framework related to AI has been introduced at the central level in India, there have been numerous policy deliberations surrounding the topic within the government. For instance, in March 2023, MeitY proposed a new legislation called the ‘Digital India Bill’, designed to replace the existing IT Act to ensure that the regulatory apparatus was adequately equipped and updated to deal with recent technological developments (Yachu, 2023). The legislation was intended to act as an all-encompassing rulebook for the governance of the entire technology ecosystem in India. One of the key provisions of the proposed act was related to ensuring online safety through regulation of AI-based ad targeting and content moderation (MeitY, 2023b). However, despite being officially proposed, its enactment has been postponed, as MeitY has shown a growing preference for introducing smaller amendments to existing laws rather than implementing an entirely new one (Sur, 2024). Most recently, in January 2025, MeitY released an advisory committee report outlining recommendations for AI governance frameworks in India. This report, anchored in a techno-legal approach towards AI governance, recommends the creation of an “AI incident database” and assessment of the suitability of technological measures to mitigate AI risks. The report has been released for public consultation and is likely to inform future legislative and regulatory actions. We provide a detailed overview in the next section under ‘government actions’ (MeitY, 2025).

Government Actions

The Government of India has undertaken certain initiatives to encourage RAI approaches within the ecosystem, frequently focusing on multistakeholder collaboration for establishing principles and sectoral frameworks. Some of these key initiatives are outlined below in reverse chronological order.

- In January 2025, MeitY released a report titled, ‘AI Governance Guidelines Development’ for public consultation. The report has been developed by an advisory committee which was constituted in November 2023 and was assigned the task of identifying key AI governance issues in India, undertaking a gap analysis of existing frameworks, and putting forward recommendations to foster trustworthiness and accountability in AI systems within the Indian context (MeitY, 2025). Some of the key recommendations of the committee are as follows:
 - MeitY and India’s Principal Scientific Adviser¹ should establish an empowered mechanism — in the form of an “Inter-Ministerial AI Coordination Committee or Governance Group” — to implement a whole-of-government approach towards AI governance.
 - MeitY should establish, and administratively house, a *Technical Secretariat* to serve as a technical advisory body and coordination focal point for the Inter-Ministerial AI Coordination Committee.

¹ The office of the Principal Scientific Adviser plays an advisory role within the domain of science and technology and provides consultation support to the executive wing of the Indian government.

- To “build evidence on actual risks and to inform harm mitigation”, the Technical Secretariat should develop and administer an “AI incident database” as a repository of AI harms experienced in the real world.
- To strengthen transparency and governance, the Technical Secretariat should engage AI developers to encourage voluntary commitments on transparency across the overall AI ecosystem and on baseline commitments for “high capability/widely deployed systems”.
- The Technical Secretariat should assess the “suitability of technological measures” to mitigate AI-related risks.
- Additional efforts need to be directed towards ensuring that proposed legislations such as the Digital India Act are suitably equipped to minimise “risks of harm from malicious use of emerging technologies, including AI”. There is a need to strengthen and harmonise India’s legal framework, regulatory, and technical capacity to ensure effective grievance redressal and ease of doing business within the digital realm.
- In November 2024, MeitY partnered with the United Nations Educational, Scientific and Cultural Organization (UNESCO) to organise a stakeholder consultation on AI ethics and safety. This event served as the inaugural session in a series of five consultations under the AI Readiness Assessment Methodology (RAM), a joint initiative by the two bodies. It seeks to develop an India-specific AI policy report aimed at highlighting the country’s strengths and opportunities for growth within the AI ecosystem. The report also intends to offer practical recommendations for the responsible and ethical adoption of AI across diverse sectors (Press Information Bureau, 2024).
- In October 2024, the IndiaAI Mission funded eight projects focused on the development of customised, indigenous tools and frameworks that can foster ethical, transparent, and trustworthy AI in India (Parasnis, 2024). The themes for the projects include Machine Unlearning, Synthetic Data Generation, AI Bias Mitigation, Ethical AI Frameworks, Privacy-Enhancing Tools, Explainable AI, AI Governance Testing, and Algorithm Auditing Tools.
 - Additionally, in October 2024, MeitY officials mentioned that they were considering the possibility of setting up an ‘AI Safety Institute’ (AISI) that could “help set standards, frameworks and guidelines for AI development without acting as a regulatory body or stifling innovation” (Agrawal, 2024).
- In August 2024, the Bureau of Indian Standards — a statutory organisation under India’s Ministry of Consumer Affairs, Food and Public Distribution — reported that it was preparing a set of standards for AI-based applications in India. The standards framework will be designed to provide guidance to AI developers across the value chain of AI applications, including model development, implementation, and post-deployment (Kumar, 2024).
- On 1 March 2024, MeitY issued an AI advisory — under the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 — advising all intermediaries and platforms to obtain ‘explicit permission’ from the Indian government before using and/or making available any under-tested/unreliable AI models, software, or

algorithms to Indian users. However, the advisory attracted sharp criticism and was withdrawn and replaced with a new one, with the requirement of obtaining government permission revoked (MeitY, 2024; Mohanty & Sahu, 2024).

- In November 2023, the Indian government signed the Bletchley Declaration that emphasises global cooperation on matters related to AI governance (UK Government, 2023).
- In October 2023, the IndiaAI Expert Group published the first edition of its report detailing India's AI strategy. The report contains recommendations from seven expert groups set up by the Indian government to examine different issues related to AI development and adoption in India (MeitY, 2023a).
- In July 2023, the Telecom Regulatory Authority of India released a series of RAI recommendations for deploying AI in telecommunications, addressing concerns related to algorithmic bias, transparency, and data privacy (Telecom Regulatory Authority of India, 2023). The recommendations highlighted the urgent need to adopt a regulatory framework applicable across sectors to ensure the development of responsible AI in India. They also suggested that the framework must facilitate the setting up of an independent statutory authority and a Multi Stakeholder Body (MSB) which will act as an advisory body to the proposed statutory authority. Additionally, it recommended the creation of different risk categories that can accordingly guide the differential application of RAI principles across various use cases. TRAI also emphasised the need for the government to coordinate with international agencies and other countries to form a global agency that can act as the nodal international body for the development, standardisation, and responsible use of AI.
- In March 2023, the Indian Council of Medical Research (ICMR), a government agency for biomedical research and development, released ethical guidelines for the development, deployment, and adoption of AI-based solutions for biomedical research and healthcare delivery in India (IndiaAI, 2023). The guidelines apply to all AI-based tools developed for use in biomedical and health research and applications involving human participants and/or their biological data. The guidelines identify ten ethical principles for AI in the healthcare sector which include protecting patient autonomy; ensuring safety and minimising risk; trustworthiness; data privacy; accountability and liability; optimisation of data quality; accessibility, equity, and inclusiveness; collaboration; non-discrimination and fairness; and validity.
- In July 2022, the draft National Data Governance Framework Policy, which focused on safety in non-personal data processing and collection, was put forward for public consultation (Ministry of Electronics and Information Technology, Government of India, 2022c).

Part B: Industry and Non-government Initiatives

The private sector and non-governmental organisations have played an important role in advancing RAI practices in India. Through self-regulation, collaboration, and resource development, these entities have taken significant strides toward embedding ethical considerations into AI systems. We highlight a few examples in this section.

Industry-led initiatives

The National Association of Software and Service Companies (NASSCOM), a tech-industry-focused non-government association has emerged as a key industry body driving the operationalisation of RAI in India. In 2024, it published a comprehensive guide, ‘Developer’s Playbook for Responsible AI’ to assist developers in identifying potential risks, aligning with best practices, and adopting responsible AI methodologies (NASSCOM, 2024). As part of its Responsible AI Resource Kit, NASSCOM has also created a tool to help organisations assess and mitigate risks associated with AI systems (NASSCOM, n.d.).

Corporate Self-Regulation and Commitments

In its pest management AI solution, ‘CottonAce’, Wadhvani AI has integrated a “human-in-the-loop” mechanism to prevent farmers from being disadvantaged due to inaccurate results. This approach underscores the importance of human oversight in AI decision-making processes (Cheney, 2020).

Multistakeholder Collaborations

Coalition for Responsible Evolution of AI (CoRE-AI) is India’s first major multi-stakeholder coalition dedicated to the responsible development and deployment of AI technologies. With its secretariat at The Dialogue, a tech policy think tank, CoRE-AI brings together diverse actors, including Google, Microsoft, AWS, Infosys, Ashoka University, IIM Bangalore, and startups like Corover.ai, to foster ethical AI practices and ensure that AI benefits society at large (Krishan, 2024).

Civil Society Organisations - Research & Policy

- Digital Futures Lab: This organisation has worked extensively on the issue of Responsible AI. In 2024, they published a first-of-its-kind study on gender bias in the Indian LLM ecosystem (Aneja, Gupta, et al., 2024). Additionally, they have also launched the Responsible AI Lab in 2024, which hosted a first-of-its-kind fellowship on responsible AI for organisations in the social sector developing AI-based interventions (Aneja, John, et al., 2024).
- Carnegie India: This organisation has a Technology and Society program which focuses on five key cross-cutting areas including data, strategic technologies such as semiconductors, emerging technologies such as AI, strategic partnerships, and digital public infrastructure.
- Aapti Institute: This organisation has published a report on “Just and equitable data labelling towards a responsible AI supply chain” (Natarajan et al., 2021). The report aims to highlight the value of annotated datasets in the AI supply chain, the maturity of India’s data annotation market, and the employment opportunities within it.
- The Dialogue: This organisation has published multiple reports on Responsible AI including a report on “Principles for Enabling Responsible AI Innovations in India: An Ecosystem Approach” (Shekar et al., 2023). It also houses the Coalition for Responsible Evolution of AI (CoRE-AI).

Civil Society Organisations - Community Engagement

The Internet Freedom Foundation’s Project Panoptic is a community resource which aims to provide information about the ongoing development and deployment of facial recognition systems throughout India (Internet Freedom Foundation, 2020).

Civil Society Organisations - Advocacy

The Internet Freedom Foundation has provided legal support to the petitioner in an ongoing challenge in the High Court of Telangana against the use of facial recognition systems by the police in Telangana (Bapat, 2022).

Academia

The Centre for Responsible AI, Madras (CeRAI) is a premier research institute hosted at the Indian Institute of Technology, Madras in southern India and specialises in applied research on responsible AI. It has emerged as one of the key research institutes focused on RAI theory and practices tailored to AI systems in India (Centre for Responsible AI, 2024).

3) Key Trends & Outlook of Responsible AI in India

AI development and adoption have gained significant momentum in India, driven by both government and private sector initiatives. There is a broad sense of optimism within the country with regard to AI's transformative potential to revamp service delivery and product offerings across different sectors, including healthcare, agriculture, and education. Over the past year, policy deliberations around 'Responsible AI' have steadily intensified, although most of these are yet to be codified under any enforceable legislation. There is a growing preference within the government to approach AI regulation with pragmatism and caution while prioritising the collection of diverse perspectives and insights to develop a regulatory framework that aligns with India's unique socio-economic context. This can be seen in the numerous consultations undertaken by MeitY and IndiaAI, with the latest focused on developing a governance framework.

However, such a measured approach also comes at a cost. The prevailing uncertainty surrounding the 'Digital India Bill,' coupled with the abrupt issuance and subsequent withdrawal of the March 2024 advisory, has generated significant ambiguity within the AI ecosystem. Even the delays in the notification of rules under the 2023 data protection legislation — which were only recently announced in January 2025 — have fueled considerable and unnecessary speculation leading up to their announcement. Additionally, the sheer multitude of actors currently involved in standard- and framework-setting within the Indian context has further contributed to fragmentation and inconsistency, complicating efforts to establish a cohesive regulatory environment.

However, in parallel, the number of AI applications in India continues to grow, bolstered by state-led investment and private capital. Data privacy — the primary focus of one of the few legislations directly applicable to AI — represents only one aspect of a broader spectrum of risks that many of these applications present to the Indian population. Issues related to algorithmic accountability, exclusion, misuse, biases, and discrimination remain largely unaddressed, particularly as AI becomes increasingly integrated into high-stakes sectors such as healthcare and agriculture.

State and non-state actors around the world are racing to embrace the promise of AI and concerns have been voiced that excessive or overbroad regulation may stifle growth and innovation opportunities. While the intention of the Government of India can be clearly ascertained from the steps they have taken to create suitable regulations for the booming AI sector in India, their fragmented and hesitant efforts to put in place enforceable legislation

reveal the apprehension that strict regulation may hamper AI growth in India. However, recent actions such as the AI Advisory Group report showcase that the Government is trying to adopt a more streamlined approach to AI governance and keep democratic values at the forefront even as we aim to reap the benefits of accelerated AI growth. The report, which has been made available for public consultation, contains actionable recommendations to enable sustainable and ethical AI development in India and suggests adopting a “whole of government” approach. India’s measured approach to AI regulation also allows it to learn and benefit from the regulations put in place across the world and create AI regulation that is uniquely suited to its needs. With various opposing and equally important interests vying for attention, it seems India still has a long way to go before ‘Responsible AI’ principles are codified into enforceable legislation.

Working Draft

Responsible AI in Sri Lanka

LIRNEasia, Sri Lanka

1) Introduction

Country Overview

At the time of writing, Sri Lanka, a lower-middle-income country (Metreau et al., 2024), is gradually recovering from a debilitating economic crisis that significantly increased poverty and food insecurity. Limited digital integration in the economy and modest computer (36%) and digital literacy (60%) rates (Department of Census and Statistics, 2022) hinder AI progress.

Sri Lanka sees a modest performance of 6.2 on the Democracy Index in 2023 (on a scale of 1-10) (Our World in Data, 2023b), with a stronger score of 0.74 on the Freedom of Expression Index in 2023 (on a scale of 0-1) (Our World in Data, 2023c) (the index “captures the extent to which people can voice their views and the media can present different political perspectives”). Finally, on the Civil Society Participation Index, which indicates the “extent to which citizens are active in diverse organizations which choose and influence policymakers...[ranging] from 0 to 1 (most active)” (Our World in Data, 2023a), Sri Lanka sees a moderate performance of 0.67.

“Sri Lanka has also shown to have developed several globally competitive technology companies backed by a home-grown technically skilled workforce. The country’s digital sector is said to be “well-established and mature” (Ministry of Technology, 2024, p. 40), with reasonably well-developed digital infrastructure in the last 20 years. These baseline conditions create a unique opportunity for Sri Lanka as it seeks to become an AI-enabled nation. However, there are still challenges, and room to “improve and expand access, innovation, and capabilities of people, businesses, and the government” (Ministry of Technology, 2024, p. 2). There is a pressing need to improve connectivity island-wide and digital capacities among individuals, increase incorporation of digital among business and government, developing trust in digital systems, creating a conducive environment for entrepreneurship and a financial system with strong digital integration (Ministry of Technology, 2024, p. 2). The prior digital transformation efforts are also incomplete. There is a scarcity of advanced AI engineering skills and a lack of executive and strategic expertise to develop strong AI business cases” (Committee on Formulating a Strategy for Artificial Intelligence (CFSAI), 2024).

The State of AI

Despite these challenges, a few pilots projects in government and many applications of AI in the private sector can be identified, even though there is a dearth of strong AI business use cases and home grown AI innovation. In the 2023, Government AI Readiness Index by Oxford Insights, Sri Lanka ranked 95th out of 193 countries (Hankins et al., 2023). While official data on AI adoption is unavailable, the recently released draft National AI Strategy—developed by experts from business, government, and civil society—recognizes the country's early stages of AI adoption in both the private and public sectors.

AI Innovation and Adoption

Public Sector

Although Sri Lanka lacks large-scale AI implementations in government, there are a few notable pilot initiatives. The 1990 Suwa Seriya, the country's free island-wide emergency ambulance service, is piloting a computer vision-based AI system that allows Emergency Medical Technicians (EMTs) to remotely connect with physicians at the Emergency Command and Control Centre (ECCC), monitor vital signs, and provide specialized care before reaching the hospital—a first for emergency medical response in the region (Daily FT, 2023). Meanwhile, in the energy sector, the Ministry of Power and Energy, with support from the Asian Development Bank, is developing an AI-powered framework for microgrids to integrate distributed renewable energy systems seamlessly (Daily FT, 2023). The initiative also includes a pilot project to demonstrate an AI-enabled microgrid's benefits to consumers and the power system.

Private Sector

Although Sri Lanka is not widely recognized for home-grown AI innovation, AI adoption is steadily growing across industries. In banking, Sampath Bank introduced Sri Lanka's first AI-powered banking robot and virtual teller for tasks like account opening and fund transfers (Thavapalakumar, 2024), while Commercial Bank uses an AI assistant (AIPARA) to streamline data searches. In healthcare, Nawaloka Hospitals launched South Asia's first AI-powered Radiology Centre in 2024, leveraging AI for medical imaging, appointment scheduling, and efficient access to patient data (Daily Mirror, 2024). In renewable energy, Alta Vision is developing AI-based mini-grids to optimize and automate solar and wind energy generation, improving electricity access in rural areas (Alta Vision, 2024).

Similarly, while not widely reported, several other sectors in Sri Lanka are likely adopters AI. PickMe, a local ride-hailing platform, likely employs AI for route optimization, driver-rider matching, demand prediction, and dynamic pricing. Possible AI deployments are also evident among mobile network operators, particularly in customer retention, personalized recommendations, network optimization, and fault prevention—all prime candidates for AI, given current technological capabilities and the data collected by those companies. Others use cases of AI have been reported in insurance, agriculture, and the apparel industry. However, some of these claims should be approached with caution, as AI-washing and exaggerated reporting are common.

Responsible AI

Sri Lanka on GIRAI

On the Global Index of Responsible AI (GIRAI), Sri Lanka ranks 72nd out of 138 countries. The graph below shows Sri Lanka's overall score in Responsible AI within Asia: below the 50th percentile, with a score of 10.9, which is 8.9 points lower than the global average of 19.8 (refer to Section 2 of this report for the scoring methodology).

Two key points deserve attention. First, Sri Lanka's score is heavily driven by a small number of initiatives and frameworks that, due to their broad applicability across multiple thematic areas, contribute disproportionately to the overall score. Second, the draft National AI Strategy was released after the evidence cutoff period for this index. As a result, Sri Lanka's overall GIRAI score does not account for this development and would likely be higher in a future update.

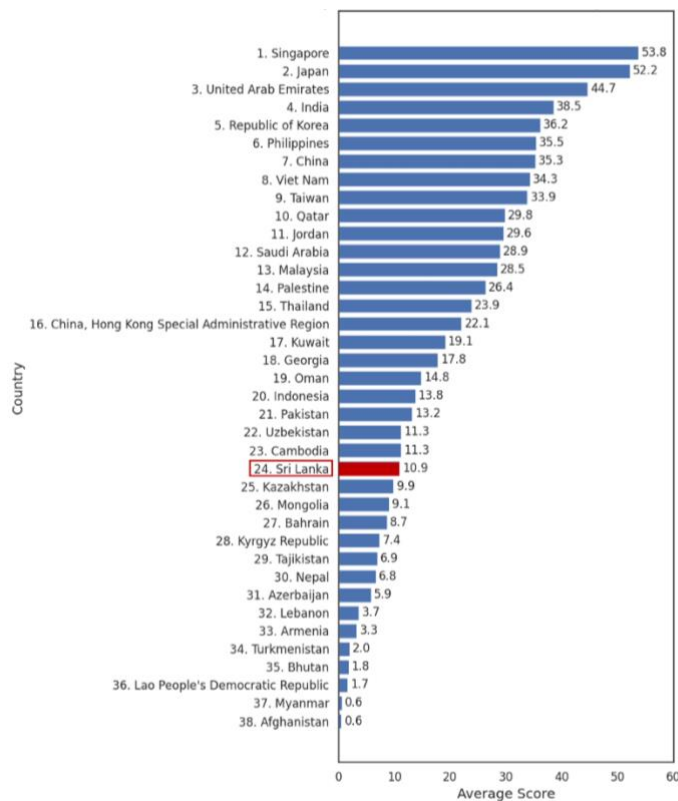


Figure 27: GIRAI country scores in Asia

PILLAR SCORE			DIMENSION SCORE		
Government frameworks	Government actions	Non-state actors	Human rights and AI	Responsible AI capacities	Responsible AI governance
11.56	10.13	11.23	9.95	12.83	11.04

Figure 28: Sri Lanka RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 23)

As shown in Figure 33 above, no specific actor—state or non-state—currently dominates Responsible AI (RAI) efforts in Sri Lanka, which remain in their infancy. However, with effective implementation of the National AI strategy, there is potential for improvement across all pillars and dimensions.

2) Government Policy and Initiatives

National AI Strategy

Sri Lanka’s draft National AI Strategy, published after the evidence cutoff for this index, marks a significant step in advancing Responsible AI (Weerasinghe, 2024). It was developed by a Committee on Formulating the Strategy for Artificial Intelligence (CFSAI) that included volunteers from government, academia, the private sector, and civil society, and has remained resilient against changes to Sri Lanka’s political climate. While meaningful progress depends on implementing the strategy’s initiatives, its approach to Responsible AI is worth highlighting.

