AI IN PUBLIC SECTOR TRANSPARENCY TOOL

Global CSO and Academic Institutions Network on AI Ethics & Policy

Subgroup on AI in the Public Sector



In support of

unesco

About the Global CSO and Academic Network on AI Ethics and Policy

Facilitated by UNESCO, the Network brings together CSOs and academic institutions from around the world to advance ethical, inclusive, and human rights–based approaches to AI governance. Launched in 2025 at the 3rd Global Forum on the Ethics of AI in Thailand, the Network serves as a platform for cooperation, knowledge exchange, and policy engagement, enabling its members to advance collaborative projects and partnerships to address the complex challenges posed by AI.

About the AI in Public Sector Subgroup

The subgroup examines the impact of AI development and deployment within the public sector, with a particular focus on strengthening transparency and accountability throughout the full development and implementation process. It also looks at the role of AI procurement processes, and at methodologies for participatory design and community-led audits.

Members of the AI in Public Sector Subgroup

Derechos Digitales; Digital Empowerment Foundation; The Open Data Institute; Comité Español de Representantes de Personas con Discapacidad (CERMI); The Ada Lovelace Institute; IT for Change; Tactical Tech; Technology and Global Affairs Innovation Hub Sciences Po; Homo Digitalis; OBVIA; The Regulatory Institute; InternetLab; Observatory on the Social and Ethical Impact of Artificial Intelligence (OdiseIA); Digital Futures Lab; AI Asia Pacific Institute; The Center for AI and Digital Policy (CAIDP); European Law Institute; AfroLeadership; GobLab UAI; Glitch Charity; Centre for Information Technology and Development (CITAD); Youth for Privacy; Observatorio de IA, Innovación, Sociedad y Gobierno; Association Al Intilaka for Development, Environment, and Culture - Afourer (AIDECA); Observatoire des algorithmes publics (ODAP).

Acknowledgements

This report was curated as part of one of the short-term work streams of the AI in the Public Sector Subgroup of the UNESCO Global Civil Society Organizations (CSO) and Academic Network on AI Ethics and Policy between September 2025 and April 2026. Our sincere gratitude to the members of the Subgroup for their inputs and questions which shaped the Repository and the report.

The UNESCO Ethics of AI team was instrumental in organizing and ensuring our work was aligned with the broader objectives of the Network and reflected the commitments of the Subgroup.

About this report

Text: Abhineet Nayyar, Amrita Sengupta and Jamila Venturini

Review: Abhineet Nayyar, Jamila Venturini and Shazade Jameson

About the Transparency Tool

Subgroup lead: Jamila Venturini

Systematization and consolidation: Amrita Sengupta

Preliminary research (Comparative repository): Dr. Grace S. Thomson, Merve Hickok and Christabel Randolph (CAIDP); Jamila Venturini and Juan Manuel García (Derechos Digitales); Clarice Tavares and Victor Pavarin (InternetLab); Abhineet Nayyar (IT for Change); Andréane Sabourin Laflamme and Sehl Mellouli (OBVIA); Dolores Abuin, Rafael González, Nicole Marie Suárez and Juan de la Cruz (OdiseIA)

Review and feedback: Abhineet Nayyar, Achim Rosemann, Andréane Sabourin Laflamme, Aarushi Gupta, Charlie Ngounou, Clarice Tavares, Dolores Abuin, Grace S. Thomson, Jamila Venturini, Juan Manuel García, Karine Gentelet, Maria Paz Hermosilla Cornejo, Rafael González, Shivangi Sharan, Victor Pavarin and YZ Ya'u

Table of Contents

Table of Contents iii

Background note 1

Scope 3

Methodology 4

Introducing the AI in the Public Sector Transparency Tool 10

How to read the tab Draft of Base Framework V1 10

How to read the AI in the Public Sector Transparency Tool 10

Recommendations 18

Background note

Artificial intelligence (AI) systems are being increasingly integrated into the public sector. Governments deploy machine learning and automated decision-making systems (ADM) to support their operations in different fields including welfare administration, taxation oversight, immigration processing, policing, urban planning, and health system management. These deployments influence administrative discretion, resource allocation, and often the delivery of services that operationalise citizens' rights. As a result, governance mechanisms that ensure transparency, accountability, and public oversight are an essential part of the deployment of AI by the public sector.

States' obligations to disclose information regarding their use of technology derives mainly from their commitment with the right to access information guaranteed in different standards including the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights, and operationalized in national Constitutions and norms. As highlighted by Global Partnership on AI, "information about how the state makes certain decisions affecting people's lives and liberties is indispensable to protecting the right to due process".[1] Additionally, it "enables citizen oversight over governmental activities and decisions associated with the adoption and implementation of ADM systems. For example, accessing meaningful information may allow civil society organizations to assess whether the use of ADM systems is compliant with the law and how public bodies allocate economic resources".

With regards to AI, the UNESCO Recommendations on the Ethics of AI, adopted by 194 member States, determines that

> The transparency and explainability of AI systems are often essential preconditions to ensure the respect, protection and promotion of human rights, fundamental freedoms and ethical principles. Transparency is necessary for relevant national and international liability regimes to work effectively. A lack of transparency could also undermine the possibility of effectively challenging decisions based on outcomes produced by AI systems and may thereby infringe the right to a fair trial and effective remedy, and limits the areas in which these systems can be legally used. (Recommendation n. 37)

And that

> [...] People should be fully informed when a decision is informed by or is made on the basis of AI algorithms, including when it affects their safety or human rights, and in those circumstances should have the opportunity to request explanatory information from the relevant AI actor or public sector institutions. In addition, individuals should be able to access the reasons for a decision

[1] GPAI (2024). "Algorithmic Transparency in the Public Sector: A state-of-the-art report of algorithmic transparency instruments". Report, May 2024, Global Partnership on Artificial Intelligence.

