HUDERIA Methodology and Model

The Committee of Ministers
approved the HUDERIA Model
on 25 February 2026
and the HUDERIA Methodology
on 26 February 2025

Council of Europe

French edition:
HUDERIA – Méthodologie et modèle

The opinions expressed in this work are the responsibility of the author(s) and do not necessarily reflect the official policy of the Council of Europe.

All other correspondence concerning this document should be addressed to the the CDNET Secretariat / Digital Development and Sport
Department of the Council of Europe,
Council of Europe, F-67075 Strasbourg Cedex,
E-mail: secretariat.cdnet@coe.int

Photo : ©Shutterstock

Cover design and layout: Publications and Visual Identity Division (DPIV)

HUDERIA was created by the **Council of Europe Committee on Artificial Intelligence (CAI)** to support **consistent and practical risk and impact assessment of AI systems** from the perspective of human rights, democracy and the rule of law. Building on established human-rights risk concepts, it translates these ideas into guidance tailored to the realities of the AI life cycle, helping public and private actors identify, assess, prevent and mitigate risks across diverse uses of Artificial Intelligence. HUDERIA combines high-level, technology-neutral guidance (the **HUDERIA Methodology**) with implementable, adaptable support materials (the **HUDERIA Model: Context-Based Risk Analysis (COBRA) Resources**), offering tools and scalable guidance that can be adjusted to different contexts, needs and levels of capacity. **As a non-legally binding instrumen**t, for Parties to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, HUDERIA offers a flexible reference that may be used or adapted – wholly or in part – together with other frameworks or tools, to support risk and impact assessment approaches aligned with the Framework Convention.

Contents

[1] The Committee on Artificial Intelligence (CAI) considered the possible inclusion of a COBRA Resource D; however, without prejudice to any future decision on its inclusion, the HUDERIA Model: Context-Based Risk Analysis Resources adopted on 5 November 2025 comprise COBRA Resources A, B, C, E and F.

Part I

METHODOLOGY FOR ASSESSING THE RISKS AND IMPACTS OF AI SYSTEMS FROM THE PERSPECTIVE OF HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW

Introduction

What is HUDERIA?

The risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law (HUDERIA) is guidance that provides a structured approach to risk and impact assessment for AI systems specifically tailored to the protection and promotion of human rights, democracy and the rule of law. It is intended to play a unique and critical role at the intersection of international human rights standards and existing technical frameworks on risk management in the AI context.

HUDERIA can be used by both public and private actors to help identify and address the risks to and the impacts on human rights, democracy and the rule of law throughout the life cycle of AI systems. It originated from the work of the Ad Hoc Committee on Artificial Intelligence (CAHAI) (2019-2021) and specifically its Policy Development Group, which mandated the Alan Turing Institute, the UK's national institute for data science and AI, to prepare an original proposal operationalising the outline for a model for a human rights, democracy and the rule of law impact assessment. The HUDERIA Methodology was adopted by the Committee on Artificial Intelligence (CAI) of the Council of Europe on 28 November 2024.

Relationship to the Framework Convention

HUDERIA is stand-alone, non-legally binding guidance that, as such, does not have legal effect. It is not mandatory, nor is it intended as an interpretative aid for the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, hereinafter referred to as "the Framework Convention". Many existing or future frameworks, policies, guidance, standards or tools may be used to assist in conducting AI risk and impact management, including HUDERIA.

Parties to the Framework Convention have the flexibility to use or adapt the guidance, in whole or in part, to develop new approaches to risk assessment or to use or adapt existing approaches in keeping with their applicable laws, provided that parties fully meet their obligations under the Framework Convention, including in particular the baseline for risk and impact management set out in its Chapter V.