The strategy has a dual focus on rapid value creation and sustained capability development and is underpinned by 7 core principles (Committee on Formulating a Strategy for Artificial Intelligence (CFSAI), 2024, p. 5). Five of those seven principles, namely, Inclusive and responsible AI for all, Trustworthy and transparent AI, Human-centric, Sustainable and Future Ready AI, Agile and Adaptive Governance, directly address principles of responsible AI. Further, the strategy’s three key pillars include enablers, accelerators and the creation of a safe and trustworthy AI ecosystem in Sri Lanka. The latter includes Sri Lanka’s approach to implementing Responsible AI (RAI principles).

The pillar on Creating a Safe & Trustworthy AI Ecosystem identifies ethical and governance considerations for Sri Lanka (p. 45). It distinguishes between ethical and governance considerations related to the development of AI technologies and products (such as transparency and explainability, fairness and equity & human centricity) and broader economic, societal, and ethical considerations that emerge from the deployment and use of AI. These include, among others, intellectual property rights, competition-related concerns and job market effects

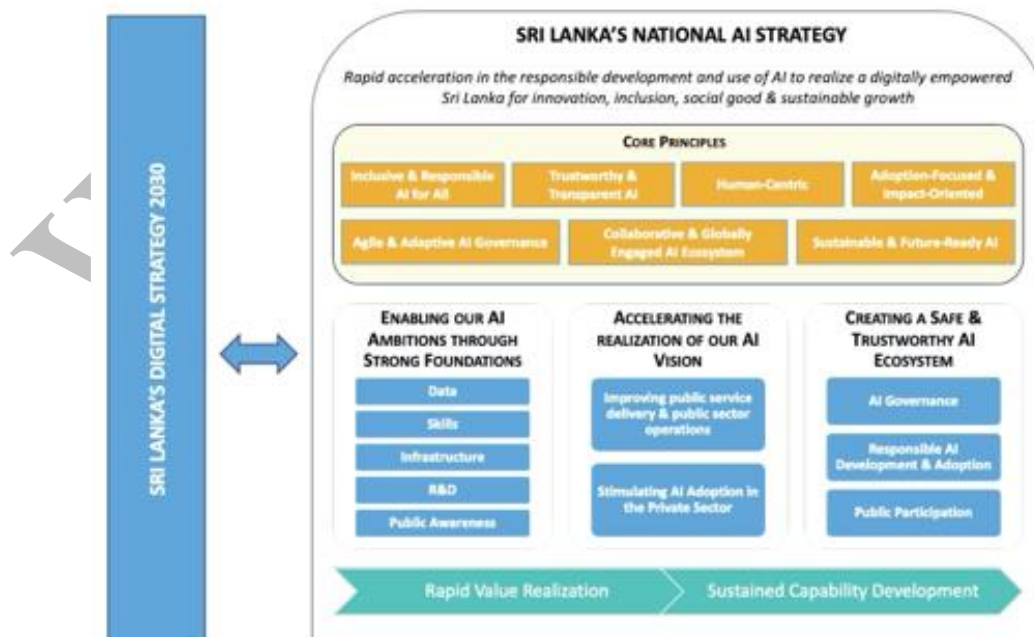


Figure 29: Strategic Framework for Advancing AI in Sri Lanka (Committee on Formulating a Strategy for Artificial Intelligence (CFSAI), 2024, p. 5)

The strategy outlines a three-pronged approach to implementing Responsible AI principles (Committee on Formulating a Strategy for Artificial Intelligence (CFSAI), 2024, p. 5). First, it emphasizes adaptive and agile governance, which is guided by measures like establishing a Responsible AI Advisory Council and developing a Responsible AI framework. Second, it promotes fostering Responsible AI development and adoption across the ecosystem through initiatives like resource toolkits and innovation sandboxes. Lastly, it highlights the importance of public engagement through awareness campaigns, inclusive consultations, and mechanisms for accountability and scrutiny.

The proposed National AI Center (NAIC) (p. 7), under the Ministry of Digital Economy, will serve as the primary body for implementing the strategy, aligning with the government's broader digitization goals. The strategy's execution is a key development to watch in Sri Lanka's short-to-medium-term progress on Responsible AI.

Government Frameworks

There currently exists no comprehensive AI regulation in Sri Lanka. However, The Personal Data Protection Act (PDPA), No. 9 of 2022 (The Parliament of Sri Lanka, 2022), the principal legislation that deals with personal data protection in Sri Lanka, has direct implications for AI development. Specifically, the provision in question provides some recourse to data subjects who object to *entirely* AI-driven decision-making (Parliament of Sri Lanka, 2022, p. 11):

Automated individual decision making

18. (1) Subject to section 19, every data subject shall have the right to request a controller to review a decision of such controller based solely on automated processing, which has created or which is likely to create an irreversible and continuous impact on the rights and freedoms of the data subject under any written law.

This places restrictions on automated decision-making, though these are less stringent than those in the EU's General Data Protection Regulation (GDPR). Notably, the PDPA does not prohibit certain automated decisions outright. While it also does not grant individuals the right to direct human intervention by the data controller as a redress mechanism, it does allow them to request a review of such decisions and seek the erasure of their data if its use infringes on their rights. Additionally, Section 18(2) requires that, in cases of automated decision-making, the data controller must follow measures and criteria to be specified by the Data Protection Authority to safeguard the rights and freedoms of data subjects.

Government Actions

Government actions addressing Responsible AI (RAI) in Sri Lanka remain limited, with most initiatives focusing on education and capacity building. Notable efforts include the launch of "AI Clubs in Schools" in July 2024 to enhance AI literacy among students, fostering innovation and preparing them for future careers in AI and technology. In 2022, the Central Bank of Sri Lanka's Centre for Banking Studies offered a course that introduced AI and data science applications that aimed at public sector skills development (Centre for Banking Studies, 2022, p. 88).

A more direct government action was the January 2025 collaboration between the Ministry of Digital Economy, LIRNEasia, and the United Nations Development Program to host a public consultation on the draft National AI Strategy (Ministry of Digital Economy, 2025). This

initiative highlighted the government's effort to gather public input on shaping AI policy, including Responsible AI, and involved representatives from the public and private sectors, academia, civil society, and development partners.

Sri Lanka's Data Protection Authority was "established under the PDPA in August 2023 to regulate personal data processing and safeguard individuals' privacy" (Ministry of Digital Economy, 2025). While the Data Protection Authority has yet to release specific AI-based directives, the recently issued Draft Directive on Specification of Instruments for Processing Personal Data Outside Sri Lanka (Data Protection Authority of Sri Lanka, 2024) and Draft Directive on Classification of Categories of Personal Data for Public Authorities (Data Protection Authority of Sri Lanka, 2024) indirectly implicate AI in their scope.

3) Non-Government Initiatives

Academia

Compared to their global and regional counterparts, Sri Lankan academia's engagement with Responsible AI (RAI) remains limited. While higher education institutions in Sri Lanka provide AI-related courses, barring a few exceptions, these are largely focused on AI innovation and technology. Nonetheless, these offerings are significant, as the National Strategy's strong emphasis on RAI may eventually influence course content and research priorities. Building on this existing technical foundation will be crucial for integrating RAI-focused research and teaching initiatives within Sri Lankan academia.

University programs in Sri Lanka that include course material related to Responsible AI include the Bachelor of Science Honors in Artificial Intelligence (University of Moratuwa, 2024a) and the Master of Science/Postgraduate Diploma in Artificial Intelligence at the University of Moratuwa (University of Moratuwa, 2024b). Additionally the Sri Lanka Institute of Information Technology (SLIIT) offers a Bachelor of Science (Hons) in Information Technology specializing in Artificial Intelligence (SLIIT, 2023). While these not explicitly referencing Responsible AI in their course outlines, other programs, such as the University of Peradeniya's "Artificial Intelligence and Signal Processing," (University of Peradeniya, 2024) aim to promote AI-related skills and knowledge, providing a foundation for future integration of Responsible AI concepts.

Going forward, academia will need to play a crucial role in advancing Responsible AI (RAI) in Sri Lanka. Key contributions would include driving RAI research, fostering multidisciplinary partnerships to develop socially beneficial AI solutions, engaging in evidence-based RAI policy advocacy, and building public awareness and capacity—all areas that are currently underdeveloped.

Private Sector & Industry

While AI adoption is progressing in the private sector (as discussed in the first section), initiatives focused on Responsible AI (RAI) remain limited. Technology and AI experts from private sector organizations and industry bodies contributed to the Draft National AI Strategy through the Committee on Formulating a Strategy for Artificial Intelligence (CFSAI) (Committee on Formulating a Strategy for Artificial Intelligence (CFSAI), 2024, p. 62). However, these are contributions by individuals and are not indicative of any broader efforts within private sector and industry to advance RAI.

Civil Society

Civil society efforts to advance responsible artificial intelligence (AI) in Sri Lanka also remain in their early stages. However, LIRNEasia, a Colombo-based regional think tank, is a notable exception. With a long history of tackling digital policy and regulatory issues in Sri Lanka, LIRNEasia has actively engaged in discussions on the ethics of new technologies and has directly advocated for a National AI Policy (Bandaranayake & Dias, 2022). As already mentioned under government actions above, in January 2025, LIRNEasia, in collaboration with the Ministry of Digital Economy and the United Nations Development Program, hosted a public consultation on the draft National AI Strategy (Ministry of Digital Economy, 2025). As Sri Lanka advances its digitization and AI adoption, active civil society engagement will be essential to safeguard citizen rights. Prioritizing this focus will be critical to ensuring an inclusive and responsible AI future.

Working Draft

Responsible AI in New Zealand

LIRNEasia, Sri Lanka

1) Introduction

Country Overview

New Zealand is a high-income country home to a population of a little over 5 million (The World Factbook, 2025), with a very strong parliamentary democracy (Global State of Democracy Initiative, 2024) (a constitutional monarchy as part of the Commonwealth) and economy.

New Zealand boasts a strong performance in the realms of human rights and democracy. The country scores a 9.6 on the Democracy Index in 2023 (on a scale of 1-10) (Our World in Data, 2023b), with a 0.95 on the Freedom of Expression Index in 2023 (on a scale of 0-1) (Our World in Data, 2023c) indicating an enabling environment in which “people can voice their views and the media can present different political perspectives” (Our World in Data, 2023c). Civil society in New Zealand enjoys a considerable degree of agency, with a score of 0.87 in 2023 in the Civil Society Participation Index (Our World in Data, 2023a) indicating New Zealand’s civil society operates [actively] “in diverse organizations which choose and influence policy-makers”.

Agriculture, manufacturing and tourism play a strong role in driving the economy, and a majority of the population is urbanized. The country is also home to a set of strong universities that invest on research alone around \$1.4 billion (Universities New Zealand - Te Pūkai Tara, 2025). Digital infrastructure in New Zealand is considered “advanced” (International Trade Administration, 2024) in terms of network coverage and internet speeds; it’s advanced computing sector, small. The digital technology sector as a whole is seen as a “a driver of future sustainable economic growth and jobs” (Ministry for the Digital Economy & Communications, 2023, p. 4), with the sector’s sizeable contribution to the economy – \$7 billion in 2021 – likely to support increasing institutional, organizational, and citizen AI-based capacities into the future.

AI Innovation and Adoption

While there is an absence of data to make definitive claims about AI adoption in New Zealand, AI systems and tools are certainly used in the country. AI is playing a pivotal role in New Zealand’s public service across multiple departments, with the aim being increases in productivity, innovation, and better service offerings for citizens (Sharon, 2024). Some examples of AI deployment include: airports employing AI-based facial recognition to validate passports and passengers (New Zealand Customs Service, 2024); grocery retailer Foodstuffs North Island trialing the use of facial recognition technology to match shoppers’ faces with a database of identified individuals with “repeated harmful behaviors” (Radio New Zealand, 2024) (the latter of which received some criticism on the grounds of possible privacy violations). AI-based facial recognition was also used in an unsanctioned instance by law enforcement (Fisher, 2023), although after much opposition this was discontinued (Pennington, 2024).

The government also actively promotes AI use among businesses, launching GovGPT in September 2024, a tool intended to “make it easier to access reliable government information

and support for businesses and people wishing to interact with government agencies” (Collins, 2024). A Datacom survey of 200 senior business leaders in New Zealand shows that adoption of AI within businesses is “relatively advanced, with almost half (48%) of business leaders stating AI is already currently in use within their organization” (Datacom, 2024), with the majority of business leaders claiming to be enthusiastic about AI and its use. Policies and plans in which to ground the “effective and safe” use of AI, however, are lax with 91% of “organizations having no targets around the use of AI”.

Academia does make some contribution towards AI innovation and adoption, with programs offered by the Universities of Auckland (2025b) and Waikato (University of Waikato, 2025), and Victoria university (Victoria University of Wellington, 2023b).

Responsible AI Overview

New Zealand received an overall score of 27.33, ahead of the *global* average 19.8 of (refer to Section 2 of this broader report for scoring methodology).

PILLAR SCORE			DIMENSION SCORE		
Government frameworks	Government actions	Non-state actors	Human rights and AI	Responsible AI capacities	Responsible AI governance
28.49	33.70	12.28	27.87	11.47	32.20

Figure 30: New Zealand RAI Pillar and Dimension Scores, GIRAI Index (Adams et al., 2024, p. 22)

On the three dimensions, Figure 26 above shows that Responsible AI governance sees the highest score, with Human rights and AI close behind, and a relatively underwhelming performance in Responsible AI capacities. While Figure 26 also indicates that government actions largely drive RAI in New Zealand, with considerably less contribution from non-state actors, Tom Barraclough, Tech Policy Researcher and Director at the Brainbox Institute, New Zealand, asserted in a key informant interview that “non-state action is comparably effective [to state action], and usually leads state action”. Tom notes that RAI among non-state actors in New Zealand is typically framed by “expressions of voluntary, non-binding commitment” by agencies and private actors, and that – particularly in the private sector – there are initiatives such as company policies, working groups, formulation of RAI principles, etc. that this score does not reflect. It would also be prudent to note a methodological limitation of the first edition of the GIRAI that may in part explain the scoring here – researchers were not directed to document an exhaustive list of state and non-state actions, which may see lesser-visibility non-state actions like assertions of commitments to RAI being uncaptured.

Overall, RAI in New Zealand is fairly limited across the board, which is worrying, given 1) substantial concern among New Zealanders that AI may be leveraged in a malicious manner; there exists no concrete regulation of AI; and that AI may incur unintended costs and harm (Matika, 2023); 2) an incognizance of the “legal and ethical implications of AI” (Ryburn et al., 2024); 3) concern that “AI development and adoption is outpacing regulation” (Knowles, 2023). There currently exists no single AI strategy that in whole or in part addresses RAI.

There is little concrete work among non-state actors, beyond commentary. Additional GIRAI research found nothing substantial in the realm of AI capacity and skills development in the public sector. All this, however, ought to be considered in the context of New Zealand's small population size and "comparatively small [economy] in the global marketplace" (Moran et al., 2025) – relative, that is, to other developed economy powerhouses at the forefront of RAI. Additionally, the fact that there at a minimum exists reasonably strong discourse on RAI signals what are likely to be more concrete developments on this front in the coming years.

2) Government Policy and Initiatives

While there isn't a single broad RAI strategy, there exist multiple outputs from various government entities that in various ways address RAI – ranging from suggestions of concrete action to brief mentions of AI. Research during the creation of the GIRAI found that the majority of frameworks that pertain to AI have a human oversight and determination leaning, along with safety, accuracy and reliability, and impact assessments. This section briefly discusses key entities and existing outputs as of December 2024.

AI "Agency"

New Zealand does not have an AI agency in the way, say, Singapore (AI Singapore, 2022) or the Philippines do (Department of Trade and Industry, 2024); however, there is a single entity that, for public sector affairs, is the entity tasked with leading the "safe and trustworthy adoption and use of AI in the public service" (New Zealand Government, 2024b) – the Government Chief Digital Officer (GCDO). The GCDO is tasked with: "leading AI use as the Digital system lead; setting standards; guiding government organisations in adopting and using AI; working with providers of AI technology". The GCDO has also done surveys among government entities to assess AI use in the public sector (New Zealand Government, 2023a) and to gather ideas on applications of AI in the public sector (New Zealand Government, 2024a). The GCDO also released guidelines on generative AI use in the public sector, which this report will discuss further on in this section.

AI Strategies and Plans

The first is a Cabinet paper titled "Approach to work on Artificial Intelligence" (Ministry of Business Innovation & Employment, 2024), agreed upon in July 2024 that lays out the ways in which governments will be expected to use AI. Broadly, the paper establishes Cabinets agreement that 1) "government agencies should be encouraged to adopt AI for its benefits, while managing the risks" (Ministry of Business Innovation & Employment, 2024); 2) the GCDO will lead "work to accelerate responsible use of AI across public services, to deliver better outcomes for all New Zealanders". The paper expresses "support for increased uptake of AI in New Zealand" (Ministry of Business Innovation & Employment, 2024, p. 1), and emphasizes the "need to stay connected to key international discussions that are establishing international norms for AI, including on AI safety". On regulation, the paper recognizes that "businesses are slow to adopt AI, due in part to uncertainty about the future regulatory environment", and that "further regulatory intervention should only be considered to unlock innovation or address acute risks and use existing mechanisms in preference to developing a standalone AI Act".

The Digital Strategy for Aotearoa, released in September 2022, maps out goals, recognized opportunities, and anticipated challenges vis-à-vis New Zealand's engagement with digital technologies and data (New Zealand Government, 2022). The document at the onset

encourages that New Zealanders consider what it would mean to embrace the “ethical deployment of Artificial Intelligence”(New Zealand Government, 2022), and emphasizes the importance of trust “when collecting data, analyzing it with artificial intelligence (AI) and other algorithms, and using it to make decisions”. Other notable mentions of AI include the need for an AI Strategy, and when stressing the need to consider trade-offs associated with digital technology use (for example, how AI use in health screening, may increase efficacy, but also collects and stores sensitive information about patients).

The Digital Technologies Industry Transformation Plan, released in May 2023 as a tool in support of the Digital Strategy for Aotearoa, seeks to build the country’s “national reputation around tech and innovation, improve the digital skills and talent pipeline, build Māori participation and activity in the sector, and support Software-as-a-Service and game development as high-growth, high potential subsectors” (Ministry for the Digital Economy & Communications, 2023, p. 4). RAI briefly features in a latter section in the document titled “Future Focus Areas” (Ministry for the Digital Economy & Communications, 2023, p. 31), with the following emphasized: the need for an AI Strategy; the importance of a “trustworthy and ethical data ecosystem” to underpin ethical AI adoption and innovation; a balance between objectives of ethical and innovative data use.

AI-Related Guides

The National AI Development Checklist (Health New Zealand, 2023) for Health New Zealand is, as the name suggests, a checklist of things to consider during the “development, validation, or implementation of a new AI” (p. 1). RAI-related questions the checklist asks people to consider are: “Describe the [AI algorithm/model’s] methods used, data and features. Has this been/will it be published or made available?”; “How would the use of the AI fit into the clinical workflow? Describe any clinical/operational input. Are there likely to be any concerns or barriers to use by clinicians/other staff? Can these be mitigated? Will training be needed? Is human oversight intended? Where will accountability lie?” (Health New Zealand, 2023, p. 2).

In September 2023, New Zealand’s Privacy Commissioner released the “Artificial Intelligence and the Information Privacy Principles” (Office of the Privacy Commissioner, 2023), built upon the Privacy Commissioner expectations around AI use (Office of the Privacy Commissioner, 2023) released previously. The document serves as a guide for all New Zealanders who use or intend to use AI-based tools in the country on how they may apply the Privacy Act’s 13 Information Privacy Principles in the context of AI. The following excerpts from the report provide a sense of the kind of advice offered across the document.

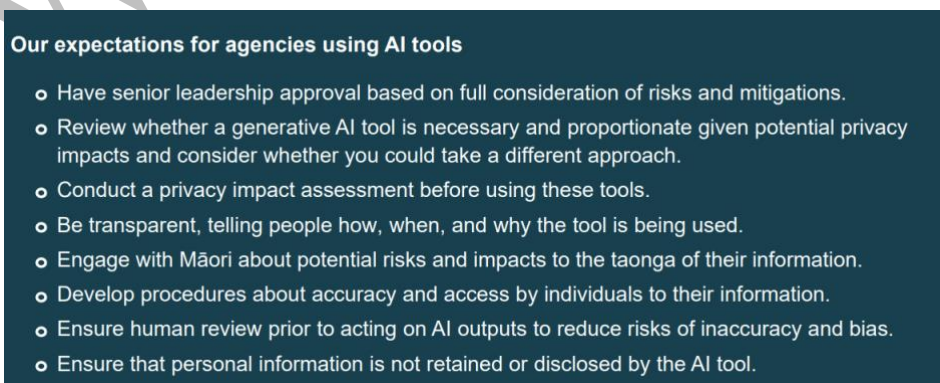
- 
- Our expectations for agencies using AI tools**
- Have senior leadership approval based on full consideration of risks and mitigations.
 - Review whether a generative AI tool is necessary and proportionate given potential privacy impacts and consider whether you could take a different approach.
 - Conduct a privacy impact assessment before using these tools.
 - Be transparent, telling people how, when, and why the tool is being used.
 - Engage with Māori about potential risks and impacts to the taonga of their information.
 - Develop procedures about accuracy and access by individuals to their information.
 - Ensure human review prior to acting on AI outputs to reduce risks of inaccuracy and bias.
 - Ensure that personal information is not retained or disclosed by the AI tool.

