Introduction

The CAI Secretariat has prepared a draft methodology that provides clear, concrete and objective criteria for identifying contexts and applications in which the deployment of artificial intelligence (AI) systems (or combinations of such systems) would likely pose significant levels of risk to the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. The methodology will ensure a uniform approach towards the identification, analysis, and evaluation of these risks and the assessment of impact of such systems in relation to the enjoyment of human rights, the functioning of democracy and the observance of rule of law.

The methodology is based on the assumption that domestic authorities are better placed to make relevant policy and regulatory choices, taking into account their country's specific political, economic, social, cultural, and technological contexts, and that they should accordingly enjoy a certain margin of appreciation in this sphere. On this view, the role of the methodology is to assist domestic authorities in establishing the procedures and mechanisms needed to identify such contexts and roles in which artificial intelligence systems, or combined technologies based on such systems, are likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, and help them manage the related risks. In supporting the latter goal, the methodology should assist domestic authorities by specifying the procedural mechanisms needed to ensure that adequate risk analysis, impact assessment, impact mitigation, access to remedy and system monitoring protocols are put in place by anyone building, procuring or using such systems or combined technologies based on them for use in areas considered as sensitive for human rights, democracy and the rule of law.

Another equally important aspect of the methodology is to ensure seamless compatibility of this approach with the existing assessment, governance, and compliance practices followed by industry.

General approach

The methodology is designed to be as "algorithm neutral" and practice-based as possible so that it can remain maximally future proof and inclusive of various AI applications. The model will also need to stay responsive to the development of novel artificial intelligence innovations and use-cases, and should be seen as dynamic and in need of regular revision. Hence, the choice of putting it in the Annex to the future legal instrument, which should facilitate the revision of the methodology in the future.

Key elements processes of the methodology

The methodology will likely comprise a number of clearly articulated and interrelated processes and instruments/steps:

A. A context-based risk analysis (COBRA) provides an initial indication of the context-based risks that AI systems or combinations of such systems could pose to human rights, democracy and the rule of law in view of the contexts of their deployment and application as well as other relevant circumstantial risk factors. The main purpose of a COBRA is to identify the extent to which, if at all, an AI system is likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, in view, in particular, of the context of its deployment. The COBRA also includes a risk calibration mechanism that integrates variables of the scale, scope, and likelihood of potential harms to help Parties establish a proportionate approach both to subsequent elements of the methodology and to the level of stakeholder engagement that is needed throughout the project lifecycle, more generally.

B. Provided that an AI system is to be deployed in the relevant contexts highlighted in the COBRA, at the next stage, the methodology requires some form of stakeholder engagement process to help project teams identify salient stakeholders and to facilitate proportionate stakeholder involvement and input throughout the project workflow.

   It is important to underline that this process aims to improve the quality of the risk analysis, impact assessment, impact mitigation planning, and determination of access to effective remedy by amplifying the perspective of the actors whose rights and interests could be potentially at stake. In any event, the level of such engagement and involvement should be proportionate to the level of risk and other relevant factors. The methodology defines certain basic principles and procedures whilst remaining flexible. In the end, it would be up to the Parties to define the exact modalities of this part of the process.

C. The core of the methodology is the actual Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA). The HUDERIA consists essentially of an obligation to address a certain number of questions regarding the contexts of design, development, procurement, and use and the potential short-, medium, and long-term impacts of the AI system under examination. These questions are designed to facilitate considerations that are specific to human rights, democracy, and the rule of law.

   The process of answering these questions, with the support of proportionate stakeholder engagement,

   a) contextualises and corroborates the potential adverse effects which have been previously identified in the COBRA,
   b) enables the discovery of further harms through the integration of stakeholder perspectives,
   c) makes possible the collaborative assessment of the severity of potential adverse impacts identified,
   d) facilitates the co-design of an impact mitigation plan,
   e) sets up access to remedy, and
   f) establishes monitoring and impact re-assessment protocols.

D. The methodology requires the implementation of an impact mitigation plan as well as, if appropriate, mechanisms providing access to remedies.

E. Lastly, the methodology contains a requirement that the carrying out of such impact assessments and mitigation procedures shall have an iterative and dynamic character. This means that assessments should be repeatable at regular intervals across the AI lifecycle in question until the system is retired or decommissioned to ensure that relevant intervening changes in both the context and the system itself are properly identified, understood and dealt with.