Principal objectives of HUDERIA

The aims of HUDERIA are:

- to help determine the extent to which risk-management activities related to human rights, democracy and the rule of law may be called for, and to offer a methodology for risk and impact identification, assessment, prevention and mitigation that is applicable to a variety of AI technologies and application contexts and is responsive to future developments in AI technologies and applications;
- to promote compatibility and interoperability with existing and future guidance, standards and frameworks developed by relevant technical, professional and other organisations or bodies (such as the International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), International Telecommunication Union (ITU), European Committee for Standardization (CEN), European Committee for Electrotechnical Standardization (CENELC), Institute of Electrical and Electronics Engineers (IEEE), Organisation for Economic Co-operation and Development (OECD), National Institute of Standards and Technology (NIST)), including the NIST AI Risk Management Framework and risk management and fundamental rights impact assessment under the European Union AI Act.

What is the approach of HUDERIA?

HUDERIA combines contemporary knowledge about the technical and socio-technical governance processes and mechanisms that can facilitate responsible activities within the life cycle of AI systems with the diligence procedures needed to protect and promote human rights, democracy and the rule of law.

HUDERIA takes as a basis well-known variables, concepts and language for the assessment of risks to human rights (scale, scope, probability and reversibility of potential adverse impacts on human rights). It aims to facilitate their examination by providing additional guidance in view of the socio-technical complexity of the AI life cycle.

Socio-technical approach

HUDERIA adopts a socio-technical approach, which views all aspects of the AI system life cycle as affected by the interconnected relationship of technology, human choices and social structures. In this approach, risk and impact management of AI systems takes account of both their technical aspects and the legal, social, political, economic, cultural and technological contexts in which they operate. Such an approach promotes the development of safe, secure and trustworthy AI that is both performant and promotes respect for human rights, democracy and the rule of law.

General and specific guidance

HUDERIA offers structure by combining general and specific guidance and flexibility by allowing room for adaptation in the practical implementation.

At the general level, the HUDERIA Methodology describes high-level concepts, processes and elements guiding risk and impact assessment activities of AI systems that could have impacts on human rights, democracy and the rule of law.

At the specific level, the HUDERIA Model will provide supporting materials and resources (such as flexible tools relevant for different elements of the HUDERIA process and scalable recommendations) that can aid in the implementation of the HUDERIA Methodology. These resources are referred to throughout the text and will provide a library of knowledge that can facilitate consideration of risks and impacts related to human rights, democracy and the rule of law, including in other approaches to risk management.

Adaptability and flexibility

Both the HUDERIA Methodology and HUDERIA Model allow room for adaptation to different contexts, needs and capacities by setting goals, principles and objectives, while leaving a margin of appreciation to decide on how to meet them and offering a range of policy and governance options that can be tailored to different contexts.

Graduated and differentiated approach

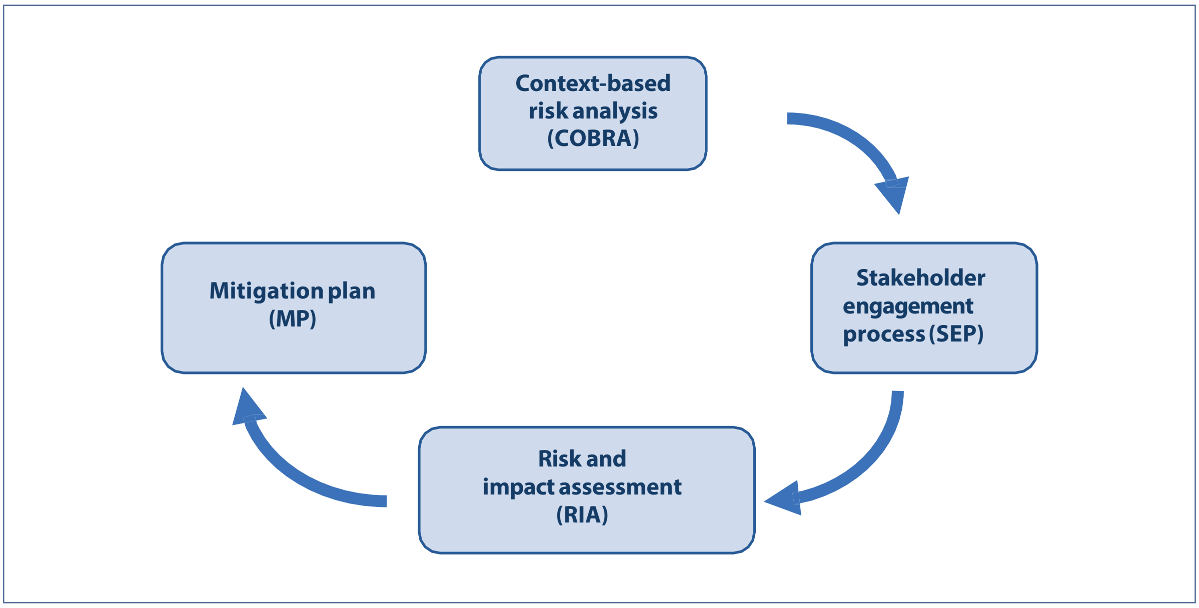
HUDERIA aims to establish a graduated and differentiated approach to measures for risk and impact identification, assessment, prevention and mitigation that takes into account the severity and probability of the occurrence of the adverse impacts on human rights, democracy and the rule of law as well as relevant contextual factors.