Figure 31: Artificial Intelligence and the Information Privacy Principles, (Office of the Privacy Commissioner, 2023, p. 1)

Key questions for collection

- Can you ensure that information about a person is collected from that person? If not, do any of the exceptions apply?
- What risks might be involved in using AI tools without understanding the training data? What steps are you taking to minimise risk? (e.g., no input of personal information)
- Can you choose to use AI tools based on responsibly collected training data?
- Are you clearly explaining to people the ways that you will use and disclose their information, including potential uses of AI tools?
- Are you engaging with communities who might have concerns or perspectives about the potential use of AI tools, including Māori?
- Is there a risk you may be using personal information obtained in a data breach? Can you do due diligence to make sure AI tools are not drawing on illegally obtained data?

Figure 32: Artificial Intelligence and the Information Privacy Principles (Office of the Privacy Commissioner, 2023, p. 6)

Use-case: Using AI tools to screen documents

You may be interested in using AI tools to make decisions based on text or documents, for example screening job applications to find people you want to interview, screening offensive comments online, or detecting text that has been generated by AI. The track record of AI tools in this area is not good, so you need to be very confident the system you want to use will be transparent, accurate, and fair before you ask anyone to rely on it.

You might want to ask:

- How can I find out about the reliability and accuracy of the AI tool for this use-case?
- Is there a risk of bias in the AI tool or the training data?
- Who can I talk with to ensure people are ok with me using this tool? Can I engage with experts? Can I engage with people and communities who might be affected?

Figure 33: Artificial Intelligence and the Information Privacy Principles (Office of the Privacy Commissioner, 2023, p. 9)

The Algorithm Charter (Figure 30 below) is intended for government agencies to “manage how algorithms will be used to strike the right balance between privacy and transparency, prevent unintended bias and reflect the principles of the Treaty of Waitangi” (New Zealand Government, 2020, p. 1).

Commitment:

Our organisation understands that decisions made using algorithms impact people in New Zealand. We commit to making an assessment of the impact of decisions informed by our algorithms. We further commit to applying the Algorithm Charter commitments as guided by the identified risk rating.

Algorithm Charter Commitments:

TRANSPARENCY

Maintain transparency by clearly explaining how decisions are informed by algorithms. This may include:

- » Plain English documentation of the algorithm,
- » Making information about the data and processes available (unless a lawful restriction prevents this),
- » Publishing information about how data are collected, secured and stored.

PARTNERSHIP

- Deliver clear public benefit through Treaty commitments by:
 - » Embedding a Te Ao Māori perspective in the development and use of algorithms consistent with the principles of the Treaty of Waitangi.

PEOPLE

- Focus on people by:
 - » Identifying and actively engaging with people, communities and groups who have an interest in algorithms, and consulting with those impacted by their use.

DATA

- Make sure data is fit for purpose by:
 - » Understanding its limitations,
 - » Identifying and managing bias.

PRIVACY, ETHICS AND HUMAN RIGHTS

- Ensure that privacy, ethics and human rights are safeguarded by:
 - » Regularly peer reviewing algorithms to assess for unintended consequences and act on this information.

HUMAN OVERSIGHT

- Retain human oversight by:
 - » Nominating a point of contact for public inquiries about algorithms,
 - » Providing a channel for challenging or appealing of decisions informed by algorithms,
 - » Clearly explaining the role of humans in decisions informed by algorithms.

Figure 34: Algorithm Charter (New Zealand Government, 2020, p. 3)

To whom does the Charter apply? When Charter signatories deem that their respective government agencies are employing algorithms in a manner that “can significantly impact on the wellbeing of people, or there is a high likelihood many people will suffer an unintended adverse impact” (New Zealand Government, 2020, they will use the Risk matrix (Figure 31 below) to determine whether or not the Charter is to be applied.

Risk matrix

Likelihood

<p>Probable Likely to occur often during standard operations</p>			
<p>Occasional Likely to occur some time during standard operations</p>			
<p>Improbable Unlikely but possible to occur during standard operations</p>			
Impact	<p>Low The impact of these decisions is isolated and/or their severity is not serious.</p>	<p>Moderate The impact of these decisions reaches a moderate amount of people and/or their severity is moderate.</p>	<p>High The impact of these decisions is widespread and/or their severity is serious.</p>

Risk rating

<p>Low The Algorithm Charter could be applied.</p>	<p>Moderate The Algorithm Charter should be applied.</p>	<p>High The Algorithm Charter must be applied.</p>

Figure 35: Algorithm Charter, Risk Matrix (New Zealand Government, 2020, p. 2)

Finally, we have the Interim Generative AI Guidance for the Public Sector (New Zealand Government, 2023c). This output is intended to help “Public Service AI practitioners and decision-makers” (New Zealand Government, 2023, p. 2) to “better understand the key risks of using GenAI within the New Zealand Public Service, and mitigations, to support [them] to develop [their policies], standards and plans for responsibly using GenAI within [their] agency context” (New Zealand Government, 2023, p. 2). After defining and highlighting the benefits, and encouraging among the public service the use of GenAI, the document lays out recommendations on how public service employees may engage with GenAI.

Examples of recommendations are provided below (New Zealand Government, 2023, p. 1).

“Avoid inputting personal data, including client data, into GenAI in your network, and exercise extreme caution if personal information is involved”

“We recommend that your agency develops a GenAI/AI policy and standards to guide your agency’s use of GenAI/AI and share this policy with the Government Chief Privacy Officer”

“We recommend understanding important context for Māori and the Crown, including why GenAI is being considered, how it could impact Māori including services to Māori, what Māori data might be involved and its status across “tapu” or “noa,” and how Māori Data Governance might apply”

“GenAI can perpetuate bias and mis/dis information. Understand the limitations and take active steps to check for accuracy when using GenAI outputs, to avoid harm”

“Be aware of the potential security, quality, intellectual property and supply chain risks of publicly available AI and mitigate risks where possible before using AI tools”

Individual Government Ministries/Departments

While many government agencies are “early in their AI journeys” (Ministry of Business Innovation & Employment, 2024, p. 2), we do see some concrete action from some entities. For example, the Ministry of Education released makes available to education professional resources on the use of AI in schools (Ministry of Education, 2024). Guidance covers how to lay out how AI can be used in schools, cultural biases in AI, the use of personal data, and the importance of reviewing AI-generated outputs. Another example is the Data Ethics Advisory Group, convened by the Government Chief Data Steward (New Zealand Government, 2023b), which does limited work in RAI. New Zealand’s Department of Conservation uses AI in multiple programs; for example, to track invasive predators using AI with output from “portable ‘cameras’ to spot predators including rats, stoats and possums which sends information from remote sites to users to allow a rapid targeted response” (Department of Conservation, 2024). The Ministry of Business, Innovation & Employment has assisted the AI Forum (more on the AI Forum in the subsequent section) to develop “cornerstones that could underpin a future AI Strategy for Aotearoa New Zealand” (New Zealand Government, 2022), with the aim of encouraging safe, ethical AI adoption and innovation that is sanctioned and trusted by New Zealanders.

International Engagements

Two noteworthy international RAI-related engagements. New Zealand is a signatory of the Digital Economy Partnership Agreement (Ministry of Foreign Affairs and Trade, 2025), which contains a module on AI to “recognise the importance of developing frameworks for the trusted, safe, and progressive use of AI” (New Zealand Ministry of Foreign Affairs and Trade, 2019), which involves knowledge sharing (including “regular dialogues to share

Singapore’s expertise in developing model AI governance framework” (New Zealand Ministry of Foreign Affairs and Trade, 2019)), aligning principles and practices, and “research collaborations on AI governance and ethics”. New Zealand is also a member of the Global Partnership on AI (The Global Partnership on Artificial Intelligence, 2025), which consists of, among other sub-bodies, a Working Group on Responsible AI that aims to “foster and contribute to the responsible development, use and governance of human-centered AI systems, in congruence with the UN Sustainable Development Goals” (The Global Partnership on Artificial Intelligence, 2023).

3) Non-government Initiatives

As stated previously, we see reasonable non-state actor RAI work largely in the form of commitments and dialogues, of which this section discusses a few.

AI Forum New Zealand is an NGO that provides resources, networking opportunities, advocacy and capacity building for a range of stakeholders to promote the “economic opportunities raised by AI, supporting great applications of AI and emerging New Zealand AI firms and...ensure that society can adapt” (AI Forum, 2022) to AI technology-based changes. On AI governance, the Forum’s AI Governance Working Group provides “thought leadership on the responsible governance of AI in Aotearoa and to develop a curated set of frameworks, tools and approaches drawn from local and international sources to meet the needs of New Zealand organizations” (AI Forum, 2024a). The Forum also organizes the annual Aotearoa AI Summit’s in collaboration with another NGO – NZTech Group (NZTech, 2022). The Summit’s goal to “[harness] the positive power of AI to enable a prosperous, inclusive and equitable future Aotearoa” (AI Forum, 2024b) does include elements of RAI, covering the intersection of AI with themes such as regulation, data privacy and security, the environment, and indigenous affairs.

The Christchurch Call Initiative on Algorithmic Outcomes, while not principally AI-based, supports the “creation of new technology to understand the impacts of algorithms on people’s online experiences” (The Christchurch Call, 2022) and to understand the “role that algorithms and AI systems play in the spread of TVEC [terrorist and violent extremist content] online” (Chen et al., 2024) following the live-streaming and sharing of the video of the March 2019 terrorist attacks on two Christchurch mosques. Maui63 sees the use of AI in wildlife conservation efforts; specifically, through the use of a tracking drone that uses AI to locate, track, and follow dolphins (MAUI63, 2022).

Finally, we see very limited RAI-related research. The think tank Koi Tū: The Centre for Informed Futures based in the University of Auckland broadly does research to address issues “rising from rapid and far-reaching social, economic, technological and environmental change” (University of Auckland, 2025a) with some attention towards AI; for example, in advocating for an “international AI framework to evaluate new digital technologies” (University of Auckland, 2023), which has RAI implications. Victoria University of Wellington’s Centre for Data Science and Artificial Intelligence leveraged AI in a project that used remote-sensing data to tag tree species to aid in the understanding of the “socio-economic impacts of different types of trees in urban spaces...[and] allow for better planning and decision making” (Victoria University of Wellington, 2023a).

4) Key Trends & Outlook of Responsible AI in the country

In sum, New Zealand ranks relatively well on RAI compared to its regional counterparts, with the kind of AI-related engagements we observe indicating serious thought intention on the part of state and non-state actors to advance RAI in the country.

Despite this progress in the form of multiple government plans and initiatives, as well as dialogue and research among non-state actors, there are a couple of points worth considering. First, the formulation of a dedicated AI strategy that lays out a comprehensive approach to Responsible AI. Borrowing best practices from countries with established, multiple-iteration AI strategies through collaboration and knowledge sharing would serve New Zealand well; for example, The Digital Economy Partnership Agreement discussed previously may provide insights from Singapore's National AI Strategy 2.0. Second, greater public-private sector partnerships on RAI in the form of dialogues/conferences would see knowledge and technology sharing that would aid both stakeholders. Third, increasing AI adoption in New Zealand should be seen in parallel capacity building at school and university level, to ensure no citizen is left behind in the wave of technological developments and increasing AI integration.

Working Draft

Part IV: Outlook for Responsible AI in Asia

LIRNEasia, Sri Lanka

As AI adoption accelerates across Asia, governments are grappling with how to govern AI effectively while fostering innovation and inclusion. This section examines four key areas shaping the future of Responsible AI in the region.

RAI and Misinformation

In this section, we examine the relationship between responsible AI and misinformation by doing the following: first, briefly establish why misinformation is a problem; second discuss the roles in which AI can play in supporting and curtailing misinformation; third, present RAI solutions geared towards countering misinformation.

1) The role of AI in misinformation

The problem of misinformation

Misinformation is on the rise (The Economist, 2024), globally. Increasing digitization and internet and mobile use (Tutella, 2024), a sharp rise in the number of social media users and proportion of people for whom content on these platforms is a primary news source (Romeo, 2022), and mounting political polarization (Osmundsen et al., 2021) (primarily in Western contexts) all help drive an increase in the proportion of citizens who are exposed to misinformation, the volume of that misinformation, the likelihood that people buy into misinformation, and the probability of further dissemination (American Psychological Association, 2023).

Why is misinformation a Responsible AI matter, anyway?

Misinformation is a responsible AI issue in two ways: through the direct harms of AI-driven misinformation, and harmful responses by governments attempting to curtail the production and spread of AI-driven misinformation, both of which can undermine human rights. Impacts of misinformation range from the widening of social cleavages, online abuse and harassment (European Foundation for South Asian Studies, 2021), and physical violence (Chopra, 2021).

For example, politicians and health workers faced abused ranging from vitriol to death threats by people driven by conspiracy theories surrounding COVID-19 countermeasures such as vaccinations and lockdowns (Glitch Charity, 2020) (the prevalence of people not adhering to these countermeasures was also an issue). In 2018, fake news spread over WhatsApp alleging a child kidnapping ring by a group of men in Kerala, India led to a spate of mob lynchings that killed around 25 people (Biswas, 2018). In Sri Lanka, instances of misinformation levelled against the Muslim population by Sinhala-Buddhist extremists led to in 2018, mob violence levelled against Muslim communities in Kandy, Sri Lanka that caused destruction of property and death (United Nations Sri Lanka, 2024) (Facebook, following blowback, apologized for the role content on the platform may have exacerbated the issue) (Al Jazeera, 2020). AI-generated misinformation can undermine democracy by distorting the information landscape during elections; for example, with deepfakes of a Bangladeshi opposition lawmaker clad in a bikini (Swenson & Chan, 2024), and deepfakes of public figures

“routinely [featuring] misleading electoral messaging” (Deeplina et al., 2024) during India’s 2024 National Election. State responses to counter misinformation can also be harmful. Due to purported AI-driven misinformation, authoritative governments can respond with blanket bans or shutdowns of social media and possibly even restrict freedom of assembly, raising concerns about freedom of expression and the digital rights of citizens (Levush, 2019); for example, as seen with the Sri Lankan government’s social media ban following the 2019 Easter Sunday Terror Attacks (Wakefield, 2019).

AI’s role in misinformation generation

Advancements in generative AI models (e.g., large language models or deepfake video tools) make it easier and cheaper to create deceptive media (Deutsche Welle, 2024). These can be deployed to spread misinformation (e.g., false speeches and actions by public figures, like a deepfake piece of audio intended to show Philippine President Ferdinand Marcos Jr ordering a military strike (Enriquez, J. M., 2024); manipulated “evidence” of events) on social media platforms. Some, however, argue that that that “current concerns about the effects of generative AI on the misinformation landscape (at least on the political front) are overblown” (Simon et al., 2023), and AI has “not fundamentally changed the landscape of political misinformation” (Kapoor & Narayanan, 2024).

AI’s role in misinformation propagation

AI-driven bots can amplify misinformation at scale, flooding discussion channels and social media with misleading or false narratives (BBC Bitesize, n.d.; Ricciardone, 2024). Research finds that, as much as human-generated content can feed into bots, content spread by bots, even “language structures” can influence human content. Programming a large number of bots to coalesce around a certain viewpoint can help lend credence to said viewpoint and do a better job at influencing real people. A notable example is a social media government agenda-aligned campaign that deployed bots on Twitter (now X) and Facebook to shape the narrative surrounding separatist movements in Indonesia’s Papua province (Strick, B., & Syavira, F., 2019). The bots “jump[ed] on to hashtags being used by groups supporting independence, such as #freewestpapua, so they swamped negative reporting with positive stories about investment in the region, a process known as “hashtag hijacking””. We also know that AI is used in ad and content targeting, for example by Meta (Meta, 2025) and Google (Google Ads, n.d.), which becomes a problem here when the content being recommended contains misinformation.

AI’s role in combatting misinformation

AI can be used in counter-misinformation initiatives. For example, Taiwanese fact-checking outfit MyGoPen finds that AI helps to “speeds up the checking process – helps with comparison, identification and translation”. The outfit also uses AI to generate transcripts, which “improves the efficiency of checking video and audio materials”, and to drive bots that help respond to requests received by the platform. Another example is Dissect, an in-development tool in Sri Lanka that aims to aid journalists in the fact-checking process by being a “first-in-line critical reader” (LIRNEasia, 2023). AI can also be used to “reverse engineer manipulated images and videos to detect deep fakes and highlight content that needs to be flagged” (Goswami, 2022).

The use of AI tools, however, poses its own challenges. For example, bias or a lack of accuracy in AI-driven tools. cultural and linguistic diversity. AI-based models trained on predominantly English-language data may not perform reliably in local Asian languages (Alghamdi et al., 2024) (Asia comprises many countries with different languages, scripts,

dialects, cultural norms, and political contexts; and there is a significant dearth of tools that capture these contexts). Therefore, AI-driven detection systems may struggle to identify misinformation in less-resourced languages, such as many South Asian languages. Additionally, AI tools that monitor, flag, or intercept communications for misinformation risk overreach, potentially infringing on individual privacy rights. It is crucial that both state and non-state actors developing these tools ensure they take measures to ensure user privacy. Failing to do so could lead to the very human rights violations that counter-misinformation efforts aim to prevent.

2) The role of Responsible AI in countering misinformation

In this section, we explore potential responses, grounded in responsible AI principles, to address the challenges at the intersection of AI and misinformation.

- **Contextualized Datasets and Local Partnerships:** Working with local community organizations, fact-checkers, linguists, and researchers is key in capturing linguistic and cultural diversity that can lead context-rich to datasets that help with developing AI-driven tools. This collaboration is vital to improve the accuracy of AI models in identifying misinformation and understanding cultural nuances. For example – recognizing rampant, socially destabilizing misinformation in Myanmar and the “lack of resources and fact checking datasets in Burmese [being] a major obstacle” (York Centre for Asian Research, 2024) to efforts from parties like Facebook leveraging AI to manage harmful content – one study created a “large fact checking dataset and NLP models for Burmese”.
- **Foster (fund, recognize, collaborate for) more local research:** Beyond richer AI datasets, but also investing in robust local research (social science, psychological etc.) that integrates social, political, and infrastructural nuances. Funding interdisciplinary teams and fostering regional expertise will ensure that efforts to slow down spread and the mitigate harms of [AI & The Misinformation Disorder] will meet the demands of complex information landscapes.
- **Algorithmic Transparency and Explainability:** As part of a Responsible AI framework, platforms and AI developers should publish policies and technical explanations of how misinformation detection systems work—especially when making moderation decisions in sensitive contexts like politics or religion. Given that there are already formal guidelines in place to mandate or encourage transparency around the use and development of AI (Chen et al., 2024; Data Guidance, 2024; Monetary Authority of Singapore, 2018; University of the Philippines, 2025), it would not be difficult to extend this to the use of AI in misinformation detection systems.
- **Educative initiatives to promote AI literacy:** May help improve the ability of citizens to determine whether a piece of media is AI-generated. As AI capabilities currently stand, there are certain qualities of AI-generated content (writing or image/video generation/manipulation) that point to its origin. Exposure to AI tools from an early age, therefore, would be a good way to familiarize oneself with the outputs of those tools and thus better recognize whether AI has a hand in generating a given piece of media. There is initial work on this, for example, MIT Media Lab’s DetectFakes experiment that allows users to understand how well they can identify AI-generated images (MIT Media Lab, n.d.) and resource on how to determine if an image is AI-generated (Kamali et al., 2024); or OECD’s indicator measuring the ability of adults to spot AI-generated disinformation online (Organization for

Economic Co-operation and Development, 2025). However, we are yet to see these scaled to concrete, large-scale initiatives at school and university levels in Asia.

- **Human-in-the-Loop Systems:** We have discussed in the previous section the use of AI in helping fact-checkers and journalists. However, the fact that “AI hallucinates and cannot be trusted fully yet” (Sperling, 2024) is an impediment to the use of unsupervised AI here. While automated tools can help moderate content at scale, human reviewers with local context are crucial for handling edge cases, highly sensitive political speech, satire, or cultural references that AI might miss. This hybrid approach would help ensure that false positives and false negatives are minimized.