> affecting their rights and freedoms, and have the option of making submissions to a designated staff member of the private sector company or public sector institution able to review and correct the decision. AI actors should inform users when a product or service is provided directly or with the assistance of AI systems in a proper and timely manner. (Recommendation n. 38)

While the transparency obligations regarding the use of AI by public institutions or in the delivery of public services are grounded on existing commitments and do not require the adoption of specific regulation on AI, evidence shows that countries are failing to comply with them.[2] This applies both to proactive and reactive measures on access to information, as well as to the disclosure of information regarding the processing of personal data often implied in the deployment of such technologies. In some cases even individuals affected are not aware of the automated interference in decision-making or the mechanisms for challenge. Broad exceptions to access to information and data protection norms facilitate denial of public interest information. When they exist, supervising authorities on data protection and access to information often lack a specific mandate, resources and capacities to monitor State's use of AI.

This doesn't mean that States are by default prepared to implement transparency measures. As the Global Partnership on AI correctly points out in its report on the matter, besides being a right from citizens and an obligation by States, algorithmic transparency also requires standards that may, for instance, "concern the quantity or quality of the disclosed information and how it is disclosed".[3]

Several accountability mechanisms have emerged in recent years in different countries, trying to respond to such needs. Examples include the Government of Canada's Algorithmic Impact Assessment and AI registry, the United Kingdom's Algorithmic Transparency Recording Standard, and the European Union's Artificial Intelligence Act. Despite these developments, existing accountability tools are unequal in scope and geographic representation. Many repositories and transparency mechanisms focus primarily on North American and European deployments of AI. Public sector innovation occurring across Arab States, Asia and the Pacific, Latin America and the Caribbean, and Africa is less systematically documented. A transparency framework that establishes references for disclosure of information incorporating perspectives from the Global South is therefore necessary to ensure that global discourse and accountability requirements reflect the diversity of governance contexts in which AI systems operate.

This report summarizes the outputs of the work undertaken by the subgroup "AI in the Public Sector" at the UNESCO CSO and Academic Network on AI Ethics and Policy. It reflects a collaborative effort developed between September 2025 and April 2026, which aimed to increase transparency and accountability on the use of AI by the public sector through the development of a framework that can serve a few outcomes:

[2] Velazco, P. & Venturini, J (2021). "Automated decision-making in public administration in Latin America". 2021, Derechos Digitales.

[3] GPAI (2024). "Algorithmic Transparency in the Public Sector: A state-of-the-art report of algorithmic transparency instruments". Report, May 2024, Global Partnership on Artificial Intelligence.

1. To provide a robust AI transparency template that can be used by both government officials and by civic actors (civil society, academics, journalists, citizens) for seeking transparency on use of AI in public sector;
2. Bridge gaps in existing frameworks and tools for seeking such accountability;
3. Ensure relevance for Global South realities of opaque and under-regulated information disclosure

Scope

The work within the AI in the Public Sector subgroup involved several stages for reaching a Transparency Tool as described further in the Methodological Note. The Tool was designed to be used by different actors including civil society, in their efforts to hold States and the private sector accountable for the use of AI. It can also inspire repositories from States and multilateral organizations to increase transparency on their practices and facilitate compliance with existing obligations the respect of human rights within AI adoption, including the UNESCO Recommendations on the Ethics of AI. Because of this, the framework was built to be flexible and adaptable to different regulatory environments, not relying exclusively on specific regional, subregional or national legislation.

This Tool builds on top of existing work on public oversight and participation to draw from observed best practices and further expand on them, such that it specifically serves accountability and transparency outcomes.

It aims to provide a comprehensive accountability mechanism for assessing artificial intelligence and automated decision-making systems deployed by government agencies across all functions, risk levels, and implementation contexts. At this stage, the scope encompasses both narrow AI systems performing specific tasks (e.g., fraud detection, application triage, risk scoring) and general-purpose AI systems adapted for public sector use (e.g., large language models deployed for citizen services, document processing, or policy analysis).

The Transparency Tool can be used to document uses of AI systems at all stages of maturity: early research and proof-of-concept projects exploring potential applications; pilot and testing deployments with limited user populations; production systems serving operational government functions; and legacy systems under review for modernization or decommissioning. It explicitly includes systems that governments have decommissioned or discontinued, so as to document failures and lessons learned is essential for avoiding repeated harms. At the same time, it also acknowledges the federal and multi-domain nature of AI deployment and maps its use in the public sector across different departments and authorities, as well as at the National and Subnational levels, globally.

Finally, it explicitly extends beyond technical assessment to encompass social, ethical, legal, and

human rights dimensions. While we do not explore this feature in detail, the tool's modularity potentially allows enforcers and CSOs to use it in complementarity with existing and tested evaluation methods, such as UNESCO's Ethical Impact Assessment (EIA) or country-specific Human-Rights Impact Assessment frameworks (HRIA).

Methodology

This section describes the process through which the UNESCO AI Transparency Framework for the Public Sector was developed within the Subgroup AI in the Public Sector. The work departed from a collective benchmarking exercise involving members from different regions who mapped almost 30 similar initiatives and their main characteristics in order to draw learnings from their methodology, scope and structure. An analysis of them allowed the subgroup to identify strengths, weaknesses and the voids to be filled by an academic and civil society led template. This evolved into a framework that later inspired the current version of the Transparency Tool, as described below.

The development process went through key steps: the construction and analysis of a comparative repository of existing initiatives; the development of the assessment framework and the transparency template. Each step was guided by in-depth review and discussion with subgroup members including academic organizations and civil society organizations, mostly from the Global South, in all cases with expertise in analyzing AI deployment in the public sector in their countries and regions.

Stage 1: Comparative Repository

The first stage involved the collaborative construction of a structured repository of existing AI transparency, accountability, and algorithmic governance initiatives. This work was carried out collectively by subgroup members, drawing on publicly available sources including government registries, academic publications, civil society databases, and multilateral policy instruments.