A: COBRA (context-based risk analysis)

The main purpose of a COBRA is to identify the extent to which, if at all, an AI system, or a combination of such systems, in view of the context(s) of its deployment and application and other relevant risk factors, is likely to pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. A COBRA helps Parties establish a proportionate approach to subsequent elements of the methodology and to the level of stakeholder engagement that is needed throughout the project lifecycle.

It is up to the Parties to define the exact modalities of this part of the process on the basis of the indicative list of rights and freedoms (contained in Annex 1) and the indicative list of risk factors and basic parameters of the risk calibration mechanism that are referred to within this section, provided that the following basic approach regarding the understanding of “significant levels of risk” is respected.

The deployment and application of AI systems or combinations of such systems may pose significant levels of risk to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, provided that the outputs of such systems are employed in one of the sectors/domains (listed in Annex 2), with a view to:

(a) substantially informing the process of establishing or revising laws and policies,

(b) substantially informing or taking decisions affecting human rights and fundamental freedoms or connected legal rights or interests of individuals or legal persons, or

(c) substantially informing or influencing decisions, behaviours, or actions taken by proxy through the deployment of such systems (i.e. where such systems carry out cognitive or perceptual functions, in the place of humans, like reasoning, communication, prediction, planning, classification, problem solving, pattern recognition, or kinematic, visual, auditory, or haptic analysis) that affect human rights and fundamental freedoms or connected legal rights or interests of individuals or legal persons.

General approach to identifying risk factors:

A central aim of a COBRA is to help Parties identify risk factors that indicate the potential harms to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law that may result from the design, development, and deployment of an AI system, or combinations of such systems.

The term ‘risk factor’ is used here to refer to the antecedent characteristics or properties of an AI innovation context that are associated with a higher likelihood of some outcome (or outcomes) that negatively impact human rights, democracy, and the rule of law. For example, inadequate technical skills in an AI project team, limited project resources, or poor governance protocols are all risk factors that increase the likelihood of developing or deploying an unsafe AI system that may harm fundamental rights and freedoms.

Risk factors may be classified as either circumstantial or modifiable. Circumstantial risk factors emerge externally from the technical, sociotechnical, historical, legal, economic, or political environments in which the design, development, and deployment of AI systems are undertaken and that are thereby less controllable. Modifiable risk factors emerge internally from the actual practices of producing and using AI technologies, and that are thus more controllable. The distinction between circumstantial and modifiable risk factors is important for all downstream elements of the methodology, because the identification of modifiable risk factors should, as part of subsequent risk management processes, trigger governance interventions that enable more responsible assurance practices and allow for the avoidance of potentially harmful impacts, whereas the identification of circumstantial risk factors, which are less controllable, should trigger proportionate impact mitigation measures.

The following non-exclusive list contains high level descriptions of various types of modifiable and circumstantial risk factors to help the Parties in the elaboration of the modalities of their respective risk analyses. These factors are not necessarily to be treated as causes of adverse impacts but rather as conditions that are correlated with an increased chance of harm and that need to be anticipated and considered in risk management and impact mitigation efforts.

The list focuses on the spectrum of risk factor types that surround the practical contexts of designing, developing and deploying AI systems as well as the rights and freedoms context surrounding the production and use of such systems rather than on the technical details underlying their specifications. For this reason, the list is organised into two categories: (1) Risk factors arising in the practical context of the AI project lifecycle (including the application context in which the system is conceived and built, the data lifecycle context, the project design context, the model development context, and the system deployment context); and (2) Risk factors arising in the ways that the production and use of such systems specifically impact human rights, fundamental freedoms, democracy, and the rule of law.