Working Draft

AI & Data Protection in Asia

1) Introduction

In the current AI paradigm—dominated by machine learning—data serves as the foundational resource, powering everything from social media recommendation systems to healthcare diagnostics (King & Meinhardt, 2024, p. 17). As the line between personal and non-personal information continues to blur, these systems frequently clash with established data protection principles. Meanwhile, the volume of data collected has grown exponentially, spurred by both private-sector platforms (social media, ride-hailing, e-commerce) and governments embracing digitization for public service delivery (Asian Productivity Organization, 2024). The result is an environment where large-scale datasets fuel AI innovation but also raise complex ethical and policy questions.

Asia's data governance landscape is characterized by a surge in new data protection regulations and significant regulatory diversity (Morrison Foerster, 2023). While some economies have comprehensive data protection laws, others have weak or incomplete laws and uneven enforcement capabilities (Squire Patton Boggs LLP, n.d.). This shortfall in regulatory expertise and resources can significantly undermine efforts to govern AI responsibly. Additionally, data localization and sovereignty requirements—often enacted to protect citizens' data—can restrict cross-border data flows and complicate international collaboration and data security concerns. Layered onto these challenges are diverse cultural and social norms surrounding privacy and consent, which influence how data protection policies and responsible AI frameworks are adopted locally. Furthermore, public-private partnerships are evolving rapidly as governments collaborate with tech companies to expand digital services, necessitating meticulous oversight to safeguard citizen rights and public interest.

The Global Index on Responsible AI identifies data protection and privacy as an essential consideration under the dimension of Human Rights. Likewise, other leading frameworks—such as the OECD AI Principles (OECD, n.d.) and UNESCO's Recommendation on the Ethics of AI (UNESCO, 2021)—highlight the importance of safeguarding privacy while ensuring accountability, transparency, and security throughout the AI lifecycle. However, blanket bans or overly restrictive regulations on personal data usage are neither feasible nor desirable. Striking a balance between privacy protection and the responsible use of personal data for AI remains a core challenge for policymakers and practitioners in Asia and beyond.

This section explores the fundamental tension between data protection and AI, highlights a pivotal development in Asia that heightens the importance of safeguarding personal data in the age of AI, and discuss key actions to ensure AI-driven innovation is pursued responsibly while protecting individual rights.

2) The Fundamental Tension & Why it Matters

AI's reliance on personal data challenges traditional notions of data protection. As the line between personal and non-personal data blurs, seemingly benign information can be re-identified when combined with direct identifiers (Lubarsky, 2017). AI often relies on vast,

diverse datasets that extend beyond their original purpose of collection. In healthcare, for example, AI can analyze past electronic health records—highly sensitive personal data initially collected without AI use cases in mind—to identify older adults at high risk of heart attacks. However, such use may conflict with the principle of purpose specification, a cornerstone of most data protection laws. Similar tensions arise with other core principles, such as collection limitation, storage limitation, and data minimization, which can constrain AI's ability to process diverse data effectively.

At the core of this debate lies the tension between data use and protection. AI holds significant promise in fields like healthcare, agriculture, education, and social assistance, where it can deliver life-changing benefits—especially in resource-constrained regions of Asia. In Malaysia, the Automatic Road Incident Detection System (ARIDS) is being tested as an AI-based solution for identifying traffic anomalies and accidents, with early results suggesting it could reduce emergency response times significantly (Digital Watch Observatory, 2024). However, these promises come with questions about privacy and data protection, such as how and where vehicle and pedestrian data are collected and stored. In India, “Outgrow”, provides AI-powered agricultural advisory services, offering insights on crop health, soil conditions, weather, and market prices to farmers across Tamil Nadu, Karnataka, Maharashtra, Andhra Pradesh, and Telangana (Outgrow, n.d.). However, there are important considerations regarding the collection, storage, and potential misuse of farmers' personal, geospatial, and agricultural data, risks of unauthorized access, and lack of transparency in how AI-driven insights are generated and shared with third parties. This inherently creates a complex challenge for policymakers who must balance innovation with the protection of individual rights.

3) Digitization and Digital Public Infrastructure: Innovation Meets Intrusion

Increasingly embraced across Asia as a key driver of efficient public service delivery, Digital Public Infrastructure (DPI) and government digitization efforts promise improved governance and AI-driven innovation. India has emerged as a global leader in this space, with initiatives like Aadhaar, UPI, and Co-WIN demonstrating the transformative power of DPI in areas such as identity, payments, and public health. AI is already part of these systems. In India, Bhashini is currently being built, an AI-led language translation system that aims to integrate with India's other DPI systems. Bhashini utilizes natural language processing (NLP), a machine learning technique most used with text, to perform rapid translations between Indic languages (Stanly, 2023). Singapore is using AI to combat fraud in DPI, as Singapore's national DPI, Singpass, recently began using a machine learning to authenticate the users of the DPI system, helping improve citizens' trust in the system, which can facilitate broader uptake (Theodorou et al., 2024).

Following these examples, many Asian countries are adopting or adapting similar digitization efforts packaged with AI, accelerating the spread of DPI in the region (Hirdaramani, 2024). However, many such systems collect large volumes of sensitive citizen data, often bypassing informed consent. While proponents argue these initiatives support governance or national security, they can also enable profiling, surveillance, and other intrusive practices. As a result, data gathered under the guise of public interest may be repurposed for AI training without sufficient transparency or accountability.

AI systems built on DPI data amplify these risks. Without robust safeguards, they may inherit or worsen societal biases, further marginalizing vulnerable populations (Nagar & Eaves, 2024). Excessive data collection—which can be justified under different grounds despite the existence of principles like data minimization—creates opportunities for misuse and undermines public trust. In regions with low trust in government, these factors discourage participation, resulting in datasets that are neither representative nor equitable. Transparent governance frameworks are thus essential to ensure AI systems derived from DPI data remain both effective and ethical, balancing their transformative potential with the protection of individual rights.

4) Progress, Gaps, and Challenges of Data Protection Laws in the AI Age

Due to long-standing discussions on digital privacy since the early internet era, data protection and privacy consistently score high in the Global Index on Responsible AI as a thematic area—every region except the Americas has it ranked among the top scoring thematic areas. Reflecting this trend, many Asian countries already have data protection laws or are in the process of drafting them. India’s Digital Personal Data Protection Act (2023) and Vietnam’s Law on Data (2024) are recent examples of a broader push to strengthen data governance with implications for AI.

However, having a law does not guarantee its effectiveness. Some frameworks, like Singapore’s Personal Data Protection Act (PDPA), directly address AI’s data-heavy nature by mandating accountability, consent, and breach notifications. Singapore’s Personal Data Protection Commission (PDPC) further supports transparency, fairness, and risk management through dedicated AI governance guidelines (Personal Data Protection Commission Singapore, 2024). In contrast, Sri Lanka’s Personal Data Protection Act includes basic safeguards for AI-related risks but is overseen by a newly formed Data Protection Authority that has yet to prioritize AI oversight. On the other end of the spectrum, Nepal’s data protection laws remain in early development, with observers citing gaps in clarity and comprehensive provisions.

In practice, many Asian countries resemble Sri Lanka or Nepal more than Singapore, facing persistent obstacles to regulating AI through a data protection lens. Data Protection Authorities (DPAs) often work with limited budgets, hampering their ability to hire skilled staff, invest in necessary technologies, and conduct thorough investigations. The shortage of AI and data protection expertise further restricts DPAs’ capacity to understand and manage complex AI systems (Paulger et al., 2024). These constraints underscore the region’s ongoing struggle to ensure that data protection laws meaningfully safeguard citizens in an AI-driven world.

5) Looking Ahead

In the AI era, data governance demands concerted efforts from policymakers, regulators, industry, civil society, and international organizations. A critical first step is to strengthen existing data protection laws by refining statutory definitions, expanding breach notification requirements, and clarifying provisions on automated decision-making. Equally important are robust enforcement mechanisms, including clear penalties and proactive investigations, to demonstrate genuine commitment to safeguarding personal data.

As already seen in some countries in the region, regulators can further support compliance through advisories, opinions, and guidelines that promote privacy-by-design principles, anonymization practices, and ethical AI deployment (National Privacy Commission, 2024). Meanwhile, Data Protection Authorities (DPAs) require substantial resources and specialized expertise to conduct thorough investigations and adapt to emerging AI challenges. Civil society groups and the media play a vital role as well, educating citizens on data rights and risks while monitoring potential abuses, ensuring marginalized voices are heard (EngageMedia, 2021).

Finally, international collaboration, such as those seen in the ASEAN Guide on AI Governance and Ethics (ASEAN, n.d.) is essential for consistent AI governance across borders. Aligning data protection standards, sharing best practices, and working with agencies responsible for competition, human rights, and consumer protection help create holistic regulations that uphold public trust. By bridging gaps and coordinating efforts globally, Asian countries can harness AI's transformative potential without undermining fundamental privacy rights.

Working Draft

Regulating AI in Asia: Guiding Questions and Key Trends

1) Introduction

As conversation on Responsible AI take shape across the world, governing AI through regulation has become a topic of great interest in the Asian region. Although the European Union’s risk-based AI Act provides a global reference point, scholars have argued that its impact in Asia may be limited because many countries prioritize AI-driven economic growth and are cautious of restrictive rules (Pagallo, 2023). At the same time, limited regulatory capacity in Asia complicates efforts to oversee AI effectively. As a result, some governments are turning to softer tools, such as industry codes and non-enforceable guidelines (Sia, 2024). This shift takes place against a backdrop of global semi-conductor competition, an emerging race to build frontier AI models, and growing investments in digital public infrastructure, adding urgency to discussions on AI regulations.

Despite the general reticence, Asia already exhibits a diverse mix of approach to regulating AI, ranging from formal, AI-specific legislation to adaptations of existing laws in areas like cybersecurity, data protection, and intellectual property. In some cases, regulations focus narrowly on specific technologies, such as recommender systems or generative AI. To navigate this complex landscape, stakeholders in the region could benefit from a structured framework to evaluate regulatory decisions that directly or indirectly impact AI development and application in Asia. This section clarifies AI regulation by addressing two key aspects: first, it explores three essential questions policymakers should consider, drawing from UNESCO’s Consultation Paper on AI Regulation; second, it examines current regulatory approaches across several Asian countries, illustrating how the unique country contexts can lead to different approaches to regulation.

2) Part I: the why, the when, and the how

The UNESCO’s Consultation Paper on AI Regulation (UNESCO, 2024, p. 40) offers three key questions to consider before adopting AI regulations.

1. Why regulate?
2. When to regulate?
3. How to regulate?

The question of “**why regulate?**” examines the conditions under which AI regulation might be justified. The paper identifies three primary reasons for regulation. First, it can address public problems such as market failures, state failures, or unacceptable risks posed by AI. Second, regulation may be necessary to protect, respect, and promote fundamental and collective rights, a rationale closely aligned with the approach advocated by the Global Index on Responsible AI. Third, policymakers may see regulation as a means to achieve societal goals, such as improving quality of life or advancing developmental objectives—ambitions reflected in AI strategies of many developing countries across Asia. These reasons are not mutually exclusive and must be tailored to the specific context. Importantly, whether regulation is the right approach depends on the institutional, legal, political, economic, social, and cultural factors unique to each jurisdiction.

If regulation is indeed justified as desirable in a given jurisdiction, the second question is **"when to regulate?"** (UNESCO, 2024, p. 41). This is, in fact, a series of structured sub-questions, as illustrated in the decision tree below. This decision tree points to when regulation is the most appropriate course of action and guides the actors in selecting the right regulatory tool. The process begins by assessing whether there is a clear justification for regulation and identifying the specific objectives it aims to achieve. If a justification exists, it must be evaluated whether suitable regulatory tools are available to meet those objectives.

Next, it needs to be considered whether other policy tools—such as incentives, guidelines, or voluntary standards—might be more effective, efficient, or equitable than regulation. If regulatory tools are still deemed necessary, the feasibility of regulation must then be assessed, accounting for legal, political, and administrative factors. Only if regulation is both necessary and feasible should policymakers proceed to explore specific regulatory approaches. At each stage, the framework encourages revisiting earlier steps if conditions are not met, ensuring a thorough and context-sensitive evaluation process. (UNESCO, 2024, p. 42).

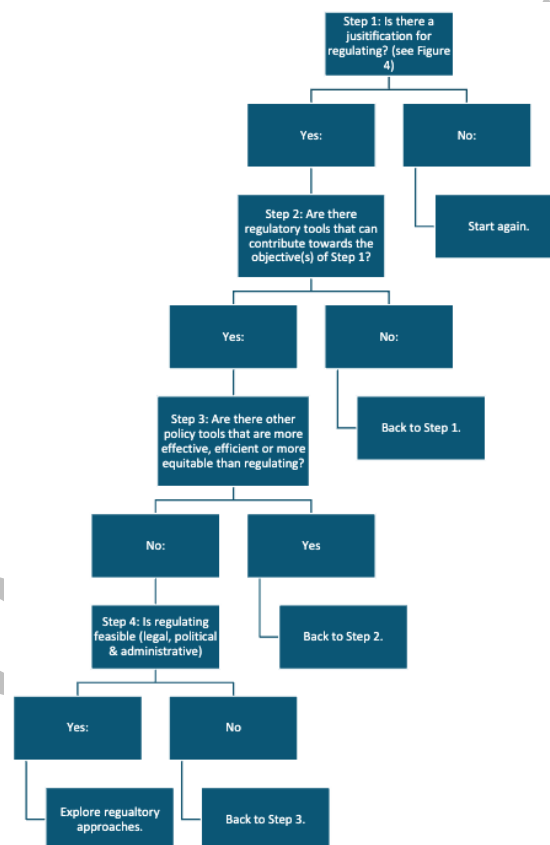


Figure 36: Decision tree for deciding to regulate (UNESCO, 2024, p. 43)

Having addressed the motivations and timing for AI regulation, the third guiding question is **"how to regulate?"** (UNESCO, 2024, p. 44). A tailored approach is essential, combining AI-specific legislation with adaptations of existing laws in cybersecurity, data protection, and intellectual property. In this context, it's crucial to address a given jurisdiction's specific societal challenges—such as addressing labor market implications of AI, curbing the spread of misinformation, and ensuring data privacy in biometric surveillance systems. Policymakers should also consider the impacts on marginalized communities, ensuring that AI benefits are equitably distributed and that regulations actively mitigate digital divides.

To keep pace with rapid technological advancements, agile regulation is key. This means adopting flexible regulatory tools, like sandboxes and testbeds, to experiment with policies in controlled environments before broad implementation. Learning from global sandboxing experiences — such as those in Singapore (Tan & Taeihagh, 2019) — can provide valuable insights into how to adapt these models for local contexts. Equally important is process, ensuring that diverse voices—from civil society, industry, and affected communities—are meaningfully included in the regulatory dialogue. Finally, rather than imposing a one-size-fits-all framework, countries are best served by focusing on targeted interventions based on specific policy challenges, continually monitoring, and adjusting regulations as AI technologies and societal impacts evolve.

3) Part II: AI regulation in Asia

Having addressed the foundational questions of why, when, and how to regulate AI, we now turn to what AI regulation looks like in practice across Asia. With its diverse socio-economic contexts and varying levels of AI adoption, the region presents a wide spectrum of regulatory approaches. This section explores the regulatory environments and discourse in key countries—South Korea, Japan, Singapore, China, and India—ranging from binding laws to non-binding guidelines and frameworks. While the treatment is not in-depth, the discussion outlines the key features of each country’s approach, providing a foundation for understanding the diversity and complexity of AI governance in the region.

South Korea’s AI Basic Law, aka the South Korean AI Act (SKAIA) is the second AI-specific piece of legislation after the EU’s AI Act (Artificial Intelligence Act, 2025). SKAIA, which will take effect in January 2026, aims to “protect citizens' rights, improve quality of life, and boost national competitiveness” (Data Guidance, 2025). The law considers three key areas: the launching of an “organizational system” (Artificial Intelligence Act, 2025) (including a National AI Committee), supporting the development of AI, and ensure safe and reliable bases for high-risk and generative AI”. Unlike the EU AI Act, which sets out “substantive laws and regulations”, SKAIA provides “operating rules placing obligations on the Ministry of Science and ICT and other related authorities”. South Korea’s 2019 National Strategy for Artificial Intelligence (The Government of the Republic of Korea, 2019) aims to position the country as a leader in AI by 2030, and calls for, among many things, the improvement of AI infrastructure, an emphasis on AI R&D, and the development of AI skills. South Korea sees a number of laws that indirectly apply to AI (International Association of Privacy Professionals, 2024, p. 22), like the Personal Information Protection Act, Monopoly Regulation and Fair-Trade Act, and Copyright Act.

Japan, on the other hand, currently lacks any laws that directly regulate AI and is said to be “taking an indirect approach to support the policy goal of prioritizing innovation while minimizing foreseeable harms” (White & Case LLP, 2024). The country’s 2022 AI Strategy (Council for Science, Technology and Innovation Policy, 2022) reflects this ethos, calling for “agile governance” (International Association of Privacy Professionals, 2024); i.e., where government “provides nonbinding guidance and defers to the private sector's voluntary efforts to self-regulate”. The Strategy contains the expected points on safe AI systems, building infrastructure to support those systems, and R&D. In 2019, the Japanese government released the Social Principles of Human-Centered AI, which contains principles that stress that AI must not violate human rights, privacy must be protected, and that AI innovation be “promoted through the close collaboration and cooperation of domestic and international

stakeholders” (Nayak, 2024). Other laws such as the Improving Transparency and Fairness of Digital Platforms Act, Financial Instruments and Exchange Act, and Protection of Personal Information Act also indirectly pertain to AI (International Association of Privacy Professionals, 2024, p. 16).

Singapore has a fairly rich AI regulatory environment that, much like Japan, does not resort to direct AI-related legislation, instead adopting “soft law” (Chng & Jones, 2024) approaches. Much of the country’s AI regulation is supported by the National AI Strategy 2.0 (NAIS 2.0). There is also the Model AI Governance Framework for Generative AI, which aids private sector entities deploying AI solutions in addressing ethical and governance issues (Infocomm Media Development Authority, n.d.). Among those principles is an emphasis on ensuring accountability among those in AI development, addressing security issues linked to generative AI, and supporting R&D. The Monetary Authority of Singapore (MAS) released a set of “generally accepted Principles for the use of artificial intelligence and data analytics (“AIDA”) in decision-making in the provision of financial products and services” (Monetary Authority of Singapore, n.d., p. 3). As stated in the Singapore case study component of this report, Singapore’s AI Verify tool goes a step further than standard soft-approach regulatory tools by also providing a clear incentive to entities to adhere to those principles through the self-assessment – the report serves as a badge or certificate of sorts which entities. Other laws like Singapore’s Personal Data Protection Act, Computer Misuse Act, Copyright Act and Cybersecurity Acts also do indirectly apply to AI (IAPP Research and Insights, n.d., p. 21).

China’s approach is generally perceived to be ‘proactive’ in introducing legislation that pertains to AI. The Artificial Intelligence Law of the People’s Republic of China (Center for Security and Emerging Technology, 2024) – a preliminary draft last documented as being circulated for legal scholars – specifies various scenarios in which AI developers, providers, or users are liable for misuse of AI tools” and “provides intellectual property protections for content created with the assistance of AI technology”. The draft also contains provisions on how state organs may use AI in administrative work, the misuse of AI, the protection of rights of AI users, and building AI skills. Other laws relevant to AI are the Interim Measures for the Management of Generative Artificial Intelligence Services (China Law Translate, 2023), Internet Information Service Algorithmic Recommendation Management Provisions (DigiChina, 2022), and Deep Synthesis Management Provisions (Zhang, L.,2023) that addresses concerns arising from specific types of AI.

India is said to have, since 2022, “oscillated between a hands-off approach to AI regulation and one that is more direct and interventionist” (Mohanty & Sahu, 2024). As things stand, the country sits more on the ‘non-binding’ rules side of the spectrum. The National Strategy for Artificial Intelligence is “centered around leveraging AI to drive economic growth, improve social outcomes, and address key national challenges across various sectors” (The Digital Watch, 2018). The Strategy consists of three Pillars that concern the economic impact of AI, AI for “social development and inclusive growth”, and India as a “global hub for developing scalable AI solutions that can be applied in other emerging and developing economies”. The Principles for Responsible AI delves in to the importance of responsible AI, and details systems and societal considerations (NITI Aayog, 2021). Other laws that indirectly relate to AI are the Information Technology Act (Government eMarketplace - Central Public Procurement Portal, 2000), Competition Act (Ministry of Corporate Affairs, 2002), and Digital Personal Data Protection Act (Ministry of Electronics & Information Technology, 2023).

AI regulation is a balancing act between having some degree of control over an ever-pervasive technology while avoiding creating an environment that stifles innovation and adoption. As AI governance gains traction in Asia, regulatory approaches remain diverse, shaped by economic priorities, institutional capacity, and policy preferences. While some countries, such as South Korea, have introduced AI-specific legislation, others, like Japan and Singapore, rely on non-binding guidelines and industry-driven frameworks. We have yet to see rigorous study that evaluates the impacts of these regulatory approaches. Where these approaches converge appears to be in the principles they espouse and the kind of actions they call for – improving socioeconomic outcomes, research and development, and innovation conducive to growth. With increasing AI innovation and integration in society, where on the spectrum of approaches governments may decide to lean has yet to be seen.