The repository documents 27 initiatives spanning multiple continents and governance contexts.

Table 1: List of mapped initiatives focused on increasing transparency on AI adoption by public institutions

Name of existing initiative	Source	Hosting organization(s)
Confia – Generalitat Valenciana Public Algorithm Register	LINK: https://gvaoberta.gva.es/es/confia/cerca PDF: https://www.uv.es/cotino/publicaciones/informe_gvav3logos.pdf?utm_source=chatgpt.com	Generalitat Valenciana, Dirección general de transparencia y participación (Directorate-General for Transparency and Participation) Spain

Public Registry of Algorithms of the Basque Country - Registro Público de Algoritmos de Euskadi	LINK: https://www.euskadi.eus/web01-a3strbil/es/?fullText=%20&r01kLang=es&r01kQry=tC:euskadi;tF:informacion;tT:informacion;m:contentName.LIKE.adimen_artifiziala_algorit_,documentLanguage.EQ.es;o:contentCreateDate.DESC;p:Inter;pp:r01NavBarBlockSize.10,r01PageSize.10	Gobierno Vasco, Departamento de Gobernanza, Administración Digital y Autogobierno (Basque Government, Department of Governance, Digital Administration and Self-Government) Spain
AI 4 Bharat	LINK: https://ai4bharat.iitm.ac.in/	Indian Institute of Technology, Madras
Observatório Brasileiro de Inteligência Artificial (OBIA)	LINK: https://obia.nic.br/	Núcleo de Informação e Coordenação do Ponto BR (NIC.br)
CAIDP Index (AI and Democratic Values)	https://www.caidp.org/reports/caidp-index-2025/	Center for AI & Digital Policy
AIKosh	https://aikosh.indiaai.gov.in/home	AIKosh mission
Mapping Automated Decision Making in Australian Social Services	https://apo.org.au/sites/default/files/resource-files/2022-10/apo-nid321337_0.pdf	Australian Research Council (ARC) Centre of Excellence for Automated Decision-Making and Society
Observatoire des Algorithmes Publics	https://odap.fr/	ODAP (independent)
Government of Canada AI Register (Minimum Viable Product) - GC AI Register (MVP)	https://open.canada.ca/data/en/dataset/fcbc0200-79ba-4fa4-94a6-00e32facea6b/resource/369f6f34-148a-42ed-b581-8c164e941a89?view_id=cd6b5d4e-1943-43d1-966b-f1b6cdd7625a	Government of Canada
Ethical AI Alliance, AI Harm Map	https://ethicalaialliance.org/	
Portrait public des utilisations de l'IA dans l'administration publique québécoise	https://www.donneesquebec.ca/recherche/dataset/portrait-public-des-utilisations-de-l-ia-dans-l-administration-publique-quebecoise	Government of Québec

Barcelona City Council Algorithmic Transparency Register	https://ajuntament.barcelona.cat/digital/en/technology-accessible-everyone/ethical-use-artificial	Ajuntament de Barcelona (Barcelona City Council) Spain
Tarragona AI Algorithm Register - Registre d'algorismes d'intelligencia artificial (IA) de l'Ajuntament de Tarragona	https://www.tarragona.cat/governobert/tgn-dades/registre-dalgorisme-dinteligencia-artificial	Ajuntament de Tarragona (Barcelona) Spain
AESIA AI Regulation Sandbox	https://aesia.digital.gob.es/es/guias	AESIA (Spanish Agency for Supervision of AI)
Civio Foundation - Algorithmic Accountability Reporting	https://civio.es	Fundación Ciudadana Civio
ONTSI - AI adoption reports	https://www.ontsi.es/es/publicaciones/indicadores-de-uso-de-inteligencia-artificial-en-espana-2024	Red.es / Ministry for Digital Transformation
AEPD Agencia española de protección de datos - Spanish Data Protection Agency	https://www.aepd.es/guias-y-herramientas/guias	Agencia Española de Protección de Datos (AEPD)
CTT - Technology Transfer Centre	https://administracionelectronica.gob.es/ctt/CTTprincipalEs.htm?urlMagnolia=/pae_Home/pae_SolucionesCTT.html	Centro de Transferencia de Tecnología - Ministry for Digital Transformation
OEIAC - Observatory of Ethics in AI of Catalonia	https://oeiac.cat/es/	University of Girona / Generalitat de Catalunya
OdiseIA (Observatorio del Impacto Social y Ético de la IA)	https://www.odiseia.org	OdiseIA
BSC - Proyectos ANIA	https://projecteaina.cat/	Generalitat de Catalunya and Barcelona Supercomputing Center (BSC)
Sistema de Algoritmos Públicos	https://algoritmos.uniandes.edu.co/	Universidad de Los Andes (Colombia)
Repositorio de Algoritmos Públicos	https://algoritmospublicos.cl/repositorio and https://www.consejotransparencia.cl/wp-content/uploads/estudios/2021/10/ESTUDIO-TRANSPARENCIA-ALGORITMICA-EN-EL-SECTOR-PUBLICO-GOBLAB-cambio-tablas-1.pdf	Universidad Adolfo Ibañez (Chile)
FairLAC	https://fairlac.iadb.org/pilotos	Inter American Development Bank

AI & Inclusion	https://ia.derechosdigitales.org/en	Derechos Digitales
Tracking Automated Government (TAG) Register Canada	https://tagcanada.shinyapps.io	Starling Centre at Western University
Algorithmic impact assessment (AIA) Reports - Open Government Canada		Government of Canada

For each initiative, subgroup members recorded the hosting organization type, geographic coverage, parametric coverage across six variables, methodology, data update frequency, and known limitations. Selected initiatives were subject to more detailed analysis, with narrative summaries recorded in a dedicated annotation sheet. This document was a very useful starting point to think about what exists, what is missing and therefore how the transparency framework/tool/template needs to be built out.