1. **Risk factors arising in the practical context of the AI project lifecycle**

   a. Application context:

      i. Sector or domain in which the system is being built, in particular, where the system will serve primary or critical functions in high impact, safety critical, or historically highly regulated sectors or domains or those sensitive to human rights, democracy and rule of law;

      ii. Existing law and regulatory environment of the sector or domain, in particular, where the legal basis or lawfulness of the application must be established or where the system could be repurposed or used in ways that are prohibited under existing statute and regulation;

      iii. The scope of deployment (numbers of rights-holders affected), in particular, where, in the event that the AI system optimally scales, it will directly or indirectly affect large portions of local, national, or global populations or large numbers of rights-holders within or across local, national, or global populations;

      iv. The scope of deployment (breadth and temporality), in particular, where the direct or indirect impacts of the use of the AI system could affect rights-holder, communities, or the environment at a timescale that Parties deem significant (e.g. long-term, generational, or intergenerational impacts);

      v. Technological maturity, in particular, where the AI system's design is not wholly based on well-understood techniques that have previously been in operation and externally validated for a similar purpose and in the same sector;

      vi. Existing system (human or technological) that the application is replacing, in particular, where the AI system is replacing a human, technical, or hybrid system that serves the same or similar function, because the existing system is considered flawed or harmful;

      vii. Existing legacies of bias and discrimination in the sector or domain context, in particular, where the sector(s) or domain(s) in which the AI system will operate, and from which the data used to train it are drawn, contain historical legacies and patterns of discrimination or unfair treatment of minority, marginalised or otherwise disadvantaged groups that could be replicated or augmented in the functioning of the system or in its outputs and short- and long-term impacts;

      viii. Environmental context, in particular, where the AI system could have significant adverse impacts on the environment and transparently reported measures are not in place to ensure that the system, in both its production and use, complies with applicable environmental protection standards and supports the sustainability of the planet;

      ix. Cybersecurity context, in particular, where

         1. the AI system could present motivations or opportunities for malicious parties to hack or corrupt it to achieve substantial financial gains, political goals, or other perceived benefits,
         2. cybersecurity measures that are put in place to safeguard the system’s safety, security, and robustness are not appropriately proportional to potential risks of hacking, adversarial attack, data poisoning, model inversion, or other cybersecurity threats, or
         3. measures are not in place to stress test the system for cybersecurity vulnerabilities and resilience.

   b. Data lifecycle context:

i. Data quality, integrity, provenance, and protection, in particular, where detailed and transparently reported processes are not in place to ensure that data used in producing, tuning, operating, or re-training the AI system, are

1. representative, accurately measured, reliable, relevant, appropriate, up-to-date, and of adequate quantity and quality for the use case, domain, function, and purpose of the system,
2. attributable, consistent, complete, and contemporaneous with collection,
3. properly recorded, traceable, and auditable in terms of their provenance and lineage, and
4. consistent with data protection laws, where personal or sensitive data is collected or procured;

ii. Data types, in particular, where special considerations about responsible data management are necessitated by the use of

1. dynamic data, collected and processed in real time, for continuous learning,
2. data that could be subject to rapid or unexpected shifts or drifts in underlying data distributions which adversely impact the accuracy and performance of the AI system, or
3. unstructured data, or a combination of structured and unstructured data, for AI systems that process social or demographic data and that consequently pose risks of algorithmic bias and lurking discriminatory inferences;

iii. Dataset linkage, in particular, where personal or sensitive data is collected or procured and there is a possibility for deanonymizing or identifying data subjects through data linkage with existing data, publicly available datasets, or data that could be easily obtained by malevolent third parties;

iv. Data labelling and annotating practices, in particular, where social or cultural biases could influence the way data labellers or annotators label, categorize, or classify data or where automated labelling or annotation could import or replicate historical patterns of discrimination and social or cultural bias.

c. Project design context

i. Decision to design, in particular, where transparent processes are not in place to ensure that the decision to build the AI system is appropriate given available resources and data, existing technologies and processes, the complexity of the use-contexts involved, and the nature of the policy or social problem that needs to be solved;

ii. Definition of problem and outcome, in particular, where input from stakeholder engagement and public scrutiny is incorporated neither into processes of selecting and formulating the problem to be solved by the AI system and nor into processes of defining its target variable (or measurable proxy);

d. Model development context:

i. Pre-processing and feature engineering, in particular, where data pre-processing and feature engineering involves the grouping, disaggregating, or excluding of input features related to protected or potentially sensitive characteristics (e.g. decisions about combining or separating sub-categories of gender or ethnic groups) or proxies for these;