Working Draft

Responsible AI in Government

From social welfare delivery (Aiken et al., 2022) to disaster management (Yossi, 2020) to public health (Sarkar & Singh, n.d.), Artificial Intelligence is seeing increasing adoption in government. In developing country contexts, especially in Asia, these conversations are further energized by ongoing interest in and the development of digital public infrastructure (DPI). AI applications built on DPI hold immense potential for improving public services, yet they also present a range of challenges.

Concerns include the risk of AI solutions being wasteful, harmful, or exclusionary to certain communities—challenges that are further intensified in settings with low digital literacy, where individuals may struggle to understand or challenge AI-driven decisions. Poor deployments of AI and the misuse of the technology by authoritarian governments—often justified under the guise of national security or operational efficiency—add further complexity. As citizens increasingly transition into digital spaces, governments gain access to vast, detailed, and multi-faceted datasets, which can underpin impactful AI applications. However, ensuring that these innovations do not lead to harm, misuse, or failure is critical, particularly for historically marginalized populations. Striking a balance between the promise of innovation and the risks associated with misuse is essential for responsible deployment of AI in the Asian region (World Bank, n.d.).

Why Does This Matter in the Public Sector?

AI adoption in the public sector comes with distinct dynamics that set it apart from private sector applications (qBotica, 2023). Resource constraints, including limited budgets, time, and skilled personnel, frequently hinder governments from leveraging AI effectively. At the same time, public services operate at scale, impacting large and diverse populations. Marginalized communities, often the primary beneficiaries of certain types of public programs (e.g., social assistance for the poor), are particularly vulnerable to inequities, raising the stakes for AI deployment. Even well-intentioned initiatives can lead to harm, such as discrimination, resource misallocation, or safety concerns, when sufficient capacity is lacking.

More alarmingly, AI systems can be misused as tools for authoritarianism, surveillance, or social control in less transparent governance environments. Resistance to change from those benefiting from legacy systems can further complicate adoption efforts, perpetuating inefficiencies and missed opportunities. Understanding these unique challenges underscores why responsible AI adoption is critical for ensuring public trust and maximizing societal benefits. Furthermore, the public sector carries a higher burden of accountability, as failures in AI systems can even result in erosion of public trust in the democratic process.

Governments adopting AI must navigate the delicate balance between mitigating the aforementioned harms and fostering innovation. Responsible AI adoption requires thoughtful strategies that align with ethical principles, ensuring AI systems are equitable, transparent, and accountable. The Global Index on Responsible AI includes several thematic areas that can guide the development of such systems (Adams et al., 2024). While many of the thematic areas have a bearing on the public sector adoption of AI, in the sections below, we identify a crucial selection of them, as key principles, implementation focus areas:

Key Principles for Responsible AI Adoption in Government

- 1. Bias and Unfair Discrimination:** Bias in AI systems occurs when outcomes are unfairly prejudiced against individuals or groups based on characteristics such as race, gender, or nationality (Gray et al., 2024). This issue is especially concerning in public sector AI applications, where historical inequalities embedded in society and traditional systems can be reinforced by AI. While it may not be possible to eradicate bias from AI systems, it is possible to mitigate some of the harms by understanding: (1) where bias originates and how it manifests; (2) who bias harms and who it benefits; (3) how bias appears in the data; and (4) how bias can reinforce discriminatory practices. With this understanding will come greater clarity about what measures governments can take to leverage the potential of AI while mitigating the risks to ensure responsible and ethical use of AI.
- 2. Transparency and Explainability:** Transparent AI systems enable citizens and stakeholders to understand how decisions are made. While there is a tendency to reduce transparency and explainability to technical explanations the minute details of the "inner workings" of AI systems, far more important, especially within AI use cases in government, is the quality of enabling people to understand how a particular system works and/or how a particular outcome was achieved by providing information that is sensible and easy-to-understand. Further, as scholars have pointed out, explainability and transparency requirements, even within a single public sector AI sector application can be different. The level of simplicity and transparency required by policymakers and regulators will be different from that is needed by frontline public sector workers or individuals affected by the decisions of AI systems (Amarasinghe et al., 2023).
- 3. Responsibility and Accountability**

When AI systems do not function as expected, a fundamental question arises around who should be held accountable and responsible. This is all the more important in AI adopted in government, to mitigate the risk of harm and to uphold the promise and potential of trustworthy AI. Many international frameworks such as UNESCO Recommendation on the Ethics of AI (UNESCO, n.d.), and OECD's AI principles (OECD.AI, n.d.) recognize accountability as one of the core principles of ethical adoption of AI. Within the context of public sector AI applications, it is essential to establish clear roles and accountability structures to ensure responsible management and oversight. This includes defining ownership and decision-making hierarchies at every stage of the AI lifecycle. Additionally, implementing auditable AI systems is crucial for tracking decisions and ensuring transparency (Loi & Spielkamp, 2021). To further strengthen accountability, establishing legal and regulatory frameworks, where they are lacking, provides a foundation for holding parties responsible for AI system outcomes.

Key Implementation Focus Areas for Responsible AI Adoption in Government

1. Algorithmic Impact Assessments

Impact assessments are crucial to designing AI systems that reflect the needs and values of all citizens (Treasury Board of Canada Secretariat, 2024). In the context of public sector AI applications, impact assessment tools, such as those unveiled by the data authority of New Zealand (data.gov.nz, n.d.), can be used as a tool for both predicting the anticipated, and assessing the actual, consequences to the public of an AI system in terms of benefits and harms. When assessing anticipated risks, impact

assessments require users of AI technologies to consider the type and scope of impact before and after deployment. Accordingly, there are impact assessments of two main types. Ex-ante impact assessments involving analysis potential benefits and harms of an AI system prior to deployment and Ex-post impact assessments analyze the actual benefits and harms imposed by an AI system after making the decision to deploy.

2. Public Participation and Awareness

Public participation as a principle refers to the process of engaging ordinary people in the decision-making processes that will affect their lives and giving full consideration to their input before making such decisions (Gilman, 2023). Public participation with respect to AI is important at several levels (Hu & Singh, 2024). All citizens should have sufficient knowledge and understanding of Artificial Intelligence, be familiar with it to identify ordinary ways of using it in their everyday life and avert harmful uses and applications of AI. Second it is also important that, the public is aware of the broader emerging consequences of AI and take part in the policy make processes as stakeholders and engage in collective decision-making processes which serve their lives. However, public sector applications of AI present a third important dimension of participation, where the public gets involved in developing and shaping individual applications of AI through their expertise and viewpoints. Recruiting ordinary citizens, especially users and stakeholders, to partake in AI design has become increasingly prevalent within the AI community, giving rise to participatory AI projects. The forms of citizen participation include commenting on design ideas through questions and interviews, group discussions with project teams, and collaborative prototyping and decision making (Delgado et al., 2023). However, as critiques have pointed out, meaningful participation requires thinking through many dimensions to ensure that the participatory process of AI application development and development is effective (Feffer et al., 2023).

3. Public Sector Skills Development

Frontline government workers—ranging from social workers and law enforcement officers to hospital staff—make critical decisions affecting people's lives daily. Operating under growing demand and limited resources, these workers must make quick, important decisions with little information. In this environment, artificial intelligence (AI) can support decision-making in public service delivery. While these workers don't need advanced technical knowledge to use AI effectively, they do require competencies beyond basic education to function as users and facilitators of AI systems in public service. For example, under Phase 3 of its Digital India program, the Ministry of Electronics and Information Technology (MeitY) recently launched nationwide training programs to equip Central and State officials with skills to manage large digital transformation projects (Singha, 2024). These programs cover topics such as Digital Public Infrastructure, Contract and Procurement Management, AI & Machine Learning, Digital Governance, and Data Management. Similar training initiatives will be essential moving forward for the successful adoption of AI in Government in the Asian region.

4. Public Procurement

Public procurement plays a crucial role in the responsible adoption of AI in government, as it can significantly improve public service delivery when implemented properly. However, due to the emerging nature of AI technology, adhering to standard procurement processes, engaging effectively with suppliers, and

establishing appropriate criteria and terms for ethical and effective deployment can be challenging. Additionally, when procuring AI systems, it is essential to create a level playing field for all stakeholders. This means the procurement process should be inclusive and competitive, giving due consideration to smaller companies such as startups and SMEs. Well-developed public procurement guidelines for AI, like those implemented in the UK (UK Government, 2020) and that are being discussed in the ASEAN region (Access Partnership, 2024), would be invaluable in overcoming capacity constraints within the public sector in Asian countries. These guidelines would help ensure that AI is adopted responsibly, delivering the intended benefits to the public while minimizing potential harm.

To operationalize these principles, and focus areas for implementation, governments can start with pilots, testbeds, and sandboxes to experiment safely, identify risks, and build confidence in AI applications. These approaches foster both effective AI solutions and flexible, iterative policies that allow governments to adapt to new challenges and opportunities in AI, ensuring governance structures evolve with technology. When used responsibly, AI can transform public service delivery, improve decision-making, and tackle pressing societal challenges. To achieve meaningful impact, governments must focus on pragmatic, adoption-oriented strategies that bring abstract ethical principles to life.

Working Draft

References

- Aaronson, S. (2024). *The Age of AI Nationalism and Its Effects*. Centre for International Governance Innovation. <https://www.cigionline.org/publications/the-age-of-ai-nationalism-and-its-effects/>
- Aboitiz Data Innovation. (2024). *Advanced Enterprise GenAI Platform*. <https://aboitizdatainnovation.com/avaai>
- Access Partnership. (2024). *Advancing Responsible AI Public Sector Procurement and Adoption in ASEAN*. Access Partnership. https://cdn.accesspartnership.com/wp-content/uploads/2024/05/Advancing-Responsible-AI-Public-Sector-Procurement-and-Adoption-in-ASEAN_final.pdf?hsCtaTracking=c72d920c-697f-489d-b983-e512263c8d6a%7C4704b85e-6789-4e56-91e2-7be89b805346
- Adams, R., Adeleke, F., Florido, A., Larissa Galdino de Magalhães, S., Grossman, N., Junck, L., & Stone, K. (2024). *Global Index on Responsible AI 2024 (1st Edition)*. Global Center on AI Governance. <https://girai-report-2024-corrected-edition.tiiny.site/>
- Aggarwal, M., & Frayer, J. (2024, June 6). *How AI is changing elections around the world*. NBC News. <https://www.nbcnews.com/news/world/india-ai-changing-elections-world-rcna154838>
- Agrawal, A. (2024, October 13). *Govt mulls setting up Artificial Intelligence Safety Institute*. Hindustan Times. <https://www.hindustantimes.com/india-news/govt-mulls-setting-up-artificial-intelligence-safety-institute-101728833433153.html>
- AI Bias Examples* / IBM. (2024, August 21). <https://www.ibm.com/think/topics/shedding-light-on-ai-bias-with-real-world-examples>
- AI Forum. (2022). *AI in Action: Exploring the Impact of Artificial Intelligence on New Zealand's Productivity*. AI Forum. <https://aiforum.org.nz/ais-impact-on-productivity-in-aotearoa/>
- AI Forum. (2024a). *AI Governance*. AI Forum New Zealand. <https://aigovernance.nz>
- AI Forum. (2024b). *Aotearoa AI Summit*. Aotearoa AI Summit. <https://aotearoaai.nz/>
- AI Forum. (2025). *Harnessing the power of AI to enable a prosperous, inclusive and equitable future Aotearoa*. *AI Forum New Zealand*. <https://aiforum.org.nz/about/>
- AI Singapore. (2022a, May 3). *AI Singapore*. <https://aisingapore.org/>
- AI Singapore. (2024). *Anchoring Deep National Capabilities in AI*. AI Singapore.
- AI Singapore. (2022b, May 3). *Home—AI Singapore*. <https://aisingapore.org/>
- AI Singapore. (2025). *AI Singapore*. <https://aisingapore.org/>

AI Verify Foundation. (2023). *AI Governance Testing Framework and Toolkit*.

AI Verify Foundation. (2023). *Summary Report*.

AI Verify Foundation. (2024). *What is AI Verify*. AI Verify Foundation. <https://aiverifyfoundation.sg/what-is-ai-verify/>

White & Case LLP. (2024, July 1). *AI Watch: Global regulatory tracker - Japan*. <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-japan>

Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603(7903), 864–870. <https://doi.org/10.1038/s41586-022-04484-9>

Alghamdi, J., Lin, Y., & Luo, S. (2024). Fake news detection in low-resource languages: A novel hybrid summarization approach. *Knowledge-Based Systems*, 296, 111884. <https://doi.org/10.1016/j.knosys.2024.111884>

Al Jazeera. (2020, May 13). *Sri Lanka: Facebook apologises for role in 2018 anti-Muslim riots*. <https://www.aljazeera.com/news/2020/5/13/sri-lanka-facebook-apologises-for-role-in-2018-anti-muslim-riots>

Alonso, C., Bhojani, T., & Hanedar, E. (2023). *Stacking Up the Benefits: Lessons from India's Digital Journey*. <https://www.imf.org/en/Publications/WP/Issues/2023/03/31/Stacking-up-the-Benefits-Lessons-from-Indias-Digital-Journey-531692>

Alta Vision. (2024). *Alta Vision | Power the Future*. Alta Vision. <https://altavision.lk/>

Amarasinghe, K., Rodolfa, K., Lamba, H., & Ghani, R. (2023). Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions. *Data & Policy*, 5, e5. <https://doi.org/10.1017/dap.2023.2>

American Psychological Association. (2023, November 29). *How and why does misinformation spread?* <https://www.apa.org/topics/journalism-facts/how-why-misinformation-spreads>

Analytics & Artificial Intelligence Association of the Philippines. (2025). *AAP PREVIOUS EVENTS*. AAP. <https://www.aap.ph/previous-events>

Aneja, U., Gupta, A., Jain, A., & John, S. (2024). *From Code to Consequence: Interrogating Gender Biases in LLMs within the Indian Context*. Digital Futures Lab. <https://www.notion.so>

Aneja, U., John, S., Sen, S., & Jain, A. (2024, March 31). *Responsible AI Impact Lab (RAIL)*

Fellowship. Notion. <https://www.notion.so>

Arayata, C. (2021, April 9). DOST introduces 9 AI R&D projects. *Philippine News Agency*. <https://www.pna.gov.ph/articles/1136226>

Artificial Intelligence Act. (2025). *South Korean AI Basic Law*. Artificial Intelligence Act. <https://artificialintelligenceact.com/south-korean-ai-basic-law/>

ASEAN. (n.d.). *ASEAN Guide on AI Governance and Ethics*. ASEAN. <https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics-beautified-201223-v2.pdf>

Asian Productivity Organization. (2024). *Digital Transformation in Asian Economies: Enhancing Productivity, Socioeconomic Impacts, and Policy Insights*. Asian Productivity Organization. https://www.apo-tokyo.org/wp-content/uploads/2024/07/Digital-Transformation-in-Asian-Economies_PUB.pdf

Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). *Which Humans?* OSF. <https://doi.org/10.31234/osf.io/5b26t>

Ateneo de Manila University. (2023, August 24). *GenAI Task Force (Memo U2324-034) / Memorandum*. Ateneo de Manila University. <http://www.ateneo.edu/document/memorandum/2023/08/24/genai-task-force-memo-u2324-034>

Bajwa, A. (2024). *Nvidia to supply chips to Reliance, other Indian companies in AI push* / Reuters. https://www.reuters.com/technology/artificial-intelligence/nvidia-rolls-out-hindi-language-ai-model-india-ceo-huang-visits-2024-10-24/?utm_source=chatgpt.com

Bandaranayake, R., & Dias, V. (2022). *Towards a Realistic AI Policy for Sri Lanka*. LIRNEasia.

Bañez, E. S. (2024). Notes on pending AI legislation. *UP College of Law*. <https://law.upd.edu.ph/faculty-portfolio/notes-on-pending-ai-legislation/>

Bapat, K. (2022, January 3). *Telangana HC issues notice in challenge to FRT*. Internet Freedom Foundation. <https://internetfreedom.in/telangana-high-court-issues-notice-in-indias-first-legal-challenge-to-the-deployment-of-facial-recognition-technology/>

BBC Bitesize. (n.d.). *What are “bots” and how can they spread fake news?* BBC Bitesize. Retrieved January 28, 2025, from <https://www.bbc.co.uk/bitesize/articles/zjhg47h>

Biswas, S. (2018, August 20). *On the frontline of India’s WhatsApp fake news war*. <https://www.bbc.com/news/world-asia-india-45140158>

Business Standard. (2024a, September 5). *Telangana aims to improve delivery of services to citizens using AI*. *Business Standard India*. https://www.business-standard.com/india-news/telangana-aims-to-improve-delivery-of-services-to-citizens-using-ai-124090500573_1.html

Business Standard. (2024b, December 5). *Google signs MoU with AP govt to accelerate AI adoption and innovation*. https://www.business-standard.com/companies/news/google-signs-mou-with-ap-govt-to-accelerate-ai-adoption-and-innovation-124120500932_1.html

[Business Today. \(2023, November 9\). AI Manipulation Raises Alarms, From Rashmika Mandanna To Katrina Kaif. Business Today. https://www.businesstoday.in/bt-tv/video/ai-manipulation-raises-alarms-from-rashmika-mandanna-to-katrina-kaif-405260-2023-11-09](https://www.businesstoday.in/bt-tv/video/ai-manipulation-raises-alarms-from-rashmika-mandanna-to-katrina-kaif-405260-2023-11-09)

Center for Security and Emerging Technology. (2024, May 2). *Artificial Intelligence Law of the People's Republic of China (Draft for Suggestions from Scholars)*. Center for Security and Emerging Technology. <https://cset.georgetown.edu/publication/china-ai-law-draft/>

Centre for Banking Studies. (2022). *Course Catalogue 2022*.

Centre for Responsible AI. (2024). *Centre for Responsible AI*. Centre for Responsible AI. <https://cerai.iitm.ac.in/>

Chaudhuri, D., Ghosh, P., & Kumar, A. (2024, March 1). *Overcoming digital divide key to achieving Digital India goal*. Down To Earth. <https://www.downtoearth.org.in/science-technology/overcoming-digital-divide-key-to-achieving-digital-india-goal-94758>

Chen, J., Bandy, J., Buckley, D., & Bhatia, R. (2024a). *AI Transparency in practice*.

Cheney, C. (2020, February 25). *To build responsible AI for good, keep humans in the loop*. Devex. <https://www.devex.com/news/sponsored/to-build-responsible-ai-for-good-keep-humans-in-the-loop-96578>

China Law Translate. (2023, July 13). *Interim Measures for the Management of Generative Artificial Intelligence Services*. China Law Translate. <https://www.chinalawtranslate.com/generative-ai-interim/>

Chng, D. G., & Jones, J. (2024). *Global AI Governance Law and Policy: Singapore*. <https://iapp.org/resources/article/global-ai-governance-singapore/>

Chopra, R. (2021, November 18). *Misinformation and Violence*. <https://www.scu.edu/ethics/internet-ethics-blog/misinformation-and-violence/>

Chua, L., & Castillo, B. D. (2024). *A Position on the Initial Consideration of Artificial Intelligence Development and Regulation Act of the Philippines* (SSRN Scholarly Paper No. 4818674). Social Science Research Network. <https://doi.org/10.2139/ssrn.4818674>

Chua, L., Aquino, F. I., Ato, N. B., Borrelli, M., Musch, S., Kerrigan, C., & Heinrich, A. (2023). *Joint Recommendation: Open Comment on University of the Philippines'*

Principles for Responsible AI (SSRN Scholarly Paper No. 4546411). Social Science Research Network. <https://doi.org/10.2139/ssrn.4546411>

CIVICUS. (2024, June 10). *SINGAPORE: 'The entrenched system instils fear, making progress for civil society slow and difficult.'* <https://www.civicus.org/index.php/media-resources/news/interviews/7077-singapore-the-entrenched-system-instils-fear-making-progress-for-civil-society-slow-and-difficult>

Collins, J. (2024, September 11). *Government unlocking potential of AI*. New Zealand Government. <https://www.beehive.govt.nz/release/government-unlocking-potential-ai>

Columbia University School of Professional Studies. (2024, January 2). *The Real Impact of Fake News: The Rise of Political Misinformation—and How We Can Combat Its Influence*. <https://sps.columbia.edu/news/real-impact-fake-news-rise-political-misinformation-and-how-we-can-combat-its-influence>

Committee on Formulating a Strategy for Artificial Intelligence (CFSAI). (2024). *National AI strategy for Sri Lanka*.

Competition Act 2002 (2003).