Following the completion of the repository, a qualitative-quantitative analysis was undertaken to identify patterns, gaps, and areas of relevance for the framework development process. The analysis, led by the subgroup research assistant, examined both the coverage and the methodological approaches of the documented initiatives and was critical for further discussion on the scope of the Transparency Tool.

The majority of documented initiatives are concentrated in North America, Western Europe, and Latin America. South Asian initiatives such as AI4Bharat and AIKosh are present but tend to focus on AI development and cataloguing rather than on transparency mechanisms. In terms of parametric coverage, most consistently document system names and deploying agencies. Coverage of procurement details, vendor identity, and contractual arrangements is less consistent. Impact metrics and human oversight mechanisms are rarely disclosed. Data on demographic bias, model performance disaggregated by population group, and grievance mechanisms are largely absent across the repository.

A closer analysis was conducted on the Canadian AI in Public Sector database. While it was not possible to extensively review in detail all mapped databases, initial analysis helped identify what could be useful within the Transparency Tool and existing gaps in current disclosures for AI use by governments. Preliminary findings pointed out that:

- First, there is limited attention to post-deployment monitoring, including model drift, ongoing bias auditing, and incident reporting.
- Second, redressal and grievance mechanisms are not well documented.
- Third, few initiatives address the accountability implications of public-private partnerships or third-party vendor arrangements.
- Fourth, the application layer, including explainability to affected citizens, is poorly covered.
- Finally, sector-specific accountability requirements, particularly for high-risk areas such as law enforcement, immigration, and health, are not systematically addressed in most existing frameworks.

Further discussion within the subgroup confirmed a perception that existing frameworks tend to assess only one or two stages, most commonly focusing on deployment-time transparency while neglecting both upstream accountability (who decided to use AI, on what legal basis, with whose data) and downstream accountability (what actually happened, and can affect parties seek redress). Members also flagged the need for a framework that could be used not only by government officials conducting internal assessments but also by civil society actors, journalists, and affected communities seeking external accountability

Stage 2: Development of an Assessment Framework

Prior to the design of the first version of the Transparency Tool, an assessment framework was proposed by the subgroup research assistant and discussed with members. It was structured around five phases of the AI lifecycle: pre-design, pre-deployment, design, deployment, and post-deployment. Within each phase, assessment dimensions were identified based on previous analysis and discussions, drawing on existing frameworks and standards where relevant.

The primary reference frameworks consulted in the development of the assessment criteria included: the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021); the OECD Principles on Artificial Intelligence (2019, updated 2024); the European Union Artificial Intelligence Act (Regulation (EU) 2024/1689); EU Public Sector Tech Watch dashboard, G7 reporting framework – Hiroshima AI Process (HAIP) international code of conduct for organizations developing advanced AI systems (2024), the Government of Canada Directive on Automated Decision-Making and Algorithmic Impact Assessment; the UK Algorithmic Transparency Recording Standard; the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679); the UN Fundamental Principles of Official Statistics; and Open Contracting Standards. In order to ensure the framework reflects governance contexts beyond North America and Europe, additional reference was made to the NITI Aayog Approach Document for India: Principles for Responsible AI (2021); the African Union Continental Artificial Intelligence Strategy (2024); and the Brazilian Framework for the Development, Use and Governance of AI Systems (Bill No. 2338/2023, approved by the Federal Senate in December 2024). Several countries have mandated use case inventories (UK, France, US, Canada) and are publishing on their website or in the case of US in GitHub.

For each assessment question, the framework specifies the expected response format (Yes/No/Partial/Not Sure/Not Applicable), examples of evidence required. The framework applies across all AI system types, including narrow AI systems and general-purpose AI adapted for public sector use, and across all stages of system maturity from research and proof-of-concept through to decommissioned systems.

Stage 3: Development of the Transparency Template

Building on the assessment framework, a public disclosure template was developed to enable standardised system-level reporting. The template is structured to allow one row per AI system and organises disclosure fields by lifecycle phase.

The template draws on the parametric coverage analysis from the repository, incorporating fields that were consistently present in existing registers while adding fields identified as absent. Key additions include fields for impact assessments, demographic disaggregation of performance metrics, vendor audit rights, human override procedures, grievance mechanisms, and incident reporting.

The template is designed for use in multiple formats, including as a spreadsheet for government self-reporting, as a structured data file for civil society and researcher use.

Through a series of discussions and rounds of feedback with the working group, a first version of the Transparency Tool was finalised. A pilot test of the Tool was conducted by subgroup members on cases previously analyzed or of knowledge. They collectively illustrate how the template performs across different AI system types, governance contexts, risk levels, and disclosure environments. They surface critical commonalities: limited pre-deployment impact assessment, weak grievance and redress mechanisms, poor documentation of vendor accountability, and insufficient transparency about data retention and bias testing.

On the Tool itself, feedback and discussions included several key factors. On usability, reviewers from the subgroup noted that several fields presuppose institutional arrangements that may not exist in all governance contexts, particularly in the Global South, and that guidance should reflect these realities. On depth versus accessibility, they identified a consistent tension between comprehensive accountability data and usability for civil society actors with limited access to government documentation, with the Core and Supplementary distinction serving as the primary balancing mechanism and also the need to reduce the number of fields for greater accessibility and usability. On definitional clarity, reviewers called for clearer definitions of key terms including legal basis for data use, training versus inference infrastructure, and the distinction between performance monitoring and audits. On human rights centrality, detailed discussions were held on how best to incorporate this section and the present version reflects the consensus that was built in the last meeting for this work, in April, 2026.

While the Tool is mature to be presented as a critical advance from the subgroup in this first phase of collaborative work, members expect to continue to refine versions of it after further iteration with the full UNESCO CSO and Academic Network on AI Ethics and Policy, as well as other stakeholders.