ii. AI model characteristics and model selection, in particular, where

1. algorithmic model(s) or technique(s) to be used by the AI system have a non-deterministic, probabilistic, evolving, or dynamic character that prevents or hinders the system's intended functionality from being formalized into specific and checkable design-time specifications (or that impairs commonly accepted methods of formal verification and validation), or
2. the algorithmic model(s) or technique(s) to be used by the AI system have a complex, high-dimensional, or non-linear character that impairs or prevents the interpretability and explainability of the system—especially if the system is (a) to operate in a safety critical or high impact sector, (b) to process social or demographic data that my contain discriminatory bias, or (c) to operate in a sector or domain in which reasonable expectations of intelligibility and accessibility accompany the system's function (e.g. a diagnostic AI model used in the health sector will be accompanied by the expectation that its outputs inform evidence-based clinical decision making and are hence explainable);

iii. Model inference, in particular, where inferences generated from the model's learning mechanisms could contain discriminatory correlations or influences of hidden proxies for protected or sensitive characteristics that may act as discriminatory factors in the generation of its output;

iv. Model verification and validation, in particular, where transparently reported processes of external peer review and evaluation by independent domain and technical experts are not in place as part of the verification and validation of the AI model;

v. Model accuracy and performance metrics, in particular, where the model could perform differentially for affected sub-populations (e.g. a model that has disproportionately higher error rates for protected or disadvantaged groups) and transparently reported processes are not in place to test the system for differential performance.

e. Model deployment context:

i. System-implementer/user interface, in particular, where the deployment of the system could harm the physical, psychological, or moral integrity of implementers or adversely impact their dignity, autonomy, and ability to make free, independent, and well-informed judgements;

ii. Level of automation/level of human involvement and choice in system outcomes, in particular, where processes are in place neither to ensure the competent involvement of human implementers or users in respect to the responsible deployment of the system nor to train implementers or users to fully understand

1. the strengths and limitation of the system and its outputs
2. the potential conditions of situational complexity, uncertainty, anomaly, or system failure that may dictate the need for the exercise of human judgment, common sense, and practical intervention.

2. **Risk factors arising in the ways that the production and use of AI systems impact human rights, democracy, and the rule of law**

a. Potential adverse impacts that the design, development and deployment of the AI system could have on each of the rights and freedoms included in the indicative list contained in Annex 1 given the system’s intended purpose and the contexts in which it will be used:

   i. Potential adverse impacts on the rights and freedoms of AI system designers and developers across the AI project lifecycle, in particular, where the activities of data collection, procurement, labelling, or annotation, project design, and model development could harm the rights and freedoms of any individual involved in the production of the system;
   ii. Potential adverse impacts on the rights and freedoms of AI system implementers or users, in particular, where activities of implementing or using the system could harm the rights and freedoms of any indivdual involved in its deployment;
   iii. Potential adverse impacts on the rights and freedoms of affected rights-holders, in particular, where (a) the collection, procurement, labelling, or annotation of the data used to train, test, validate, and operate the AI system or (b) the deployment of the system and the conveyance and use of its outputs could harm the rights and freedoms of any impacted rights-holder;
   iv. Potential adverse impacts on the rights and freedoms of affected rights-holders in the event of system malfunction, misuse, or abuse, in particular, where the breakdown or failure of the system, the use of the system out-of-the-scope of its intended purpose, or its malicious misapplication could harm the rights and freedoms of any impacted rights-holder.

b. Potential adverse impacts that the design, development and deployment of the AI system could have on the functioning of democracy, and the observance of the rule of law:
   i. Potential adverse impacts on the functioning of democracy, in particular, where the use of an AI system, or a combination of such systems, plays a role in substantially influencing or informing the democractic processes listed in Annex 2 (4a-f), or where the use or misuse of an AI system, or a combination of such systems, could lead to interference with free and fair election processes or with the ability of impacted individuals to participate freely, fairly, and fully in the political life of the community through,
      1. Mass deception, misinformation, or disinformation at local, national, or global levels caused by the deployment of the system
      2. Mass manipulation, at local, national, or global levels enabled by the deployment of the system
      3. Mass intimidation or behavioural control, at local, national, or global levels enabled by the deployment of the system
      4. Mass personalized political targeting or profiling, at local, national, or global levels enabled by the deployment of the system;
   ii. Potential adverse impacts on the observance of the rule of law, in particular, where the use of an AI system, or a combination of such systems, plays a role in substantially influencing or informing the process of decision-making in establishing or revising laws and policies in the domains/sectors included in 3a-b (Administration of Justice: Institutional aspects of organisation of the judiciary) of Annex 2 or where the deployment of an AI system, or a combination of such systems,