Council for Science, Technology and Innovation Policy. (2022). *AI Strategy 2022*. <https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf>

Daily FT. (2023, December 13). *1990 Suwa Seriya innovates with Connected Ambulance powered by Mixed Reality and AI in Asia first*. <https://www.ft.lk/front-page/1990-Suwa-Seriya-innovates-with-Connected-Ambulance-powered-by-Mixed-Reality-and-AI-in-Asia-first/44-756275>

Daily Mirror. (2024, October 22). *Nawaloka Hospitals celebrates anniversary with AI-driven healthcare initiatives*. <https://www.dailymirror.lk/business-news/Nawaloka-Hospitals-celebrates-anniversary-with-AI-driven-healthcare-initiatives/273-294220>

Data Guidance. (2024, August 13). Philippines: Bill on Deepfake Accountability and Transparency Act introduced to House. *DataGuidance*. <https://www.dataguidance.com/news/philippines-bill-deepfake-accountability-and>

Data Guidance. (2025, January 2). South Korea: National assembly passes basic law on development of AI | News. *DataGuidance*. <https://www.dataguidance.com/news/south-korea-national-assembly-passes-basic-law>

Data Guidance. (n.d.). Philippines: Bill on Deepfake Accountability and Transparency Act introduced to House | News. *DataGuidance*. Retrieved January 28, 2025, from <https://www.dataguidance.com/news/philippines-bill-deepfake-accountability-and>

Data Protection Authority of Sri Lanka. (2024). *DRAFT 1.0 – Specification of Instruments for Processing Personal Data Outside Sri Lanka Directive*.
<file:///Users/sakalabujung/Downloads/DPA%20-%20Instruments%20for%20Processing%20Data%20Outside%20SL-%20Draft%20Directive%20V%201.0.pdf>

Data Protection Authority of Sri Lanka. (2024). *Draft Personal Data Protection (Classification of Categories of Personal Data Permitted to be Processed in a Third Country by a Public Authority as a Specific Controller) Directives*.
[file:///Users/sakalabujung/Downloads/DPA%20-%20Classification%20of%20Data%20Categories%20Directive%20-%202013-09-2024-%20for%20Public%20Consultations%20\(1\).pdf](file:///Users/sakalabujung/Downloads/DPA%20-%20Classification%20of%20Data%20Categories%20Directive%20-%202013-09-2024-%20for%20Public%20Consultations%20(1).pdf)

data.gov.nz. (n.d.). *Algorithm Impact Assessment toolkit*. New Zealand Government. Retrieved February 2, 2025, from <https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-impact-assessment-toolkit>

Datacom. (2024). *AI Attitudes in New Zealand*.
Datacom. <https://datacom.com/nz/en/solutions/experience/insights/ai-attitudes-research-report>

Deeplina, B., Dutta, S., & Dutta Choudhury, S. (2024, June 5). *In India Election, AI Amplifies Political Reach but Magnifies Disinformation*. Asia Pacific Foundation of Canada. <https://www.asiapacific.ca/publication/indian-election-use-of-ai-political-campaigns-voter-engagement>

Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2023). *The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice* (No. arXiv:2310.00907). arXiv. <https://doi.org/10.48550/arXiv.2310.00907>

Department of Conservation. (2024, July 11). *New AI predator detection tools available*. <https://www.doc.govt.nz/news/media-releases/2024-media-releases/new-ai-predator-detection-tools-available/>

Department of Industry, Science and Resources. (2024, December 17). *Australia and Singapore agree to continue AI cooperation* [News].
<https://www.Industry.Gov.Au/Node/94317>. <https://www.industry.gov.au/news/australia-and-singapore-agree-continue-ai-cooperation>

Department of Information and Communications Technology. (2023, April 14). *DICT Partners with ITU to Deliver “AI Dialogue: Gender-Based AI Policy in the Philippines” Workshop*. <https://ictstatistics.dict.gov.ph/dict-partners-with-itu-to-deliver-ai-dialogue-gender-based-ai-policy-in-the-philippines-workshop/>

Department of Trade and Industry. (2024a). *National Artificial Intelligence Strategy for the Philippines*. Department of Trade and Industry. https://drive.google.com/file/d/1de5kfaGi3tdUgxu1UPV8iRxKpXECTkDp/view?usp=embed_facebook

- Department of Trade and Industry. (2024b, July 2). DTI launches National AI Strategy Roadmap 2.0 and Center for AI Research, positioning the Philippines as a Center of Excellence in AI R&D. *Department of Trade and Industry Philippines*. <https://www.dti.gov.ph/news/dti-launches-national-ai-strategy-roadmap-2-0-center-ai-research-positioning-philippines-center-excellence-ai-rampd/>
- Department of Trade and Industry. (2024c, July 5). DTI drives AI innovation with National Roadmap 2.0 and CAIR launch. *Department of Trade and Industry Philippines*. <https://www.dti.gov.ph/news/dti-drives-ai-innovation-national-roadmap-2-0-cair-launch/>
- Deutsche Welle. (2024, March 26). *Generative AI is the ultimate disinformation amplifier*. Deutsche Welle. <https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890>
- Dharmaraj, S. (2024, September 13). AI to Transform Weather Forecasting in the Philippines. *OpenGov Asia*. <https://opengovasia.com/2024/09/13/ai-to-transform-weather-forecasting-in-the-philippines/>
- Dhillon, A. (2024, February 26). India confronts Google over Gemini AI tool's 'fascist Modi' responses. *The Guardian*. <https://www.theguardian.com/world/2024/feb/26/india-confronts-google-over-gemini-ai-tools-fascist-modi-responses>
- DigiChina. (2022). Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022. *DigiChina*. <https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/>
- Digital Personal Data Protection Act 2023 (2023).
- Digital Watch Observatory. (2024, November 19). *Malaysia explores AI for faster accident detection* /. Digital Watch Observatory. <https://dig.watch/updates/malaysia-explores-ai-for-faster-accident-detection>
- Dunuan, L. F. (2023, September 28). Over 1,000 Women join digital workforce through Connected Women's socially responsible AI skills training and jobs. *Connected Women - Entrepreneurs, Freelancers and Professionals*. <https://www.connectedwomen.co/magazine/over-1000-women-join-digital-workforce-through-connected-womens-socially-responsible-ai-skills-training-and-jobs/>
- Economic Development Board. (2024, June 28). *How Singapore can serve businesses as a hub for AI innovation and growth*. EDB Singapore. <https://www.edb.gov.sg/en/business-insights/insights/how-singapore-can-serve-businesses-as-a-hub-for-ai-innovation-and-growth.html>
- Economist Impact. (2022). *The Inclusive Internet Index*. <https://impact.economist.com/projects/inclusive-internet-index/2022>

- EngageMedia. (2021). *Governance of Artificial Intelligence (AI) in Southeast Asia*. EngageMedia. https://engagemedia.org/wp-content/uploads/2021/12/Engage_Report-Governance-of-Artificial-Intelligence-AI-in-Southeast-Asia_12202021.pdf
- Enriquez, J. M. (2024, October 16). *Tempering the Philippines' AI disinformation storm* | *East Asia Forum*. <https://eastasiaforum.org/2024/10/16/tempering-the-philippines-ai-disinformation-storm/>
- European Foundation for South Asian Studies. (2021). *The role of fake news in fueling hate speech and extremism online; Promoting adequate measures for tackling the phenomenon*. <https://www.efsas.org/publications/study-papers/the-role-of-fake-news-in-fueling-hate-speech-and-extremism-online/>
- Feffer, M., <https://orcid.org/0000-0002-5243-472X>, View Profile, Skirpan, M., <https://orcid.org/0009-0000-4033-1478>, View Profile, Lipton, Z., <https://orcid.org/0000-0002-3824-4241>, View Profile, Heidari, H., <https://orcid.org/0000-0003-3710-4076>, & View Profile. (2023). From Preference Elicitation to Participatory ML: A Critical Survey & Guidelines for Future Research. *Proceedings of the 2023 AAI/ACM Conference on AI, Ethics, and Society*, 38–48. <https://doi.org/10.1145/3600211.3604661>
- Fernando, R. (2024, July 13). 'AI Clubs in schools' launched. *Mirror Education*. <https://edu.dailymirror.lk/>
- Fisher, D. (2023, March 29). Police officers using facial recognition tool without recommended safeguards, prompting 'cavalier' concerns. *New Zealand Herald*. <https://www.nzherald.co.nz/nz/politics/police-officers-using-facial-recognition-tool-without-recommended-safeguards-prompting-cavalier-concerns/PGOYKQ7XJFAO7AULJWGELDB5RA/>
- Freedom House. (2021). *Singapore: Freedom in the World 2021 Country Report*. Freedom House. <https://freedomhouse.org/country/singapore/freedom-world/2021>
- Gilman, M. (2023). *Democratizing AI: Principles for Meaningful Public Participation*. Data & Society. https://datasociety.net/wp-content/uploads/2023/09/DS_Democratizing-AI-Public-Participation-Brief_9.2023.pdf
- Glitch Charity. (2020). *Response to DCMS - The impact of COVID-19 on online abuse and harassment*.
- Global State of Democracy Initiative. (2024). *New Zealand*. International Idea. <https://www.idea.int/democracytracker/country/new-zealand>
- Google Ads. (n.d.). *AI-powered Advertising and Marketing Solutions*. Retrieved January 28, 2025, from <https://ads.google.com/home/campaigns/ai-powered-ad-solutions/>
- Goswami, A. (2022, July 20). Is artificial intelligence the only antidote to disinformation? *World Economic Forum*. <https://www.weforum.org/stories/2022/07/disinformation-ai-technology/>

- Government eMarketplace - Central Public Procurement Portal. (2000). *Information Technology Act, 2000*.
- Government of Tamil Nadu. (2020). *Safe & Ethical Artificial Intelligence Policy*. https://it.tn.gov.in/sites/default/files/2021-06/TN_Safe_Ethical_AI_policy_2020.pdf
- Government Technology Agency of Singapore. (2025a, January 23). *AIBots*. GovTech Singapore. <https://www.tech.gov.sg/aibots/>
- Government Technology Agency of Singapore. (2025b, January 23). *Pair: Transforming ways of working for public officers through AI*. GovTech Singapore. <https://www.tech.gov.sg/products-and-services/for-government-agencies/productivity-and-marketing/pair/>
- Gray, M., Samala, R., Liu, Q., Skiles, D., Xu, J., Tong, W., & Wu, L. (2024). Measurement and Mitigation of Bias in Artificial Intelligence: A Narrative Literature Review for Regulatory Science. *Clinical Pharmacology & Therapeutics*, 115(4), 687–697. <https://doi.org/10.1002/cpt.3117>
- Grover, J. (2025, January 7). Will not rush in bringing Digital India Act: MeitY secretary. *FE Tech Bytes*. <https://www.financialexpress.com/life/technology-will-not-rush-in-bringing-digital-india-act-meity-secretary-3708673/>
- Gutiérrez, J. D. (2024). *Consultation paper on AI regulation: emerging approaches across the world* [Review of *Consultation paper on AI regulation: emerging approaches across the world*]. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000390979>
- Hankins, E., Nettel, P., Martinescu, L., Grau, G., & Rahim, S. (2023). *2023 Government AI Readiness Index*. Oxford Insights. <https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-1.pdf>
- Health New Zealand. (2023). *National AI development checklist*.
- Hindustan Times. (2024, November 11). India's adoption of AI technologies higher than global average, claims new report. *Hindustan Times*. <https://www.hindustantimes.com/business/indias-adoption-of-ai-technologies-higher-than-global-average-claims-new-report-101731321863675.html>
- Hirdaramani, Y. (2024, March 22). *GovInsider dives deep into the tech that's reshaping sustainable development today*. GovInsider. <https://govinsider.asia/intl-en/article/special-reports-dpi>
- Hu, W., & Singh, R. (2024). *Enrolling Citizens: A Primer on Archetypes of Democratic Engagement with AI*. Data & Society. https://datasociety.net/wp-content/uploads/2024/06/DS_Enrolling-Citizens-Primer_FINAL.pdf
- IAPP Research and Insights. (n.d.). *Global AI Law and Policy Tracker*. IAPP Research and Insights. https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf

IBM Data and AI Team. (2023, October 16). Shedding light on AI bias with real world examples. *IBM*. <https://www.ibm.com/think/topics/shedding-light-on-ai-bias-with-real-world-examples>

Implementing Rules and Regulations of the Data Privacy Act of 2012 (2016). <https://privacy.gov.ph/implementing-rules-regulations-data-privacy-act-2012/>

Inamdar, N. (2024, May 1). India's economy: The good, bad and ugly in six charts. *BBC News*. <https://www.bbc.com/news/world-asia-india-68823827>

IndiaAI. (2023). *ICMR releases ethical guidelines for AI in biomedical research and healthcare*. IndiaAI. <https://indiaai.gov.in/news/icmr-releases-ethical-guidelines-for-ai-in-biomedical-research-and-healthcare>

[IndiaAI. \(2025\). *INDIAai / Pillars*. IndiaAI. https://indiaai.gov.in/](https://indiaai.gov.in/)

Infocomm Media Development Authority. (n.d.). *Annex: Nine dimensions of the proposed Model AI Governance Framework*. Infocomm Media Development Authority. <https://www.imda.gov.sg/-/media/imda/files/news-and-events/media-room/media-releases/2024/01/public-consult-model-ai-governance-framework-genai/annex-nine-dimensions-of-the-proposed-model-ai-governance-framework.pdf>

International Association of Privacy Professionals. (2024). *Global AI Law and Policy Tracker*.

International Trade Administration. (2024a, January 23). *Philippines—Information and Communications Technology*. <https://www.trade.gov/country-commercial-guides/philippines-information-and-communications-technology>

International Trade Administration. (2024b, September 18). *New Zealand—Digital Economy*. <https://www.trade.gov/country-commercial-guides/new-zealand-digital-economy>

[Internet Freedom Foundation. \(2020\). *Project Panoptic*. Panoptic. https://panoptic.in](https://panoptic.in)

Ipac, J. C. (2024). AI and the law in the Philippines. *Asia Business Law Journal*. <https://law.asia/ai-law-philippines/>

Jain, A. (2022, January 29). *IFF explains DigiYatra: Turbulence ahead*. Internet Freedom Foundation. <https://internetfreedom.in/digiyaatra-part-2/>

James Cook University. (2024, March 12). *The need to look deeper into responsible use of artificial intelligence (AI) in higher education* [Text]. <https://www.jcu.edu.sg/news/releases/the-need-to-look-deeper-into-responsible-use-of-artificial-intelligence-ai-in-higher-education>

Jayswal, P. (2021, August 17). How the IT Industry is shaping the future of India? *The Times of India*. <https://timesofindia.indiatimes.com/readersblog/youth2020/how-the-it-industry-is-shaping-the-future-of-india-36519/>

- Jindal, S. (2023, December 19). *India's LLM Battle: Ola's Krutrim vs Zoho*. Analytics India Mag. <https://analyticsindiamag.com/ai-breakthroughs/indias-llm-battle-olas-krutrim-vs-zoho/>
- Kamali, N., Nakamura, K., Chatzimpampas, A., Hullman, J., & Groh, M. (2024). *How to Distinguish AI-Generated Images from Authentic Photographs* (No. arXiv:2406.08651). arXiv. <https://doi.org/10.48550/arXiv.2406.08651>
- Kapoor, S., & Narayanan, A. (2024, December 13). We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem. *Knight First Amendment Institute*. <http://knightcolumbia.org/blog/we-looked-at-78-election-deepfakes-political-misinformation-is-not-an-ai-problem>
- Kapoor, S., & Narayanan, A. (2024, December 14). We Looked at 78 Election Deepfakes. Political Misinformation is not an AI Problem. *AI Snake Oil*. <https://www.aisnakeoil.com/p/we-looked-at-78-election-deepfakes>
- King, J., & Meinhardt, C. (2024). *Rethinking Privacy in the AI Era Policy Provocations for a Data-Centric World*. Stanford University Human-Centred Artificial Intelligence. <https://hai.stanford.edu/sites/default/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf>
- Knowles, C. (2023, August 25). 78% of NZ small business owners worried about AI adoption. *IT Brief New Zealand*. <https://itbrief.co.nz/story/78-of-nz-small-business-owners-worried-about-ai-adoption>
- Krishan, N. (2024, July 22). Google, Microsoft, Infosys, IIM B and Indian AI startups form new major coalition for Responsible AI in India. *The Hindu*. <https://www.thehindu.com/sci-tech/technology/google-microsoft-infosys-iim-b-and-indian-ai-startups-form-new-major-coalition-for-responsible-ai-in-india/article68431246.ece>
- Kumar, D. (2024, August 28). *AI has had a free run. Now, India is weighing standards*. Mint. <https://www.livemint.com/ai/india-to-regulate-ai-ai-based-applications-bis-preparing-comprehensive-guidelines-for-ai-based-platforms-11724742116250.html>
- Kumar, R. (2024, October 8). *Artificial Intelligence Transforming Singapore's Industries Today*. Mobiloitte. <https://www.mobiloitte.com/blogs/ai-adoption-in-singapore/>
- Lai, B. (2024, August 28). Singapore Strengthens Her Commitment to Responsible AI with the Release of AI Ethics and Governance Body of Knowledge Version 2.0. *Privacy World*. <https://www.privacyworld.blog/2024/08/singapore-strengthens-her-commitment-to-responsible-ai-with-the-release-of-ai-ethics-and-governance-body-of-knowledge-version-2-0/>
- Letzing, J. (2024, November 13). What is 'sovereign AI' and why is it so appealing? *World Economic Forum*. <https://www.weforum.org/stories/2024/11/what-is-sovereign-ai-and-why-is-the-concept-so-appealing-and-fraught/>

- Levush, R. (2019). *Government Responses to Disinformation on Social Media Platforms: Argentina, Australia, Canada, China, Denmark, Egypt, European Union, France, Germany, India, Israel, Mexico, Russian Federation, Sweden, United Arab Emirates, United Kingdom*. <https://digitalcommons.unl.edu/scholcom/178>
- LIRNEasia. (2023, November 30). Introducing Dissect: An AI-empowered fact-checking tool developed by Appendix (Watchdog Sri Lanka). *LIRNEasia*. <https://lirneasia.net/2023/11/lirneasia-introduces-dissect-a-fact-checking-tool-empowered-by-ai/>
- Liu, S. (2023). *India's AI Regulation Dilemma*. <https://thediplomat.com/2023/10/indias-ai-regulation-dilemma/>
- Loi, M., & Spielkamp, M. (2021). Towards Accountability in the Use of Artificial Intelligence for Public Administrations. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 757–766. <https://doi.org/10.1145/3461702.3462631>
- Lubarsky, B. (2017). *Re-Identification of "Anonymized" Data*. <https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/>
- Morrison Foerster. (2023, January 30). *New Wave of Privacy Laws in the APAC Region*. Morrison Foerster. <https://www.mofo.com/resources/insights/null>
- Matika, C. (2023). *New Zealand's Internet Insights*.
- MAUI63. (2022). *Saving the worlds rarest dolphin with technology*. MAUI63. <https://www.maui63.org>
- MeitY. (2023a). *IndiaAI 2023: Expert Groups*. <https://www.meity.gov.in/content/indiaai-2023-expert-group-report-%E2%80%93-first-edition-the-ministry-electronics-and-information>
- MeitY. (2023b). *Proposed Digital India Act, 2023*. https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf
- MeitY. (2024). *Due Diligence by Intermediaries/Platforms under the IT Act, 2000*. <https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf>
- MeitY. (2025). *Report on AI governance guidelines development for Public consultation*. <https://www.meity.gov.in/content/report-ai-governance-guidelines-development-public-consultation>
- Meta. (2025). *Meta Advantage: Better ad performance with ai-powered tools*. Meta for Business. <https://en-gb.facebook.com/business/ads/meta-advantage>
- Metreau, E., Young, K. E., & Eapen, S. G. (2024, July 1). World Bank country classifications by income level for 2024-2025. *World Bank*

Blogs. <https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025>

Ministry for the Digital Economy & Communications. (2023). *Digital Technologies Industry Transformation Plan*.

Ministry of Business Innovation & Employment. (2024). *Approach to work on Artificial Intelligence*. Ministry of Business Innovation & Employment.

Ministry of Corporate Affairs. (2002). *Competition Act 2002*.

Ministry of Digital Development and Information. (2024). *Smart Nation 2.0: A Thriving Digital Future for All*. Ministry of Digital Development and Information.

Ministry of Digital Economy. (2025). *Data Protection Authority*. Ministry of Digital Economy. <https://mode.gov.lk/docs/institutions/DPA>

Ministry of Digital Economy. (2025, January 10). *Shaping Sri Lanka's AI Future— Consultation of the National Strategy at Water's Edge*. <https://mode.gov.lk/blog/ai-strategy-launch-program>

Ministry of Education. (2024, November 25). *Generative AI*. <https://www.education.govt.nz/school/digital-technology/generative-ai>

Ministry of Electronics & Information Technology. (2023). *Digital Personal Data Protection Act 2023*.