Introducing the AI in the Public Sector Transparency Tool

The full Transparency Tool is openly available for download and use by interested parties here. Below is a description and guidance to read and use the document.

How to read the tab Draft of Base Framework V1

This tab is the precursor to the Transparency Tool and the result of an extended analysis of existing AI standards in different regions of the world. It reflects a stage in which the subgroup was not clear about the desired outcome, but allowed it to mature collective perspectives on key elements that were later integrated into the final outcome. It is organized into multiple complementary components that should be used together for having a comprehensive AI system assessment. Some parts of the framework are still under development and all parts are open for discussion and revision.

The tab contains 55+ assessment questions organized by AI lifecycle phase: pre-design; pre-deployment; design; model development; deployment; and post-deployment. Each question specifies the expected response format (Yes/No/Partial/Not Sure/N/A) and the evidence required for verification.

How to read the AI in the Public Sector Transparency Tool

The complete Tool is openly available here for interested parties to download, peruse, and deploy at their own pace. New versions may continue to be updated in the same link, and below is a description and guidance to read and use the document.

For ease of use, we have organized the transparency tool in a series of questions that CSOs and public authorities can ask of themselves and of the AI system in question. These questions span across the procurement life-cycle, covering everything from the system's design to its deployment. Furthermore, each question elicits an objective answer, followed by an expectation of concrete evidence in support of that answer, thereby enabling subsequent quantitative analyses and comparisons between sectors, countries, and different public authorities.

The tool is structured into 14 distinct categories, with a total of 50 questions designed to capture transparency information across the lifecycle of an AI system. Aware of the diverse informational context across the Global South, we label these questions as core and supplementary, tagging 32 as the former to indicate their relative importance over the remaining 18. These 14 categories are listed below:

1. Administrative Note (1 field): Timestamp for when the tool was populated.
2. System Identification (10 fields): Basic metadata including country, administrative level,

agency in charge, and lifecycle status.

3. Problem Definition (6 fields): Focuses on the "why" and "how" of the system, including purpose, automation level, and stakeholder consultations.
4. Procurement (6 fields): Details on the vendor, procurement method, contract value, and government audit rights.
5. Data (5 fields): Information on data sources (public/private), data types, and collection methods.
6. Model & Performance (3 fields): Technical transparency regarding the model type and performance metrics.
7. Deployment (2 fields): Scope and scale of the system's actual use.
8. Human Rights & Impact Assessment (3 fields): Whether impact assessments (human rights, ethics, or privacy) were conducted and published.
9. Monitoring (2 fields): Ongoing performance tracking and oversight.
10. Audits (3 fields): Information on internal and external audits, including links to reports.
11. Incident Reporting (3 fields): Mechanisms for reporting human rights-impacting events (e.g., wrongful exclusions or discriminatory decisions).
12. Redressal Mechanism (2 fields): Procedures for affected persons to challenge decisions and seek remedies.
13. Compliance (3 fields): Governance oversight, data protection law applicability, and specific AI legislation.
14. Disclosure Notes (1 field): A critical field for identifying what information remains unavailable.

Additionally, below are some qualitative notes to further help readers understand as well as fill the tool, with some of the notes being general while others being specific to certain fields.

First, a number of fields in this Tool, particularly those relating to audits, monitoring, and incident reporting, ask for information that is frequently undisclosed, ambiguous in its public documentation, or difficult to attribute specifically to the AI system rather than to the broader policy or institutional context in which it operates. When completing these fields, respondents are encouraged to only answer where there is direct documentary evidence that the information refers specifically to the AI system being documented. Where the available evidence is indirect, assumed, or relates to the broader policy or institution rather than the system itself, select ‘Unclear’ and use the free-text fields to note what is known, what is assumed, and what remains undocumented. This approach ensures that the tool captures genuine disclosure gaps rather than obscuring them with inferred answers, while also leaving open questions for further reflection.

Similarly, throughout the Tool, a distinction is maintained between the stated purpose, function, or outcome of an AI system as documented by the responsible institution, and the observed or reported purpose, function, or outcome as documented by independent sources. Most fields should be completed using official sources, wherever possible. Where official documentation is absent, partial, or contradicted by independent evidence, respondents should note this explicitly rather than substituting their own characterisation of the system.

Second, the sector classification draws on the United Nations Classification of the Functions of Government (COFOG), the internationally adopted standard for categorising government functions, developed by the OECD and published by the UN Statistics Division. The categories have been adapted from COFOG's expenditure-based taxonomy to reflect the operational contexts in which AI systems are deployed, and are further validated by observed deployment patterns documented in existing AI registers including the Government of Canada AI Register and the UK Algorithmic Transparency Recording Standard.

Third, on accountability, throughout this tool, accountability for AI system behaviour rests with human and institutional actors; the developing agency, the deploying institution, the contracting authority, or the oversight body and not with the AI system itself. Where a column asks whether something "exists" or "is in place," this should be read as asking whether the responsible institution has established, documented, or ensured that thing. Respondents should identify the relevant actor when completing free-text fields. This note is to ensure we don't risk anthropomorphism of AI systems.

Fourth, audits and performance monitoring serve complementary but distinct accountability functions, and should not be confused. Monitoring is what a responsible agency does continuously to ensure the system keeps working as intended. Auditing is what happens when someone, internal or external, formally evaluates whether the system is working fairly, legally, and/or effectively. A system can have robust internal monitoring with no external audit, or a one-off external audit with no ongoing monitoring in place. Both fields should be filled independently.

And lastly, a note on human rights and impact assessments. Other impact assessments (such as ethics, privacy, fundamental rights) are considered into this category. If someone has done a Data Protection Impact Assessment, the identification of adverse impacts on vulnerable groups is expected, which is why they can be subsumed in the 'Human Rights & Impact Assessment' categories. We consider that if detailed human rights impact assessments have taken place, they will already cover various aspects of human rights, and hence, they have not been individually listed here, to reduce duplication.