Outline of HUDERIA

The HUDERIA Methodology contains four elements.

1. The **context-based risk analysis** (COBRA) provides a structured approach to collecting and mapping the information needed to identify and understand the risks the AI system could pose to human rights, democracy and the rule of law in view of its socio-technical context. It also supports an initial determination of whether the AI system is an appropriate solution for the problem being considered.
2. The **stakeholder engagement process** (SEP) proposes an approach to enabling and operationalising the engagement, as appropriate, with the relevant stakeholders in order to gain information regarding potentially affected persons and contextualise and corroborate potential harm and mitigation measures.
3. The **risk and impact assessment** provides possible steps for the assessment of the risks and impacts related to human rights, democracy and the rule of law.
4. The **mitigation plan** provides possible steps for defining mitigation and remedial measures, including access to remedies and iterative review.

While it is logical to carry out the COBRA element first, depending on the needs and approaches, one may choose to change the sequence of the elements and/or apply or otherwise use only certain parts of the methodology based on existing AI governance approaches and specific contexts, needs and capabilities.


Context-based risk analysis (COBRA)
Stakeholder engagement process (SEP)
Risk and impact assessment (RIA)
Mitigation plan (MP)

1. Context-based risk analysis (COBRA)

Introduction

COBRA assists in the identification of different risk factors – characteristics or properties of an AI system and its context that affect the probability of adverse impacts on human rights, democracy and the rule of law. 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 therefore need to be anticipated and considered in risk management and impact mitigation efforts. The risk factors are categorised into three broad areas: the system's application context, its design and development context, and its deployment context.

Examination of the risk factors is intended to facilitate the mapping of potential adverse impacts on human rights, democracy and the rule of law. The results of this risk factor and impact mapping analysis are intended to inform the extent of the approach to subsequent elements of HUDERIA, including by establishing the proportionality of subsequent HUDERIA activities.

The results of this risk factor and impact mapping analysis may also help to pinpoint the specific socio-technical contexts across the system's life cycle that need focused governance attention.

The COBRA element consists of four steps:

1. preliminary scoping;
2. analysis of risk factors;
3. mapping of potential impacts on human rights, democracy and the rule of law;
4. triage.

Preliminary scoping

Objectives

The main purpose of this stage is to carry out the preliminary background research needed to inform subsequent risk factor identification and impact mapping activities.

Explanations

The COBRA process begins with preliminary scoping research that:

- outlines the purpose of the system, key components of the system, the contexts in which it is intended to be used, the area/domain(s) in which it will operate, the degree of human intervention and the nature and amount of data it will process and on which it will be trained, noting any checks that may have already been done to assess bias in the dataset or model;
- identifies persons or groups who may be affected by, or may affect, the system, focusing on the relevant contextual characteristics of identified persons and groups including protected characteristics and vulnerability factors, provides a preliminary scoping of potential adverse impacts on human rights, democracy and the rule of law by exploring the illustrative areas of concern;[2]
- and provides an initial mapping of roles and responsibilities across the AI system's life cycle.[3]

[2] COBRA resources E (Illustrative areas of potential concern from the point of view of human rights, democracy and the rule of law) provides a tool that could be used to perform or inform this assessment.