Ministry of Electronics and Information Technology, Government of India. (2022a, March 30). Artificial Intelligence. *Press Information Bureau*. <https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1811372>

Ministry of Electronics and Information Technology, Government of India. (2022b, May 24). MeitY holds Brainstorming with Researchers & Startups to shape strategy for Digital India BHASHINI - India's AI led language translation platform. *Press Information Bureau*. <https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1827997>

Ministry of Electronics and Information Technology, Government of India. (2022c, July 27). National Data Governance Framework Policy. *Press Information Bureau*. <https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1845318>

Ministry of Foreign Affairs and Trade. (2025). *DEPA text and resources*. New Zealand Ministry of Foreign Affairs and Trade. <https://www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-in-force/digital-economy-partnership-agreement-depa/depa-text-and-resources>

Ministry of Health, Health Sciences Authority, & Integrated Health Information Systems. (2021). *Artificial Intelligence in Healthcare Guidelines*.

Ministry of Technology. (2024). *National Digital Economy Strategy 2030*.

- Mishra, U. (2024, January 31). *AI adoption rate high among educators with 60% already using AI tools*. Fortune India. <https://www.fortuneindia.com/macro/ai-adoption-rate-high-among-educators-with-60-already-using-ai-tools/115535>
- MIT Media Lab. (n.d.). *Detect DeepFakes: How to counteract misinformation created by AI*. MIT Media Lab. Retrieved January 28, 2025, from <https://www.media.mit.edu/projects/detect-fakes/overview/>
- Mitra, B., Cramer, H., & Gurevich, O. (2024). *Sociotechnical Implications of Generative Artificial Intelligence for Information Access* (arXiv:2405.11612). arXiv. <https://doi.org/10.48550/arXiv.2405.11612>
- Mobiloitte. (2024, October 12). *AI Adoption in Singapore: A Detailed Analysis*. Mobiloitte. <https://mobiloitte.com/blog/ai-adoption-in-singapore/>
- Mohanty, A., & Sahu, S. (2024, November 21). *India's Advance on AI Regulation*. Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en>
- Monetary Authority of Singapore. (2018, November 12). *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*. <https://www.mas.gov.sg/~media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf>
- Mohanty, A., & Sahu, S. (2024). *India's Advance on AI Regulation*. Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en>
- Monetary Authority of Singapore. (2019, November 13). *MAS Partners Financial Industry to Create Framework for Responsible Use of AI*. <https://www.mas.gov.sg/news/media-releases/2019/mas-partners-financial-industry-to-create-framework-for-responsible-use-of-ai>
- Monetary Authority of Singapore. (2024, July 18). *MAS Commits up to S\$100 Million to Support Quantum and Artificial Intelligence Capabilities in the Financial Sector*. [https://www.mas.gov.sg/news/media-releases/2024/mas-commits-up-to-s\\$100-million-to-support-quantum-and-artificial-intelligence-capabilities](https://www.mas.gov.sg/news/media-releases/2024/mas-commits-up-to-s$100-million-to-support-quantum-and-artificial-intelligence-capabilities)
- Moran, W., Vowles, J., Blyth, C. A., Dalziel, R., Oliver, W. H., & Sinclair, K. (2025). *Economy of New Zealand*. In *Britannica*. <https://www.britannica.com/place/New-Zealand/Economy>
- Nagar, S., & Eaves, D. (2024). *Interactions Between Artificial Intelligence and Digital Public Infrastructure: Concepts, Benefits, and Challenges* (No. arXiv:2412.05761). arXiv. <https://doi.org/10.48550/arXiv.2412.05761>
- Nandi, A., & Yadav, S. (2024). *Digital Dreams, Real Challenges: Key Factors Driving*

India's AI Ecosystem. Observer Research Foundation.
https://www.orfonline.org/research/digital-dreams-real-challenges-key-factors-driving-indias-ai-ecosystem#_edn84

NASSCOM. (2024). *The Developer's Playbook for Responsible AI in India*. NASSCOM.
<https://nasscom.in/ai/pdf/the-developer's-playbook-for-responsible-ai-in-india.pdf>

NASSCOM. (n.d.). *NASSCOM Responsible AI Resource Kit*. NASSCOM. Retrieved 17 December 2024, from <https://indiaai.gov.in/responsible-ai/homepage>

Natarajan, S., Mishra, K., Mohamed, S., & Taylor, A. (2021). *Just and equitable data labelling towards a responsible AI supply chain*. Aapti Institute.
<https://aapti.in/blog/just-and-equitable-data-labelling/>

National Citizens' Movement for Free Elections (NAMFREL). (2024). *NAMFREL Position Paper on the Use of Artificial Intelligence In Elections and Election-Related Activities*. https://drive.google.com/file/d/1m8n_QqVrJu97OO5A5VcMHBXWuJ7QsrqI/view?usp=embed_facebook

National Privacy Commission. (2024, December 19). *GUIDELINES ON THE APPLICATION OF REPUBLIC ACT NO. 10173 OR THE DATA PRIVACY ACT OF 2012 (DPA), ITS IMPLEMENTING RULES AND REGULATIONS, AND THE ISSUANCES OF THE COMMISSION TO ARTIFICIAL INTELLIGENCE SYSTEMS PROCESSING PERSONAL DATA*. National Privacy Commission.
<https://privacy.gov.ph/wp-content/uploads/2024/12/Advisory-2024.12.19-Guidelines-on-Artificial-Intelligence-w-SGD.pdf>

National University of Singapore. (2025a). *AI Governance and Policy*. <https://ai.nus.edu.sg/ai-governance-and-policy/>

National University of Singapore. (2025b). *NUS Artificial Intelligence Institute*. <https://ai.nus.edu.sg/>

National University of Singapore. (2025c). *Projects*. <https://ai.nus.edu.sg/projects/>

Nayak, R. (2024, July 18). Japan's AI regulations: Agile governance in action. *Diligent*. <https://www.diligent.com/resources/blog/japan-ai-regulations>

New Zealand Customs Service. (2024, November 12). *eGate*. New Zealand Customs Service. <https://www.customs.govt.nz/personal/travel-to-and-from-nz/travelling-to-new-zealand/egate/>

New Zealand Government. (2020). *Algorithm charter for Aotearoa New Zealand*.

New Zealand Government. (2022, September 14). *The Digital Strategy for Aotearoa*. New Zealand Digital Government. <https://www.digital.govt.nz>

New Zealand Government. (2023a). *Full results: 2023 cross-agency survey for artificial intelligence (AI)*. New Zealand Digital Government. <https://www.digital.govt.nz>

- New Zealand Government. (2023b, July 20). *Data Ethics Advisory Group*. New Zealand Digital Government. <https://data.govt.nz/leadership/advisory-governance/data-ethics-advisory-group>
- New Zealand Government. (2023c, July 26). *Understanding the benefits of GenAI to the public service*. New Zealand Digital Government. <https://www.digital.govt.nz/standards-and-guidance/technology-and-architecture/artificial-intelligence/interim-generative-ai-guidance-for-the-public-service/understanding-the-benefits-of-genai-to-the-public-service>
- New Zealand Government. (2024a). *Full results: 2024 cross-agency survey for artificial intelligence (AI) use cases*. New Zealand Digital Government. <https://www.digital.govt.nz>
- New Zealand Government. (2024b, September 19). *Government Chief Digital Officer's role in artificial intelligence (AI)*. New Zealand Digital Government. <https://www.digital.govt.nz/standards-and-guidance/technology-and-architecture/artificial-intelligence/gcdo-role>
- New Zealand Government. (2023). *Initial advice on Generative Artificial Intelligence in the public service*.
- New Zealand Ministry of Foreign Affairs and Trade. (2025). *DEPA text and resources*. New Zealand Ministry of Foreign Affairs and Trade. <https://www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-in-force/digital-economy-partnership-agreement-depa/depa-text-and-resources>
- New Zealand Ministry of Foreign Affairs and Trade. (2019). *Digital Economy Partnership Agreement*.
- Ng, M., Khoo, M., Haridas, G., & Toh, J. T. (2023). *The Economic Impact of Generative AI: The Future of Work in the Philippines*.
- Niti Aayog. (2018, June). *National Strategy for Artificial Intelligence, 2018*. Niti Aayog. <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf>
- NITI Aayog. (2021, February). *Approach Document for India Part 1 – Principles for Responsible AI*. NITI Aayog. <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>
- NZTech. (2022). *NZTech*. NZTech. <https://nztech.org.nz/>
- OECD.AI. (n.d.). *Accountability (Principle 1.5)*. OECD.AI. Retrieved February 2, 2025, from <https://oecd.ai/en/dashboards/ai-principles/P9>
- OECD. (n.d.). *AI principles*. OECD. Retrieved February 2, 2025, from <https://www.oecd.org/en/topics/ai-principles.html>

- Office of the Privacy Commissioner. (2023, May 25). *Privacy Commissioner outlines expectations around AI use*. <https://privacy.org.nz/news/media-releases/privacy-commissioner-outlines-expectations-around-ai-use/>
- Office of the Privacy Commissioner. (2023). *Artificial Intelligence and the Information Privacy Principles*.
- Organisation for Economic Co-operation and Development. (2025). *Ability of adults to identify online disinformation created by generative AI*. OECD Going Digital Toolkit. <https://goingdigital.oecd.org/en/indicator/81>
- Osmundsen, M., Bor, A., & Petersen, M. B. (2021, May 13). *How partisan polarization drives the spread of fake news*. Brookings. <https://www.brookings.edu/articles/how-partisan-polarization-drives-the-spread-of-fake-news/>
- Our World in Data. (2023a). *Civil society participation index*. Our World in Data. <https://ourworldindata.org/grapher/civil-society-participation-index?tab=table>
- Our World in Data. (2023b). *Civil society participation index*. Our World in Data. <https://ourworldindata.org/grapher/civil-society-participation-index?tab=table>
- Our World in Data. (2023c). *Democracy Index*. Our World in Data. <https://www.eiu.com/n/campaigns/democracy-index-2023/>
- Our World in Data. (2023d). *Democracy Index*. Our World in Data. <https://ourworldindata.org/grapher/democracy-index-eiu>
- Our World in Data. (2023e). *Democracy Index*. Our World in Data. <https://ourworldindata.org/grapher/democracy-index-eiu>
- Our World in Data. (2023f). *Democracy Index*. Our World in Data. <https://ourworldindata.org/grapher/democracy-index-eiu>
- Our World in Data. (2023g). *Freedom of expression index*. Our World in Data. <https://ourworldindata.org/grapher/freedom-of-expression-index?tab=table>
- Our World in Data. (2023h). *Freedom of expression index*. Our World in Data. <https://ourworldindata.org/grapher/freedom-of-expression-index?tab=table>
- Outgrow. (n.d.). *Agri Services—Outgrow*. WayCool. Retrieved February 2, 2025, from <https://waycool.in/>
- Pagallo, U. (2023). *Why the AI Act Won't Trigger a Brussels Effect* (SSRN Scholarly Paper No. 4696148). Social Science Research Network. <https://doi.org/10.2139/ssrn.4696148>
- Pal, M. (2024, August 12). *Here's How AI Is Revolutionizing India's Banking Sector; Driving Customer Experience And Operational Efficiency*. Times Now. <https://www.timesnownews.com/technology-science/heres-how-ai-is-revolutionizing->

indias-banking-sector-driving-customer-experience-and-operational-efficiency-article-112472156

- Parasnis, S. (2024). *India AI Mission picks 8 projects to boost ethical AI development*. Medianama. <https://www.medianama.com/2024/10/223-indiaai-mission-selects-eight-proposals-responsible-ai-development/>
- Paulger, D., Mohamed, B., Shivhare, S., & Tan, B. (2024). *Regulatory Strategies of Data Protection Authorities in the Asia-Pac Region: 2024, and Beyond*. Future of Privacy Forum. <https://fpf.org/wp-content/uploads/2024/09/FPF-Issue-Brief-APAC-DPA-Strategies-2024-1.pdf>
- Pennington, P. (2024, August 20). NZ Police finally has facial recognition policy—But is it strict enough? *RNZ*. <https://www.rnz.co.nz/news/media-technology/525642/nz-police-finally-has-facial-recognition-policy-but-is-it-strict-enough>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>
- Personal Data Protection Commission Singapore. (2024, March 1). *ADVISORY GUIDELINES ON USE OF PERSONAL DATA IN AI RECOMMENDATION AND DECISION SYSTEMS*. Personal Data Protection Commission Singapore. <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/advisory-guidelines-on-the-use-of-personal-data-in-ai-recommendation-and-decision-systems.pdf>
- Philippine Institute for Development Studies. (2024, January 29). *Beyond remittances: PIDS study calls for holistic policymaking for OFWs*. <http://pids.gov.ph/details/news/press-releases/beyond-remittances-pids-study-calls-for-holistic-policymaking-for-ofws>
- Porcalla, D. (2024, August 12). Bill filed to regulate artificial intelligence use in 2025 polls. *The Philippine Star*. <https://www.philstar.com/headlines/2024/08/12/2377229/bill-filed-regulate-artificial-intelligence-use-2025-polls>
- Press Information Bureau. (2024). *UNESCO and the Ministry of Electronics and Information Technology, Host Multi-Stakeholder Consultation on Safety and Ethics in Artificial Intelligence (AI)*. <https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=2073920>
- Press Insider. (2024). India allocates \$1.25 billion for AI mission. *PRESS Insider | India's Global Voice*. <https://pressinsider.com/technology/india-allocates-1-25-billion-for-ai-mission/>
- qBotica. (2023, September 22). *The Roadmap to AI Adoption in Government: Challenges and*

Solutions. qBotica. <https://qbotica.com/the-roadmap-to-ai-adoption-in-government-challenges-and-solutions/>

- Radio New Zealand. (2024, February 8). Privacy commissioner keeps close eye on supermarkets' facial recognition trial. *RNZ*. <https://www.rnz.co.nz/news/national/508613/privacy-commissioner-keeps-close-eye-on-supermarkets-facial-recognition-trial>
- Ramotowski, S. (2025, January 15). Future Of BPO Industry in The Philippines 2024. *Nexford University*. <https://www.nexford.edu/the-future-of-bpos-in-the-philippines-and-growth-opportunities>
- Remulla, M. M., & Medina, G. M. (2012). *Measuring the Contribution to the Philippine Economy of Information Technology- Business Process Outsourcing (IT-BPO) Services*.
- Ricciardone, S. M. (2024, July 24). *How AI bots spread misinformation online and undermine democratic politics*. *The Conversation*. <http://theconversation.com/how-ai-bots-spread-misinformation-online-and-undermine-democratic-politics-234915>
- Romeo, J. (2022, October 11). Disinformation is a growing crisis. Governments, business and individuals can help stem the tide. *World Economic Forum*. <https://www.weforum.org/stories/2022/10/how-to-address-disinformation/>
- Ryburn, A., Stiles, R., Steel-Baker, D., Chapman, A., Dysart, A., & Rose, E. (2024). Regulatory update: Cabinet paper on strategic approach to work on artificial intelligence. *Lexology*. <https://www.lexology.com/library/detail.aspx?g=29abc3db-606f-413a-9cb1-c0ebba46c5fa>
- Sanchez, C. (2021, July 8). *Civil society can help ensure AI benefits us all. Here's how*. *World Economic Forum*. <https://www.weforum.org/stories/2021/07/civil-society-help-ai-benefits/>
- Sarkar, A., & Singh, P. (n.d.). *Progress of Healthcare Artificial Intelligence in India*. Retrieved February 2, 2025, from <https://www.isb.edu/en/research-thought-leadership/research-centres-institutes/isb-institute-of-data-science/Opinion-Pieces/Progress-of-Healthcare-Artificial-Intelligence-in-India.html>
- SCAI. (2023). *About SCAI*. <https://www.scai.gov.sg/about/>
- Sebastian, M. (2024, May 16). AI and deepfakes blur reality in India elections. *BBC News*. <https://www.bbc.com/news/world-asia-india-68918330>
- Sharon, A. (2024a, November 8). Singapore and UK: Advancing Responsible, Secure AI. *OpenGov Asia*. <https://opengovasia.com/2024/11/08/singapore-and-uk-advancing-responsible-secure-ai/>
- Sharon, A. (2024b, November 16). New Zealand: AI and GenAI Transforming Citizen Services. *OpenGov Asia*. <https://opengovasia.com/2024/11/16/new-zealand-ai-and-genai-transforming-citizen-services/>

- Shekar, K., Sahiba, J., & Birla, B. (2023). *Principles for Enabling Responsible AI Innovations in India: An Ecosystem Approach*. The Dialogue. <https://thedialogue.co/publication/principles-for-enabling-responsible-ai-innovations-in-india-an-ecosystem-approach/>
- Sia. (2024, July 29). *The Artificial Intelligence (AI) regulatory landscape in APAC*. <https://www.sia-partners.com/en/insights/publications/artificial-intelligence-ai-regulatory-landscape-apac>
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-127>
- Singapore Academy of Law. (2020). *Applying Ethical Principles for Artificial Intelligence in Regulatory Reform*. https://www.sal.org.sg/Resources-Tools/Law-Reform/AI_Ethical_Principles
- Singapore Computer Society. (2025). *About Us—Singapore Computer Society*. <https://www.scs.org.sg/about>
- Singapore Government Developer Portal. (2023, September 12). *Data & AI Literacy ePrimer*. Singapore Government Developer Portal. <https://www.developer.tech.gov.sg/products/collections/data-science-and-artificial-intelligence/data-ai-literacy-eprimer.html>
- Singapore Institute of Technology. (2025, September 1). *Applied Artificial Intelligence | Singapore Institute of Technology*. <https://www.singaporetech.edu.sg/undergraduate-programmes/applied-artificial-intelligence>
- Singh, K. R., & Rekhi, D. (2024, September 3). Governments say ‘aye’ to sovereign AI. *Economic Times*. <https://economictimes.indiatimes.com/tech/artificial-intelligence/governments-say-aye-to-sovereign-ai/articleshow/113014070.cms?from=mdr>
- Singha, K. (2024, September 10). *Ministry of Electronics and Information Technology launches Phase III of capacity-building initiatives under Digital India vision*. Ministry of Electronics & IT. <https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=2053495>
- SLIIT. (2023). *BSc (Hons) in Information Technology—Artificial Intelligence*. SLIIT. <https://www.sliit.lk/computing/programmes/artificial-intelligence-degree/>
- Smart Nation and Digital Government Office. (2025a, January 10). *National AI Strategy*. Smart Nation Singapore. <https://www.smartnation.gov.sg/nais/>
- Smart Nation and Digital Government Office. (2025b, January 10). *Our Smart Nation Vision in 2014*. Smart Nation Singapore. <https://www.smartnation.gov.sg/sn1/>
- Smart Nation and Digital Government Office. (2023). *Singapore National AI Strategy 2.0*. Smart Nation and Digital Government Office.