Table 2: AI in the Public Sector Transparency Tool[4]

S.N.	Category	Field name	Field description
1.	**Administrative Note**	Date of review	Date or week when this template is being populated.
2.	**System Id**	Country	Name of country where this AI-enabled service is deployed.
3.	**System Id**	Administrative level	Sub-national level, if applicable. *Example: Federal, state, city, and municipal*

[4] The highlighted fields reflect supplementary information.

4.	**System Id**	Name of the AI-Enabled public service[5]	Descriptive name of the AI-enabled public service or automated process. *Example: 'Automated welfare eligibility scoring', and not 'IBM Watson'.*
5.	**System Id**	AI system type	What kind of an AI system is it? *Categories: i) Generative AI (LLMs, multimodal, text to image etc.); ii) Predictive & scoring systems (mainly machine learning-driven); iii) Decision support & rules-based systems; iv) Computer vision tools; v) Natural Language Processing systems; vi) Others; vii) Unclear*
6.	**System Id**	Public register URL	Direct link to this entry in any government algorithm register or transparency portal.
7.	**System Id**	Government branch	Which government branch is the AI system being deployed in? *Categories: i) Executive; ii) Legislative; iii) Judiciary; iv) Independent body; v) Multiple branches (add specifics); vi) Others (add specifics)*
8.	**System Id**	Implementing institution	Which government agency is primarily accountable for this system's deployment and operation? *Accountability may be distributed across multiple agents, including different Ministries and Statutory Boards. Note all relevant institutions where applicable*
9.	**System Id**	Sector/Area of implementation	Which sector or public service area is the AI system being implemented in? *Categories: i) Immigration, refugees & citizenship; ii) Employment & social development/welfare; iii) Revenue agency/taxation; iv) Law enforcement & criminal justice; v) Health & medical services; vi) Education & skills training; vii) Employment & workforce management; viii) Statistics & national surveys; ix) Scientific research & innovation; x) Natural resources & environment; xi) Foreign affairs & diplomacy; xii) Defense & national security; xiii) Others (add specifics)*
10.	**System Id**	Lifecycle status	What is the current status of the AI system? Categories: i) Planning; ii) Under consultation; iii) Under approval; iv) In development; v) Pilot/Testing; vi) Deployed; vii) Under review; viii) Decommissioned/discontinued
11.	**System Id**	Service start date	Date the AI-enabled service began operating or was piloted with the public.
12.	**Problem Definition**	Intended and stated purpose[6]	What problem does this AI system address? *Describe the specific decision, task, or process being automated or assisted. Examples include welfare eligibility determination, clinical decision support, visa application triage, or tax fraud detection.*

[5] If a public register entry exists, use the name as recorded there. If the system name is unknown, note the initiative it is being used in.

[6] This field should reflect the officially stated purpose of the AI system as documented by the deploying agency and not an independent assessment of what the system actually does in practice. The primary sources for this field should be official government documents such as procurement notices, ministerial announcements, legislative mandates, or agency websites; published contracts or tender documentation that describe the system's intended function; parliamentary records, budget documents, or official press releases where the above are unavailable. Where the official stated purpose is absent or vague, note this in the Key Disclosure Gaps field — the absence of a clear, publicly documented purpose statement is itself a significant accountability gap.

13.	**Problem Definition**	Automation level	Select the level of human involvement. *Categories: i) Fully automated (no human in the loop); ii) AI recommendation - human must act on it; iii) AI recommendation - human may override; iv) Primarily human decision with AI acting as a support or augmenting system; v) Unclear*
14.	**Problem Definition**	Approval authority	Name and title of the official who authorised deployment.
15.	**Problem Definition**	External consultation	Was a public or stakeholder consultation conducted? *Categories: i) Yes - open public consultation; ii) Yes - targeted stakeholder consultation; iii) Yes - consultation with affected communities (incl. indigenous groups if relevant); iv) No - no consultation conducted; v) Not applicable; vi) Unclear*
16.	**Problem Definition**	Link to consultation documentation	URL to published consultation report, public notice, or community engagement record.
17.	**Problem Definition**	Date of consultation	Date of consultation.
18.	**Procurement**	Development or procurement approach[7]	How was the AI system developed and/or procured? *Categories: i) In-house/government-built; ii) Vendor-procured (off-the-shelf); iii) Custom development by vendor; iv) Third-party model or API; v) Hybrid; vi) Unclear*
19.	**Procurement**	Vendor or developer	Name of the vendor and developer, or 'In-house', if government-built. List all, if multiple vendors.
20.	**Procurement**	Procurement method[8]	How did the government procure this AI system? *Categories: i) Competitive tender; ii) Sole source/direct award; iii) Framework agreement; iv) Public-private partnership; v) Open source*

[7] The categories listed for this field are explained below:
i) In-house/government-built: A public agency built the system itself using its own staff and resources
ii) Vendor-procured (off-the-shelf): The government bought an existing commercial product — a system that was already built and is sold to multiple clients.
iii) Custom development by vendor: A vendor was contracted to build a system specifically for the government's requirements, typically through a formal tender.
iv) Third-party model or API: The government is using an AI capability that is owned and operated by an external party — for example, plugging into a foundation model via an API (like using GPT-4 or a similar model for a citizen-facing service).
v) Hybrid: A combination of the above. for instance, a government-built interface sitting on top of a third-party model, or an in-house team working alongside a vendor.
vi) Unclear: The procurement route is not publicly documented or cannot be determined from available sources.