[3] The "Roles and responsibilities" section in the HUDERIA Model [to be developed and adopted by the CDNET in 2026] will provide guidance in connection with this aspect of the Methodology.

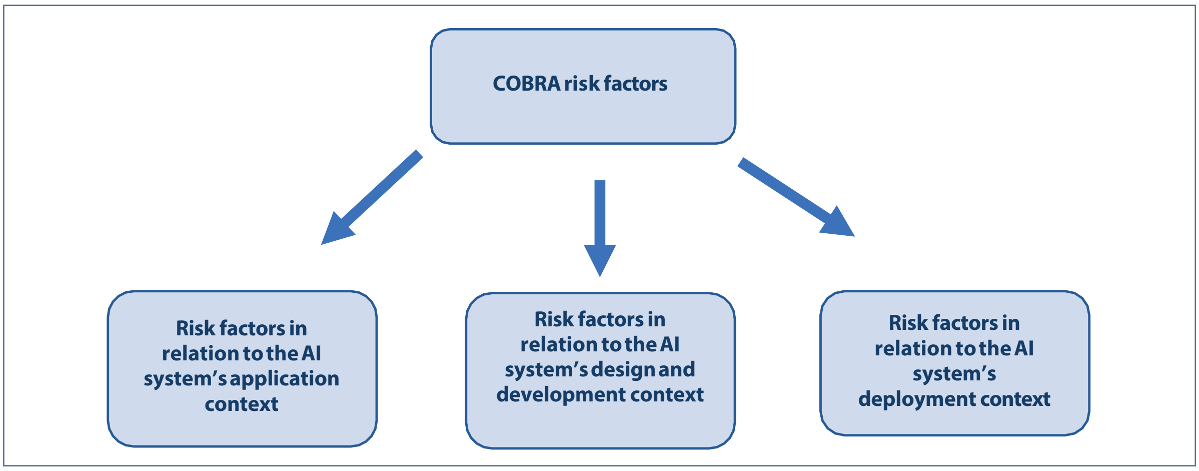
This preliminary scoping activity could draw on organisational documents (the project business case, proof of concept or project charter), collaboration and desk research (if necessary). This preliminary scoping activity, and subsequent elements of the HUDERIA process, should take place, as appropriate, in a multidisciplinary team, consisting of experts with a range of complementary specialisations[4] and both technical and non-technical backgrounds.

Analysis of risk factors

Objectives

The main purpose of this stage is to collect the relevant information about risk factors related to the system's intended application context, design and development context and deployment context. These risk factors will facilitate the mapping of potential adverse impacts on human rights, democracy and the rule of law and the subsequent assessment of key risk variables: severity (scale, scope and reversibility)[5] and probability.

Explanations


COBRA risk factors
Risk factors in relation to the AI system's application context
Risk factors in relation to the AI system's design and development context
Risk factors in relation to the AI system's deployment context

AI systems are designed, developed and used in a wide variety of contexts and in numerous different ways, making it important to holistically assess various factors related to the system's application context, its design and development context and its deployment context.

The AI **system's application context**[6] includes information about the system's application sector and domain, the legal and regulatory environments in which the system is being developed and used, the system's intended purpose, and other relevant details of the system's application context, such as any known legacies of bias of discrimination.

The AI **system's design and development context**[7] includes the relevant technical characteristics of the system. This may include known limitations of the system, considerations related to data collection, enrichment, storage, use and retirement; and considerations related to the algorithm or model itself. Particularly relevant considerations include technical characteristics related to privacy and data protection, bias and discrimination, and explainability and interpretability.

[4] Relevant domain expertise may include, as appropriate, issues of human rights, privacy and personal data protection, data science, data set management, security, AI risks, and AI testing, evaluation, verification and validation.