could harm impacted individuals' right to effective remedy or right to a fair trial (equality of arms, right to a natural judge established by law, the right to an independent and impartial tribunal, and respect for the adversarial process);

iii. Potential weakening of accountability of the executive authority for its actions through the undermining of the role of democratic institutions and the judiciary, in particular, where the use on an AI system, or a combination of such systems, brings about changes of rules or procedures related to democratic and juridical institutions and functions that weaken accountability mechanisms within the processes, domains, and sectors listed in 3a-b and 4a-f of Annex 2.

Basic parameters of the risk calibration mechanism:

The present description of the risk calibration mechanism is intended to help Parties determine the procedure according to which organisations that are planning to build or procure an AI system can establish a proportionate approach to project governance activities and stakeholder engagement. It sets out the elements needed to index the risk level of each of the potential harms to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law that have been identified in the COBRA but does not dictate how to combine these through a risk calibration formula or calculus. The final determination of this formula, whether qualitative or quantitative or a combination of these, is left to the discretion of the Parties.

Calculating the risk level of the potential harms to human rights, democracy, and the rule of law that may result from the design, development, and deployment of an AI system, or a combination of such systems, involves detailed analysis of each of these identified harms. Following the language of the United Nations Guiding Principles on Business and Human Rights and ISO/IEC/IEEE 16085:2021, the variables that would need to be part of the calculus are:

1. The scale or gravity of the potential harm (i.e. the seriousness of the potential harm's expected consequence);
2. The scope of the potential harm (both the number of rights-holders affected and the timeframe of the effects (e.g. short-term, medium-term, generational, intergenerational);
3. The likelihood of the potential harm.

The combination of these variables into a risk calibration mechanism is intended to provide organisations with a preliminary estimation of risk levels so that they can determine initial proportionality recommendations for appropriate risk management and assurance practices and stakeholder engagement. The potential harms and adverse impacts that have been distinguished in the COBRA, and that are analysed through the risk calibration mechanism, are then to be properly re-visited and re-evaluated in a contextually sensitive way (and with stakeholder input) in the stakeholder engagement process and in the HUDERIA.

B: Stakeholder engagement process

The purpose of the stakeholder engagement process (SEP) is to identify stakeholder salience and to facilitate proportionate rights-holder involvement and input throughout the project workflow. A diligent SEP is essential for ensuring that rights-holders' views are appropriately incorporated in the assessment and governance of the project and that any potential risks of adverse impacts are identified and mitigated across the system's lifecycle.

It is up to the Parties to define the exact modalities of this part of the process, provided that the following basic principles are respected:

1. Stakeholder engagement may take various forms, but the exact level of rights-holder participation should be proportionate to risks identified in the COBRA and other relevant factors;
2. The process should involve five key activities:
    a. stakeholder analysis (identification of stakeholder groups who may be affected by, or may affect, the design, development, and deployment of the system; assessment of the relative interests, rights, vulnerabilities, and advantages of identified stakeholders; analysis of the salience of identified stakeholder groups, with a view to the meaningful inclusion of rights-holders who (1) are disproportionately at risk from the use of the system, (2) are vulnerable to potential harms, and (3) have limited ability to influence project outcomes, e.g. historically marginalised, disadvantaged, or underrepresented groups);
    b. positionality reflection (reflection on the positional standpoint of project team members vis-à-vis affected stakeholders with a view to recognizing the limitations of team members' perspectives and identifying missing stakeholder viewpoints that would strengthen assessment of the system's potential impacts; this includes assessment of team members' self-ascribed identity and demographic characteristics, education and training, socioeconomic status and history and institutional and team context);
    c. establishment of engagement objectives (establishing clear and explicit stakeholder participation goal(s), which ensure the inclusive, informed, and meaningful involvement of affected rights-holders);
    d. determination of engagement method (evaluation and accommodation of rights-holder needs, taking into consideration other relevant factors such as resources, capacities, timeframes etc.).
    e. Initiation and implementation of proportionate engagement processes consistent with the results of the stakeholder analysis, positionality reflection, and established engagement objectives and methods.
3. The SEP is subject to the iterative requirements listed in section E below, as it is crucial to revisit and revise the results of the SEP and of other project governance activities to ensure that approaches continue to reflect the perspectives and interests of relevant stakeholders and that stakeholder input remains sufficiently responsive across the project lifecycle to any changes occurring both in the production and implementation contexts of the system and in the real-world environments in which it is embedded.