[8] The categories listed for this field are explained below:
i) Competitive tender: The government publicly advertised the contract, multiple vendors submitted bids, and the contract was awarded based on defined criteria (price, capability, track record, etc.).
ii) Sole source/direct award: The government selected a specific vendor without a competitive process.
iii) Framework agreement: The government is purchasing from a pre-approved list of vendors that was itself established through a competitive process at an earlier point in time.
iv) Public-private partnership: The AI system is being delivered through a formal long-term arrangement in which a private entity shares in both the investment and the operation of the system; often in exchange for a revenue share, data access, or other benefits.

			adoption; vi) Other, vii) Unclear
21.	**Procurement**	Contract Value	Total contract value in local currency.
22.	**Procurement**	Link to procurement documentation	URL to tender notice, contract award, or any publicly available procurement documentation.
23.	**Procurement**	Government Audit rights	What auditing rights does the government have over the system? *Categories: i) Yes, contractual right to audit vendor's system, code, or data; ii) Yes, statutory audit body oversight; iii) No; iv) Unclear*
24.	**Data**	Public data use	Are public databases used to train, adapt, or operate the system? *Categories: i) Yes (list which databases in the next field); ii) No; iii) Unclear*
25.	**Data**	Databases used[9]	List specific public databases used for training, adaptation, or real-time operation.
26.	**Data**	Personal data used	Are personal databases used to train, adapt, or operate the system? *Categories: i) Yes; ii) No; iii) Unclear*
27.	**Data**	Legal Basis for personal data use	What is the legal basis on which personal data was used to train, adapt, or operate the system? *Categories: i) Data Protection Law; ii) Another law (cite the specific law or regulation or set of laws); iii) No legal basis; iv) Others (add any comments); v) Unclear*
28.	**Data**	Data retention period	How long is personal data retained? *Specify in months/years or mark as 'not disclosed'*
29.	**Model and Performance**	Model card	Is a 'Model Card' or 'Technical Specification' publicly available? *Categories: i) Yes; ii) No; iii) Partially available; iv) Unclear*
30.	**Model and Performance**	Link to model card	URL to published model card, technical datasheet, or equivalent.

v) Open source adoption: The government is using a publicly available, openly licensed AI model or tool; one that anyone can inspect, use, and modify.
vi) Other: The procurement route does not fit any of the above categories, or involves a combination of mechanisms that cannot be neatly classified.
vii) Unclear The procurement method is not publicly documented or cannot be determined from available sources.

[9] Consult the applicable legal framework in your country to determine whether the data used by this system qualifies as personal data (data that identifies or can identify an individual) or public data (data that is non-identifiable and publicly available by law or policy).

31.	**Model and Performance**	Decision-making logic	Is the logic and rationale of the AI system's decision logged and traceable? *Categories: i) Yes, explanations provided to affected individuals; ii) Yes; explanations available to operators only; iii) No; system outputs are not explainable; iv) Unclear*
32.	**Deployment**	AI infrastructure type[10]	Where is the inference infrastructure of this AI system located? *Categories: i) On-premises (government-owned); ii) Cloud (commercial provider); iii) Cloud (government cloud); iv) Hybrid; v) Unclear*
33.	**Deployment**	AI disclosure format[11]	Does the deploying agency notify individuals that an AI system is involved in decisions affecting them? *Categories: i) Yes, all affected individuals are notified that AI is involved in the decision; ii) Yes, some individuals are notified; iii) No; iv) Unclear*
34.	**Human Rights & Impact Assessment**	Human Rights Impact assessment(s)	Was there an human rights impact assessment prior to the deployment of the system? *Categories: i) Yes; ii) No; iii) Unclear*
35.	**Human Rights & Impact Assessment**	Link to impact assessment(s)	URL(s) to published assessment reports. List all links if multiple assessments exist.
36.	**Human Rights & Impact Assessment**	Impact assessment(s) conducted[12]	Select all that apply. *Categories: i) Human Rights Impact Assessment (HRIA); ii) Ethical Impact Assessment (EIA) - e.g. UNESCO framework; iii) Fundamental Rights Impact Assessment (FRIA); iv) Data Protection Impact Assessment (DPIA); v) Algorithmic Impact Assessment (AIA); vi) Others (add specifics); vii) No impact assessment was conducted; viii) Unclear*
37.	**Monitoring**	Performance monitoring process[13]	What is the status of the performance monitoring process? *Categories: i) Existing and ongoing, with results published externally; ii) Existing and ongoing, with results undisclosed to the public; iii) Existing and incident-driven, with results published externally; iv) Existing and incident-driven, with results undisclosed to the public; v) Existing and audit-linked, with results published externally; vi) Existing and audit-linked, with results undisclosed to the public; vii) Does not exist; viii) Unclear*
38.	**Monitoring**	Link to monitoring results	URL to published performance monitoring reports or dashboards.

[10] This field refers largely to inference infrastructure; the computing environment in which the AI system processes data and generates outputs during live operation. Inference infrastructure is what determines where citizens' data is directed and processed at the point of decision-making, and is therefore the most relevant indicator for data sovereignty, jurisdictional accountability, and ongoing operational risk.

[11] This is notification of AI involvement in decision-making, not data processing consent. The obligation to notify lies with the deploying agency or responsible institution; not the system itself.

[12] For reference, CoE HUDERIA methodology is available here; UNESCO EIA is available here; UNDP HRIA toolkit is available here; and the EU FRIA guide from Denmark is available here.

[13] Performance monitoring refers to the ongoing, operational tracking of how an AI system is functioning after it has been deployed. This includes regularly checking whether the system continues to perform accurately, whether its outputs remain consistent over time, and whether any drift, errors, or disparate impacts across population groups are emerging.This is distinct from a formal audit. If documentation is not available, select 'unclear' rather than 'does not exist'.