[5] Following the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, the United Nations Office of the High Commissioner for Human Rights and the United Nations Guiding Principles on Business and Human Rights, for the purposes of HUDERIA the term "severity" is understood to be composed of a combination of the variables of scale, scope, and reversibility.

[6] COBRA Resources A (List of risk factors arising in the system's application context) provides a tool that could be used to perform or inform this assessment.

[7] COBRA Resources B (List of risk factors arising in the system's design and development context) provides a tool that could be used to perform or inform this assessment.

The AI **system's deployment context**[8] includes factors that govern how potential risks may manifest and be managed in practice, such as steps that will be taken to protect privacy and personal data, mitigate harmful bias, ensure proper training, guard against unintended uses and ensure accountability and legal compliance.

Mapping of potential impacts on human rights, democracy and the rule of law

Objectives

The mapping step identifies potentially affected persons or groups and makes the initial assessment of the key risk variables – severity (scale, scope, reversibility) and probability. The mapping helps to inform subsequent elements of the methodology and the extent of governance intervention and mitigation measures that may be appropriate (see the "Triage" section below). The analysis of the key risk variables is crucial for a clear, structured overview of where threats are most likely to occur and their potential impact.

Explanations

COBRA Resources E and F could be used to identify potentially sensitive application areas and potentially relevant areas of concern related to human rights, democracy and the rule of law[9].

Using the information collected in previous steps:

a. determine whether the system will operate in proximity to activity(ies) (decision making or actions) which may produce impacts on affected persons in the relevant sectors/domains;

b. identify and enumerate the relevant areas of concern[10] and, with this in mind, answer the question of whether the system may have potential or actual impacts on specific human rights, democracy and the rule of law;

c. for each potential and actual impact identified, describe the nature of the potential and actual impact,[11] taking account of differential impacts on affected persons and groups given relevant contextual characteristics, including protected characteristics and vulnerability factors.

Analysing these points will provide information for the initial assessment of key risk variables

- severity (scale, scope, reversibility) and probability – that assist in determining risk and choosing the right approach to the subsequent elements of the methodology, which will help ensure that governance interventions and mitigation measures are aligned with the needs throughout the AI system's life cycle.

The results of this analysis can also help to identify opportunities for using the AI system to support positive actions that advance human rights, including promoting and ensuring non-discrimination.

Determination of risk level

The following variables may be employed to index the risk level of each of the potential adverse impacts to human rights, democracy and the rule of law that have been identified in as the result of the mapping exercise.

1. The **scale**[12] of the potential adverse impacts (the seriousness of the potential harm).

2. The **scope** of the potential adverse impacts (including the number of persons affected, the protected characteristics or vulnerability of individuals or groups and the time frame of the impacts).

[8] COBRA Resources C (List of risk factors arising in the system's application context) provides a tool that could be used to perform or inform this assessment.

[9] References in the COBRA Resources to human rights as set forth in various international human rights instruments are included for illustrative purposes. Those references only apply to states that are parties to those instruments. Each state is expected to apply its own applicable laws and international obligations.

[10] COBRA Resources E and F could be used to identify potential areas of concern related to human rights, democracy and the rule of law as well map relevant sectors and domains.

[11] For clarity of assessment both adverse or restrictive impacts as well as beneficial, enhancing or otherwise positive impacts produced by AI systems should be accounted for, since various issues with bias and discrimination may arise in respect of systems that produce both types of impacts.

[12] The term "scale" may sometimes be referred to as "gravity" in risk-assessment contexts.

3. The **reversibility**[13] of the potential adverse impacts is the information on the possible reparability or restoration for affected persons to their pre-impact situation or equivalent.

4. The **probability**[14] of the potential adverse impacts.

Relevant teams should go through each of the potential impacts that have been identified and consider for each area of concern related to human rights, democracy and the rule of law, and each affected group, the scale, scope, reversibility and probability of the potential or actual adverse effect. Domestic law or policy may provide more detailed definitions that can be used to inform this determination of risk and to determine suitable and proportionate approaches to subsequent HUDERIA activities (such as stakeholder engagement).