The SEP shall result in an iteratively updated project report containing the summary of the findings of the SEP and documentation of other project governance activities (e.g. the HUDERIA) as well as actions taken concerning impact mitigation, risk management, and assurance measures.

C: Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA)

The purpose of the Human Rights, Democracy and the Rule of Law Impact Assessment is to provide detailed evaluations of the potential and actual impacts that the design, development and application of an AI system could have on human rights and fundamental freedoms, democracy, and the rule of law. With the support of proportionate stakeholder engagement, this process contextualises and corroborates potential adverse effects identified at the previous stages, enables the establishment of an impact mitigation plan, and sets up access to remedy.

It is up to the Parties to define the exact modalities of the HUDERIA, provided that the following basic principles are respected:

1. The HUDERIA should re-examine and re-evaluate the potential harms to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law previously identified in the COBRA, exploring in more detail how the right, freedom, or dimension of democracy and the rule of law under consideration could be adversely impacted. This enables more open-ended and flexible deliberation on the range of potential adverse impacts at the same time as it allows participants in the HUDERIA process to focus on the specific contexts of the impacts to facilitate better, more granular understandings of their scope, scale, and remediability.
2. The HUDERIA should contextualise and corroborate these potential harms in dialogue with stakeholders through the engagement activities determined by the SEP.
3. The HUDERIA should allow for the identification and analysis of further potential harms by enabling project team members to engage in extended reflection and gap analysis and by giving stakeholders the chance to uncover new potential harms that have not yet been explored and to pinpoint deficits in the completeness and comprehensiveness of the previously enumerated harms.
4. The HUDERIA should allow for the exploration, with appropriate stakeholder participation, of the severity (scope, scale, and remediability) of the potential adverse impacts, so that the risks of these can be better assessed, prioritised, managed, and mitigated.

The HUDERIA process takes into account the information collected across the other elements of the methodology, resulting in a more extensive, detailed, and iteratively updated impact assessment summary of its findings that is included in the project report alongside the SEP findings and other documentation of project governance activities. The HUDERIA also facilitates the drawing up of an impact mitigation plan and ultimately results in the adoption of necessary mitigation measures which should address specific risks that have been identified.

D: Impact mitigation plan and access to remedies

Once potential adverse impacts have been mapped out and organised, and mitigation actions have been considered, an impact mitigation plan should be drawn up. It will be the part of the HUDERIA that specifies the actions and processes needed to address adverse impacts and, as such, will serve a crucial documenting function.

The impact mitigation plan should, as a general rule, include:

(a) a summary of combined impact findings;
(b) an assessment of the severity (scale, scope, and remediability) of the potential adverse impacts that the HUDERIA has identified, which is completed with input from impacted stakeholders;
(c) a clear presentation of the measures and actions that will be taken to address the potential adverse effects;
(d) a clarification of the roles and responsibilities of the various actors involved in impact mitigation, management and monitoring;
(e) a plan for monitoring impact mitigation efforts and for re-assessing and re-evaluating the HUDERIA during subsequent development and deployment phases of the project lifecycle;

(f) an accessible presentation of access to remedy mechanisms that will be available to impacted rights-holders.

As a general rule, while impact prevention and mitigation planning may involve prioritisation of actions, all potential adverse impacts on human rights, democracy and the rule of law must be effectively addressed.

E: Iterative requirements

Carrying out HUDERIA at the beginning of AI system lifecycle is only a first, albeit critical, step in a much longer, end-to-end process of responsible evaluation and re-assessment. In the impact assessment process continuous attention should be paid both to the dynamic and changing character of the AI production and implementation lifecycle and to the shifting conditions of the real-world environments in which systems will be embedded.