39.	**Audits**	Audit frequency[14]	Select the option that best reflects the audit frequency that applies specifically to the AI system being documented — not the broader policy, programme, or institution within which it operates. *Categories: i) No audits conducted; ii) One-off audit; iii) Annual; iv) More frequently; v) Unclear* *[An AI Audit is a formal evaluation of an AI system conducted at a specific point in time, typically by a party with some independence from the day-to-day operation of the system.]*
40.	**Audits**	Types of audit	Describe the audit in question. *Categories: i) No audits; ii) Internal audit, results published; iii) Internal audit, results confidential; iv) Independent third-party audit, results published; v) Independent third-party audit, results confidential; vi) Parliamentary/legislative oversight; vii) Unclear*
41.	**Audits**	Link to most recent audit	URL to the most recent published audit. Leave blank if audits are confidential or not conducted.
42.	**Incident reporting**	Reporting of incident(s)[15]	Were there any incidents reported during the operation of the system? *Categories: i) Yes, reported publicly (add link below); ii) Yes, internal reporting only; iii) No; iv) Unclear*
43.	**Incident reporting**	Reporting of Human Rights-Impacting incidents	Were Human Rights-Impacting Incidents Reported? *Categories: i) Yes, reported publicly (add link below); ii) Yes, internal reporting only; iii) No; iv) Unclear*
44.	**Incident reporting**	Link to incident documentation	URL to incident logs, case reports, or documentation of any administrative, legal, or judicial challenges to this deployment.
45.	**Redressal mechanism**	Complaint lodging process	What is the mechanism to lodge a complaint for harm caused by an AI system in place? *Categories: i) Yes, independent administrative review; ii) Yes, judicial review available; iii) Yes, internal review only; iv) Yes, Ombudsman exists; v) No; vi) Unclear*
46.	**Redressal mechanism**	Link to documentation	Describe the appeal process and share the URL to published documentation of the appeal or review process, if available.

[14] In many cases, a government policy or programme may be subject to general institutional audits (e.g. annual parliamentary or financial audits of a ministry) that do not specifically evaluate the AI system itself. These should not be recorded here unless the audit explicitly covered the AI system's design, performance, data use, or decision-making logic. When completing this field from publicly available sources, it is expected that 'Unclear' will be the most accurate answer in many cases. Selecting a frequency option such as 'One-off audit' or 'Annual' should only be done where there is documented evidence that an audit of the system specifically took place at that frequency. Where this is an inference or assumption based on indirect evidence, select Not known and note the basis for the assumption in the free-text fields. The tool prioritises accuracy over completeness.

[15] Incidents include human rights-impacting events (e.g. wrongful exclusions, discriminatory decisions, privacy breaches), and are NOT limited to technical safety incidents.

47.	**Compliance**	Applicable data protection legislation	Name the applicable data protection law (e.g. GDPR, Council of Europe Convention 108+, national law). Enter 'None applicable' or 'No framework exists' if appropriate.
48.	**Compliance**	Applicable AI-specific legislation	What is the state of the applicable AI-specific legislation? *Categories: i) National AI Legislation (add specifics); ii) No AI-specific legislation applies; iii) In development; iv) Unclear*
49.	**Compliance**	Independent oversight[16]	Name of the independent oversight body for AI. Enter 'None' if none exists.
50.	**Disclosure Notes**	Key disclosure gaps	What information about this system is not publicly available? Note significant gaps for accountability purposes.

Recommendations

To complement the Tool, we also provide below a set of recommendations for ecosystem stakeholders on how to use it for different purposes.

Table 3: Recommendations according to stakeholder group

Stakeholder group	Recommendations
Academia and CSOs	• Leverage the framework to strengthen independent oversight over the governance of AI systems in public service and their potential rights-impacting consequences • Build collective efforts to push for transparency and accountability regarding the use of AI in the public sector, as well as to demand global mechanisms that allow an oversight into existing systems and their impacts to human rights • Build capacity in public participation of local communities, and their ability to contest and seek remediation in case of harm by AI systems • Explore the proposed framework and share methodologies and instruments that facilitate public access on the impact of AI in the public sector • Advocate for reducing the barriers to informational access faced by CSOs, grassroots organizations, and researchers working on topics of AI ethics, safety, and trust • Enable global dialogues with other CSOs in multiple territories where the framework is applied, and advocate for policies and guardrails to protect against rights-impacting technologies
Governments, policymakers, and judicial authorities	• Establish transparent institutional governance mechanisms in line with global practices to protect human rights, democratic values, and the rule of law

[16] Name of body providing oversight beyond the implementing agency (e.g. data protection authority, parliamentary committee, audit office, ombudsman, human rights office, AI boards).

	• Translate “core” parameters of the framework into different legal and policy instruments (including laws, procurement guidelines, Freedom of Information mandates, etc.) to enable information disclosures and ease of access for the public • Formulate procurement guidance that effectively serves as a regulatory tool to prevent the adoption of systems that are scientifically questionable, discriminatory or that fail to comply with international human rights standards • Improve due processes related to the adoption of AI and launching national and/or subnational repositories with key public interest information regarding such processes and the systems adopted as a first step towards promoting accountability of such adoption • Develop participatory processes involving CSOs, grassroots movements and affected communities for decision-making regarding the adoption of AI systems • Implement participatory impact assessments prior to the decision on adopting a system in order to inform it in line with Recommendations 50 to 53 of UNESCO Recommendations on the Ethics of AI and in order to comply with existing obligations regarding necessity and proportionality of interventions that may restrict human rights
Multilateral organizations	• Sustain civil society and academic meaningful participation in global AI Governance discussions and decision-making • Popularize and support initiatives for the development of collaborative research like the ones made possible within the subgroups of the UNESCO CSO and Academic Network on AI Ethics and Policy • Consider adopting core elements of the proposed framework as part of specific recommendations for States to pro-actively disclose public interest information regarding their deployment of AI systems • Promote and support academic and CSO-led initiatives in the Global South to promote and maintain public interest open repositories of AI systems used in the public sector, as well as other public interest initiatives related to the increase of transparency and accountability in this regard

Funding Acknowledgement

The presented work was funded by the Patrick J. McGovern Foundation through the Scaling Inclusive, Human Rights-Based Ethical AI Governance (2025-26) project.



Patrick J McGovern
FOUNDATION

In support of



unesco