- Sperling, J. (2024, August 5). *AI-Driven Misinformation: Challenges and Solutions for Businesses*. Columbia Business School. <https://business.columbia.edu/insights/digital-future/ai-driven-misinformation-challenges-and-solutions-businesses>
- Squire Patton Boggs LLP. (n.d.). *Overview of Privacy & Data Protection Laws: Asia-Pacific*. Privacy World. Retrieved February 2, 2025, from <https://www.privacyworld.blog/privacy-asia-pacific/>
- Stanly, M. (2023, December 20). *India turns to AI to capture its 121 languages for digital services*. IndiaAI. <https://indiaai.gov.in/article/india-turns-to-ai-to-capture-its-121-languages-for-digital-services>
- Strick, B., & Syavira, F. (2019, October 11). *Papua unrest: Social media bots “skewing the narrative.”* <https://www.bbc.com/news/world-asia-49983667>
- Sur, A. (2024, July 1). *Digital India Bill likely to be delayed, government may opt for smaller, urgent regulations*. Moneycontrol. <https://www.moneycontrol.com/technology/digital-india-bill-likely-to-be-delayed-government-may-opt-for-smaller-urgent-regulations-article-12759435.html>
- Swenson, A., & Chan, K. (2024, March 14). *Election disinformation takes a big leap with AI being used to deceive worldwide*. AP News. <https://apnews.com/article/artificial-intelligence-elections-disinformation-chatgpt-bc283e7426402f0b4baa7df280a4c3fd>
- Tan, S. Y., & Taeihagh, A. (2019). *Adaptive and experimental governance in the implementation of autonomous vehicles: The case of Singapore*.
- Telecom Regulatory Authority of India. (2023). *Recommendations on Leveraging Artificial Intelligence and Big Data in Telecommunication Sector*. https://www.trai.gov.in/sites/default/files/Recommendation_20072023_0.pdf
- Thavapalakumar, M. (2024). Sri Lanka’s banking evolution. *AB Magazine*, November 2024. <https://abmagazine.accaglobal.com/content/abmagazine/global/articles/2024/nov/comment/sri-lanka-s-banking-evolution.html>
- The Ambit Coalition - Philippines Council. (2023, May 12). *Public Call on Ethics, Safety, and Governance of AI in the Philippines*. Change.Org. <https://www.change.org/p/public-call-on-ethics-safety-and-governance-of-ai-in-the-philippines>
- The Association of Southeast Asian Nations. (2024). *ASEAN Guide on AI Governance and Ethics*.
- The Christchurch Call. (2022, September 21). Christchurch Call Initiative on Algorithmic Outcomes. *The Christchurch Call*. <https://www.christchurchcall.org/christchurch-call-initiative-on-algorithmic-outcomes/>

- The Digital Watch. (2018). *India's National Strategy for Artificial Intelligence* | Digital Watch Observatory. <https://dig.watch/resource/indias-national-strategy-for-artificial-intelligence>
- The Economist. (2024, May 1). Disinformation is on the rise. How does it work? *The Economist*. <https://www.economist.com/science-and-technology/2024/05/01/disinformation-is-on-the-rise-how-does-it-work>
- The Financial Times. (2025). *AI ascent: Inside Singapore's rise as a global hub for artificial intelligence*. <https://stb.ft.com/article/inside-singapores-rise-global-hub-artificial-intelligence>
- The Global Partnership on Artificial Intelligence. (2023). *Working Group on Responsible AI*. <https://gpai.ai/projects/responsible-ai/>
- The Global Partnership on Artificial Intelligence. (2025). *Global Partnership on Artificial Intelligence—GPAI*. <https://gpai.ai/>
- The Government of the Republic of Korea. (2019). *Korea National Strategy for Artificial Intelligence*.
- The Lawphil Project. (2012). *Data Privacy Act*.
- The Lawphil Project. (2022, July 30). *Anti-Online Sexual Abuse or Exploitation of Children and Anti-Child Sexual Abuse or Exploitation Material Act*. https://lawphil.net/statutes/repacts/ra2022/ra_11930_2022.html
- The Lawphil Project. (2022, July 30). *Philippine Digital Workforce Competitiveness Act*. https://lawphil.net/statutes/repacts/ra2022/ra_11927_2022.html
- The Parliament of Sri Lanka. (2022). *Personal Data Protection Act, No. 9 of 2022*.
- The World Bank. (2022, November 24). *PHILIPPINES: Reducing Inequality Key to Becoming a Middle-Class Society Free of Poverty* [Text/HTML]. World Bank Group. <https://www.worldbank.org/en/news/press-release/2022/11/24/ph-reducing-inequality-key-to-becoming-a-middle-class-society-free-of-poverty>
- The World Bank. (2024a). *The World Bank in Singapore* [Text/HTML]. World Bank Group. <https://www.worldbank.org/en/country/singapore/overview>
- The World Bank. (2024b, October 16). *The World Bank in the Philippines* [Text/HTML]. World Bank Group. <https://www.worldbank.org/en/country/philippines/overview>
- The World Factbook. (2025a). New Zealand. In *The World Factbook*. Central Intelligence Agency. <https://www.cia.gov/the-world-factbook/countries/new-zealand/#people-and-society>

- The World Factbook. (2025b). Philippines. In *The World Factbook*. Central Intelligence Agency. <https://www.cia.gov/the-world-factbook/countries/philippines/#economy>
- The World Factbook. (2025c). Singapore. In *The World Factbook*. Central Intelligence Agency. <https://www.cia.gov/the-world-factbook/countries/singapore/>
- Theodorou, Y., Cowie, C., & Charlwood, R. (2024, July 25). *Creating Value for Users and Governments: How AI Can Enhance Digital-ID Solutions*. Tony Blair Institute for Global Change. <https://institute.global/insights/tech-and-digitalisation/creating-value-for-users-and-governments-how-ai-can-enhance-digital-id>
- Times of India. (2024, March 12). This new AI tool can help you book train tickets, get refunds and check details on IRCTC website and app. *The Times of India*. <https://timesofindia.indiatimes.com/gadgets-news/this-new-ai-tool-can-help-you-book-train-tickets-get-refunds-and-check-details-on-irctc-website-and-app/articleshow/108431623.cms>
- Treasury Board of Canada Secretariat. (2024, May 30). *Algorithmic Impact Assessment tool*. Government of Canada. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>
- Trinchera, G. T. (2024, July 2). Innovation and the convergence of AI and ESG in the Philippines. *KPMG*. <https://kpmg.com/ph/en/home/insights/2024/07/innovation-and-the-convergence-of-ai-and-esg-in-the-philippines.html>
- Tutella, F. (2024, January 8). *Slow the scroll: Users less vigilant about misinformation on mobile phones*. <https://www.psu.edu/news/research/story/slow-scroll-users-less-vigilant-about-misinformation-mobile-phones>
- UK Government. (2020, June 8). *Guidelines for AI procurement*. GOV.UK. <https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement>
- UK Government. (2023). *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*. GOV.UK. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>
- UNESCO. (n.d.). *Ethics of Artificial Intelligence*. UNESCO. Retrieved February 2, 2025, from <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence* (p. 21). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
- United Nations Sri Lanka. (2024, April 7). *Safeguarding democracy by combatting hate speech and disinformation*. <https://srilanka.un.org/en/265414-safeguarding-democracy-combatting-hate-speech-and-disinformation>

<https://srilanka.un.org/en/265414-safeguarding-democracy-combatting-hate-speech-and-disinformation>

- Universities New Zealand - Te Pōkai Tara. (2025). *Driving research and innovation*. <https://www.universitiesnz.ac.nz/sector-research/driving-research-and-innovation>
- University of Auckland. (2023, October 25). *Call for framework to evaluate new AI tech*. Koi Tū: Technology Futures. <https://informedfutures.org/call-for-ai-framework/>
- University of Auckland. (2025a). *About – Koi Tū: The Centre for Informed Futures*. <https://informedfutures.org/about/>
- University of Auckland. (2025b). *Artificial Intelligence*. <https://www.auckland.ac.nz/en/study/study-options/find-a-study-option/artificial-intelligence.html>
- University of Moratuwa. (2024a). *Bachelor of Science Honours in Artificial Intelligence*. <https://uom.lk/itfac/bachelor-science-honours-artificial-intelligence>
- University of Moratuwa. (2024b). *Master of Science/Postgraduate Diploma in Artificial Intelligence*. <https://uom.lk/itfac/education/postgraduate/artificial>
- University of Peradeniya. (2024). *Artificial Intelligence and Signal Processing*. <https://web2.ee.pdn.ac.lk/courses/artificial-intelligence-and-signal-processing>
- University of the Philippines. (2025a). *University of the Philippines Principles for Responsible and Trustworthy Artificial Intelligence*. <https://up.edu.ph/up-principles-for-responsible-artificial-intelligence/>
- University of the Philippines. (2025b). *University of the Philippines Principles for Responsible and Trustworthy Artificial Intelligence*. <https://up.edu.ph/up-principles-for-responsible-artificial-intelligence/>
- University of Waikato. (2025). *Artificial Intelligence*. <https://www.waikato.ac.nz/int/study/subjects/artificial-intelligence/>
- UPD College of Engineering. (2025). *Artificial Intelligence*. <https://coe.upd.edu.ph/masters-of-engineering-in-artificial-intelligence/>
- Vaid, R. (2024, November 18). *Digiyaatra to support 22 languages, expand to 4 new airports*. *BusinessLine*. <https://www.thehindubusinessline.com/economy/logistics/digiyaatra-to-support-22-languages-expand-to-4-new-airports/article68880790.ece>
- Victoria University of Wellington. (2023a, March 27). *Data science for remote sensing and tree modelling*. Victoria University of Wellington; Victoria University of Wellington. <https://www.wgtn.ac.nz/cdsai/current-research-projects/remote-sensing-and-tree-modelling>

- Victoria University of Wellington. (2023b, May 18). *Centre for Data Science and Artificial Intelligence*. Victoria University of Wellington; Victoria University of Wellington. <https://www.wgtn.ac.nz/cdsai>
- Wakefield, J. (2019, April 23). *Sri Lanka attacks: The ban on social media*. <https://www.bbc.com/news/technology-48022530>
- Weerasinghe, T. (2024, September 1). *New National Artificial Intelligence Strategy aligns with national and international best practices*. The Sunday Times, Sri Lanka. <https://www.sundaytimes.lk/240901/news/new-national-artificial-intelligence-strategy-aligns-with-national-and-international-best-practices-569982.html>
- White & Case LLP. (2024, July 1). *AI Watch: Global regulatory tracker - Japan*. <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-japan>
- World Bank. (n.d.). *Artificial Intelligence in the Public Sector*. <https://documents1.worldbank.org/curated/en/746721616045333426/pdf/Artificial-Intelligence-in-the-Public-Sector-Summary-Note.pdf>
- Yachu, S. (2023, August 8). *Explained: The Digital India Act 2023*. Vidhi Centre for Legal Policy. <https://vidhilegalpolicy.in/blog/explained-the-digital-india-act-2023/>
- Yondu Team. (2024, September 3). *The State of Artificial Intelligence in the Philippines*. Yondu. <https://www.yondu.com/articles/the-state-of-artificial-intelligence-in-the-philippines>
- York Centre for Asian Research. (2024). *Fact Checking in Low-Resource Languages: A New Dataset and Transformer Model for the Burmese Language*. York Centre for Asian Research. <https://www.yorku.ca/research/ycar/events/fact-checking-dataset-transformer-model-burmese-language-lwin-moe-13112024/>
- Yossi, M. (2020, September 1). *A big step for flood forecasts in India and Bangladesh*. Google. <https://blog.google/technology/ai/flood-forecasts-india-bangladesh/>
- Zhang, L. (2023, April 26). *China: Provisions on Deep Synthesis Technology Enter into Effect* [Web page]. Library of Congress, Washington, D.C. 20540 USA. <https://www.loc.gov/item/global-legal-monitor/2023-04-25/china-provisions-on-deep-synthesis-technology-enter-into-effect/>

Appendix A: GIRAI Methodology

Overview

Data for the GIRAI was collected first-hand by 138 researchers. Researchers completed a comprehensive expert survey containing 1862 questions (98 per thematic area) designed to ascertain what conditions were in place and what actions were being taken to advance responsible AI in each country surveyed, between November 2021 and November 2023. A global team of quality assessors conducted an exhaustive review of all data collected. Country scores are derived from the primary data gathered across the three pillars. More information on methodology can be found in the Methodology and Conceptual Framework of the Global Index on Responsible AI [here](#).

The countries included in the 1st Edition of the Global Index on Responsible AI were chosen based on where suitable country-researchers were recruited and the existing research network of institutional partners. Governments were not involved in determining whether their country was included in the Index.

Methodological Limitations

Before we proceed to the data, a couple of methodological limitations that ought to be discussed here.

- **Limited Scope of Evidence Collection:** The datasheet does not provide a comprehensive record of all Responsible AI (RAI) actions within a country. Researchers were instructed to document only one piece of evidence per actor category per thematic area, categorized under each pillar and type of non-state actor.
- **Equal Weighting of Evidence Types:** The scoring system does not differentiate between types of evidence. For example, a newspaper op-ed by an academic and a newly developed Responsible AI toolkit by a practitioner are treated equally, despite their differing levels of impact.
- **No Assessment of Implementation or Enforcement:** The index does not measure the effectiveness of enforcement or the success of implementation. Policies, regulations, and strategies that meet the criteria for evidence are included, regardless of how well they are enforced or executed.

Appendix B: Heatmap of Adjusted Scores for Asia by Country by pillar sorted by score

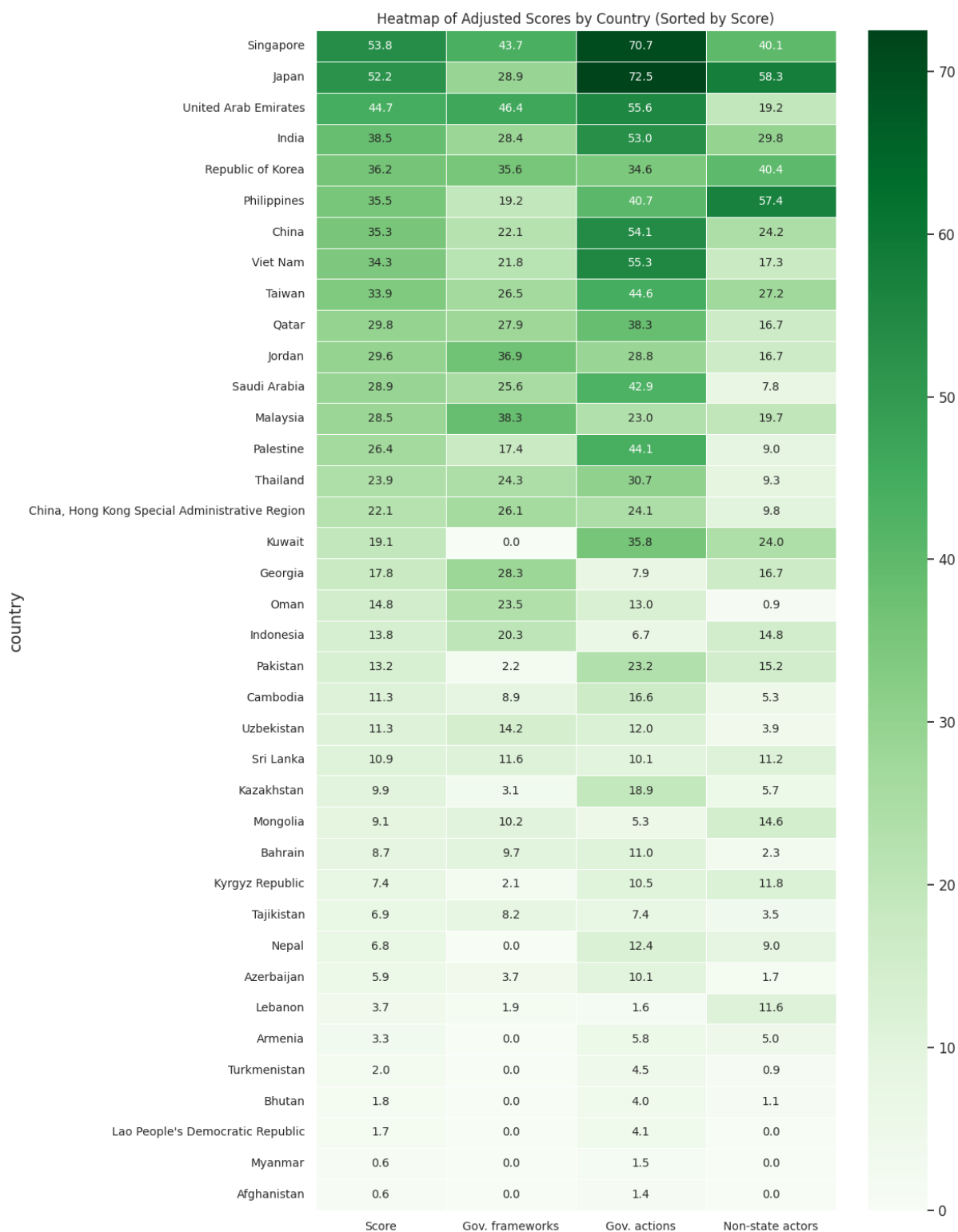


Figure 37: Heatmap of Adjusted Scores by Country by pillar sorted by score

Appendix C: Description of the scope of thematic areas

This appendix contains brief descriptions of each thematic area, either described directly linked to AI, or with the understanding that the substance of that area – for example, children’s rights – should either be upheld when that substance is desirable; circumvented when it is not, in relation to every stage and use of AI. Elaborations and justifications for choice of thematic area can be found in their respective documents.

National AI Policy

“A National AI Policy can be defined as ‘a course or principle of action adopted formally agreed to by the government’ in relation to the country’s approach to AI.”

Impact Assessments

“Impact assessments can be used as a tool for both predicting the anticipated, and assessing the actual, consequences of an AI system in terms of benefits and harms. When assessing anticipated risks, impact assessments require users of AI technologies to consider the type and scope of impact before and after deployment.”

Human Oversight and Determination

“Human oversight requires systems to be designed in a way that allows for human supervision or control and provides clear channels for intervention or interference by humans to prevent or disrupt adverse effects in the use and deployment of AI tools. Human determination enables individuals to exercise judgment, evaluate complex ethical considerations, and make informed decisions that consider the broader societal impact, ensuring that AI is aligned with human values and responsible use.”

Responsibility and Accountability

“The concept of responsibility concerns the humans or decision-making body (whether public or private) to whom causal relations to a certain outcome may be assigned. This is different from the concept of ‘responsible AI’ in that it identifies who (i.e. human legal entities, etc.) is responsible for what (i.e., decisions, outcomes, impacts), rather than ensuring the use of AI in a safe, ethical and trustworthy way.”

Proportionality and Do No Harm

“Proportionality demands the use of AI systems to not ‘exceed what is necessary to achieve legitimate aims or objectives and should also be appropriate to the context’. When read together, the principle of proportionality and ‘do no harm’ require AI systems not to be used in ways that cause or exacerbate harm, whether at an individual or communal level, and includes injuries and injustices that impact on the social, cultural, economic, natural or political environments.”

Public Procurement

“The focus on public procurement centers on ensuring that governments adhere to international procurement laws and principles of responsible AI as outlined in AI policy frameworks. These frameworks emphasize the importance of fair, transparent, inclusive, diverse, and non-discriminatory procurement of AI systems.”

Transparency and Explainability

“Transparency concerns the extent to which the “inner workings” of an AI system are open and accessible, which invariably requires the provision of “easy-to-understand” explanations of the algorithmic models, the data that drives them, and the rationale for their use. Accordingly, the principle of transparency implicates the concept of explainability, which has arisen within the AI ethics discourse and which calls for the provision of information that can be understood by a non-technical person – particularly, the process driving outcomes of AI systems.”

Access to remedy and redress

“Where persons who have suffered harm as a result of the development or use of an AI system to be able to submit complaints, to pursue legal actions in court, or to report issues to a competent authority and to have those harms addressed with due diligence.”

Safety, Accuracy and Reliability

“AI safety focuses specifically on technical solutions to ensure that AI systems and tools operate safe and reliably and do not introduce new harms or exacerbate existing risks. The principle of accuracy requires immediate steps to be taken to remove or rectify inaccurate data to ensure algorithmic outputs are both correct and precise. Reliability is the probability that an AI tool or system will perform its intended function correctly for a specific period of time and produce consistent and reproducible results.”

Gender Equality

“Gender equality is defined as the granting of “equal rights, responsibilities and opportunities” to people of all gender identities, which extends beyond the traditional binary concept of equality between women and men, and girls and boys.” In the context of AI, it is desirable that AI systems reflect – or, at a minimum, do not compromise – this principle, at every stage, from design to use.

Data Protection and Privacy

“The right to privacy is defined as the right of a person ‘to be free from intrusion into publicity concerning matters of a personal nature’ by government and/or third parties, including private entities or individuals...The right to privacy has evolved to include information privacy, which is defined as the right of an individual to have control over who has access to their personal data and for what purpose.”

Bias and Unfair Discrimination

Bias is defined as a prejudice for - or against - an individual or group in a way that oftentimes unfair...Algorithmic bias is defined as systematic and repeatable errors in an AI-power system that create unfair outcomes, such as privileging one group over others... Discrimination is defined as the practice of treating an individual or group of people a way that is different to others because of their age, disability, gender, race, sexuality, or other characteristic.”

Children’s Rights

“Children’s rights refer to the subset of human rights that recognises the need to provide children with ‘special care and protection’ given their dependence on adults for survival, protection, and development. A ‘child’ is defined as anyone under the age of 18, or until they attain majority, or legal age, earlier under an applicable law.” It would follow that AI systems ought not to compromise children’s rights.

Labor Protection and Right to Work

Labor protections are broadly defined as the ‘protection of employment condition working conditions, and labor welfare as well as occupational safety, health and environmental conditions provided for an employee’...The right to work is recognised as a fundamental right under Article 23 of the Universal Declaration of Human Rights (UDHR) which provides that ‘[e]veryone has the right to work, to free choice of employment, to just and favorable conditions of work and to protection against unemployment.’”

Cultural and Linguistic Diversity

“Cultural and linguistic diversity is an underpinning value of responsible AI and may be advanced through the active participation of all groups regardless of race, color, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability, or any other ground.”

Competitions Authorities

“A competitions authority, commonly referred to as a ‘competitions regulator’ is responsible for overseeing the fair functioning of markets in a given country. While their form may vary across countries, most competition authorities are independent ‘stand-alone’ bodies and/or agencies that have the legal authority and mandate to regulate and enforce competitions law (e.g. antitrust law), and, in some instances, consumer protection laws.”

Public Sector Skills Development

“Skills development refers to the technical knowledge, capabilities and proficiencies that are required to ‘successfully integrate AI into functions’ across various industries, in both the public and private sectors. This should not be interpreted to include educational curriculum provided at the tertiary level, which will be covered under the Education Thematic Area.”

International Cooperation

International cooperation refers to joint efforts between countries to align policy instruments in relation to AI to ensure: (1) global adherence to responsible AI principles; (2) inclusive and equitable access to the benefits and opportunities offered by AI; (3) trust and accountability mechanisms between countries; and (4) advancement of scientific research and technical knowledge.”

Appendix D: Key Informant Interviews

1. Tom Barraclough – Tech Policy Researcher and Director, Brainbox Institute, New Zealand

Working Draft