It is up to the Parties to define the exact modalities of this part of the process, provided that:

(a) continued revisitation of HUDERIA plays a pivotal role in its continued efficacy and reliability;
(b) a plan is established (e.g. as part of the impact mitigation component of the HUDERIA) for monitoring impact and impact mitigation efforts and for re-assessing and re-evaluating the HUDERIA during each phase of the project lifecycle up to system retirement or decommissioning;
(c) such processes remain as responsive as possible to the way the AI system is interacting with its operating environments and with impacted rights-holders;
(d) in rapidly evolving or changing contexts, there may be a need for more frequent re-assessment and re-evaluation interventions.

Annex 1: Indicative list of rights and freedoms:

	Human rights and fundamental freedoms	The European Convention on Human Rights	The EU Charter	The International Covenant on Civil and Political Rights[1]
1	Right to human dignity	Various provisions	Article 1	Article 16
2	Right to protection of life	Article 2	Article 2	Article 6
3	Prohibition of torture and inhuman and degrading treatment	Article 3	Articles 3 and 4	Articles 7 and 10
4	Prohibition of slavery and forced labour	Article 4	Article 5	Article 8
5	Right to liberty and security	Article 5	Article 6	Articles 9 and 11
6	Right to fair proceedings	Article 6	Article 47	Article 14
7	Right to tribunal established by law	Article 6	Article 47	Article 14
8	Right of access to court	Article 6	Article 47	Article 14
9	Right to an independent and impartial tribunal established by law	Article 6	Article 47	Article 14
10	Right to public pronouncement of a judgment	Article 6	Article 48	Article 14
11	Right to court proceedings within reasonable time	Article 6	Article 47	Article 14
12	Right to presumption of innocence	Article 6	Article 48	Article 14
13	Right to respect for private life	Article 8	Articles 7 and 8	Article 17
14	Right to respect for family life	Article 8	Articles 7 and 9	Article 23
15	Right to respect for home	Article 8	Article 7	Article 17
16	Right to respect for correspondence	Article 8	Article 7	Article 17
17	Freedom to hold or not to hold religious beliefs	Article 9	Article 10	Article 18
18	Freedom to practice or not to practice a religion	Article 9	Article 10	Article 18
19	Right to freedom of expression and access to information	Article 10	Article 10; Article 42	Article 20; Article 19
20	Freedom of assembly	Article 11	Article 12	Article 21
21	Freedom of association	Article 11	Article 12	Article 22
22	Right to marry	Article 12	Article 9	Article 23
23	Right to an effective remedy	Article 13	Article 47 Article 41 Article 42	Article 2
24	Prohibition of discrimination	Article 14	Various provisions	Article 3 Article 23 Article 26
25	Right to protection of property	Article 1 of Protocol No. 1	Article 17	
26	Right to education	Article 2 of Protocol No. 1	Article 14	
27	Prohibition of expulsion of nationals	Article 3 of Protocol No. 1	Article 19	
28	Freedom of movement	Article 2 of Protocol No. 4	Article 45	Articles 12 and 13
29	Right to free elections	Article 3 of Protocol No. 1	Articles 39 and 40	Article 25
30	Environmental protection (proxy through other human rights)	Articles 2, 3, 6, 8, 10, 11	Article 37	

[1] Ratified by Canada, Israel, Mexico, the U.S.A., not by Holy See or Japan

Annex 2: The indicative list of sectors/domains:

1. Public administration
    a) Health care, including, but not limited to, such issues as access to healthcare services, diagnostics, prognostics, and preventative care, the provision of life-sustaining treatments, treatment of life-threatening conditions, emergency care services, mental health counselling and treatment, end of life decisions;
    b) Family life and social care, including, but not limited to, such issues as mutual enjoyment of parents with children, custody, access, contract-rights, State care, foster families, access to and provision of public benefits;
    c) Immigration, including, but not limited to, such issues as expulsion, extradition, deportation, adjustments of status, denial of right to entry, notification of rights, translation/interpretation services, production of transcripts, collection and assessment of evidence, conditions and modalities of entrance to and removal from the territory of the State;
    d) Infrastructure development and maintenance, including, but not limited to, such issues as health security, and enjoyment of public space, management of environmental hazards, land and urban planning, energy management and energy consumption;
    e) Emergency services, including, but not limited to, such issues as management of rescue operations, emergency communications infrastructures, management of the aftermaths of disasters;
    f) Public education, including, but not limited to, such issues as access to educational institutions assessments, and official recognition of studies.

2. Law enforcement and security
    a) Police, intelligence and assimilated services, including, but not limited to, such issues as the use of lethal force, administration of physical force during arrests, ID checks and identification of individuals for law enforcement purposes, programmes regarding protection of persons in danger (e.g. victims of domestic violence or protected witnesses), arrests and detentions, management of programmes regarding vetting of officials, management of rescue and hostage rescue operations, crowd management during public events, predictive policing, emotion and sentiment analysis, measures entailing entering private home, surveillance of telecommunications, restrictions, bans, prohibitions, lockdowns, supervisions regarding the freedom of movement;
    b) Prosecutions, including, but not limited to, such issues as collection and assessment of evidence.

3. Administration of justice
    a) Criminal/civil/administrative/commercial/constitutional courts justice, including, but not limited to, such issues as arrests, detentions, decisions regarding bail, release on parole, conditional release and wearing of electronic bracelets, notification of rights, translation/interpretation services, production of transcripts, collection and assessment of evidence (including assessment of trustworthiness of witnesses and evidence), granting of legal aid, determination of any criminal charge, determination of civil rights and obligations, decisions regarding challenges of judges or jury members, decisions regarding access to review level of proceedings, criminal sentencing, automated proceedings;
    b) Institutional aspects of organisation of the judiciary, including, but not limited to, such issues as the management of the process of vetting, appointments and dismissal of judges/judicial officers, attribution of cases for processing to specific judges/judicial officers, case management in legal proceedings.

4. Democratic processes
    a) Electoral system, including, but not limited to, such issues as conditions and modalities of the exercise of the right to vote, eligibility age, exclusion rules, conditions and modalities of voting, voting methods and procedures, conditions and modalities of counting, the right to stand in elections, the organisation of elections and referenda, redistricting, the management of electoral disputes and effective remedies in this connection, distribution of electoral information;
    b) Institutions and political processes, including, but not limited to, such issues as the supremacy of the constitution, the role of the judiciary in the balancing of powers, delegation of the legislative function;
    c) Freedom of expression, including, but not limited to, such issues as expression of protected speech in various forms, protection of journalistic sources, information gathering activities, access collection and automated processing of data, research and investigation activities, disclosure regime concerning information received in confidence, protection of whistle-blowers;
    d) Freedom of assembly and association, including, but not limited to, such issues as time, place and manner of conduct of assemblies, conditions and modalities of the right to form or be affiliated with a group or organisation pursuing particular aims, surveillance of assemblies and identification of participants;
    e) Access to information, including, but not limited to, such issues as access to financial information and information about business dealings of individuals, duty to provide reliable and precise information, responsibilities with regard to verification and transmission of information, access to State-held information;
    f) Media, including but, not limited to, such issues as transparency with regard to media ownership, media pluralism, freedom of expression during elections (offline and online), duties and responsibilities of internet news portals automated news generation, mis/disinformation, online content moderation.

5. Prison and probation
    a) Management of prisons and detention facilities, including, but not limited to, such issues as prisoner profiling, psychological screening of potentially vulnerable inmates, management of dangerous prisoners, management of prison population, searches of visitors and inmates, surveillance of communications;
    b) Parole and probation services, including, but not limited to, such as issues as release on parole, conditional release and wearing of electronic bracelets.

6. Essential services offered by private sector, including, but not limited to, such spheres as:
    a) Communications
    b) Education
    c) Biomedical research, life sciences, epidemiology, and health care
    d) Environmental and waste management
    e) Energy management
    f) Urban infrastructure and planning
    g) Manufacturing and industrial automation
    h) Construction and building
    i) Security and public safety
    j) Domotics (smart home technologies)
    k) Human resources and labour management
    l) Finance
    m) Information technology and networks
    n) Vehicle manufacturing and transportation infrastructure.
    o) Agriculture and food supply
    p) Provision of systems procured by local, municipal, or national governments