Consideration may be given to establishing a method for combining these variables to enable the calibration of risk and the determination of suitable and proportionate approaches to

subsequent HUDERIA activities (such as stakeholder engagement) as well as the extent and depth of downstream governance interventions and risk-management and mitigation measures. This may involve the formulation of quantitative or semi-quantitative methods of risk calculation, risk matrices or more qualitative or rules-based procedures. Any risk calibration mechanism resulting from the combination of these variables for the purposes of the HUDERIA assessment may take into account:

- in terms of human rights, the low scope and high gravity effects as well as high scope and low gravity effects on each affected person;
- in terms of democracy and the rule of law, the high scope and long-lasting effects on persons, institutions and society in general.

Triage

Objectives

The main purpose of this stage is to build on the information collected in previous COBRA activities and therefore:

- to facilitate the task of identifying and triaging systems that pose significant risk, so that the HUDERIA Methodology is not onerous for minimal or low-risk AI systems;
- to make an initial determination of whether the AI system should be developed or deployed, based on whether the benefits of developing or deploying the AI system outweigh its risks, particularly given its potential impacts on human rights, democracy and the rule of law, as well as whether the use of the AI system is incompatible with respect for human rights, democracy and the rule of law.

Adaptable approach to triaging

The prior activities in this stage provide a preliminary picture of the risk profile of the AI system.

The information gathered may, for example, be sufficient to determine that the system is unlikely to have any impact on human rights, democracy or the rule of law, making the subsequent elements of HUDERIA unnecessary. A similar conclusion could be reached if the identified impacts are insignificant or unlikely. If the identified impacts lead to a decision that an AI system will not be developed or deployed because it is considered incompatible with respect for human rights, democracy and the rule of law, subsequent elements of HUDERIA are also unnecessary. Finally, in cases where serious potential impacts are identified, a range of risk-management strategies and responses (including the stakeholder engagement process described in the next section) may be justified. To address this complexity, HUDERIA does not prescribe detailed guidance for adjusting risk-management efforts, but simply sets out proposed elements that may be applied as appropriate based on the risk of adverse impacts to human rights, democracy and the rule of law. Different approaches to determining risk-management steps based on the potential and actual impacts identified – or a combination of them – may be applied based on the specific domestic regulatory framework or environment, industry, system and context (for example threshold-based, scenario-based, proportionality, dynamic or context-specific approaches).

The final determination of whether to use a qualitative, quantitative, mixed or any other method is left to the discretion of the authorities or, where applicable, the AI project teams responsible for the system.

[13] The term “reversibility” may sometimes be referred to as “remediability” in risk-assessment contexts.

[14] The term “probability” may sometimes be referred to as “likelihood” in risk-assessment contexts.

“Zero questions”

To help determine whether the benefits of building or deploying the AI system, including additional social benefits that may result from the use of the system beyond its primary purpose, outweigh its risks given the risk factors and potential impacts identified, consider:

- whether the use of the system is appropriate considering the nature of the problem that the AI system is trying to solve –the extent to which existing technologies and processes already in place to solve the problem under consideration are better placed to do so, considering the risk profile and potential adverse impacts of the prospective system with a particular focus, where appropriate, on any marginal risk added by introducing AI into the current context;
- the extent to which the prospective system will be able to meet the deployer’s needs and expectations;
- the extent to which the impacts of the prospective system will be equitable across affected groups;
- the extent to which the quality and representativeness of currently or potentially available data is sufficient for the prospective system to be effective and safe and to reasonably avoid potential harmful bias;
- the extent to which sufficient resources (human and material) are available and able to meet technical requirements and perform technical and governance actions to adequately mitigate identified risks;
- the system’s potential use contexts and risks for misuse or abuse, including through deployment beyond its intended purpose.

2. Stakeholder engagement process (SEP)

Introduction

A decision on whether to run this step can be considered in order to improve the quality of information for the next element of HUDERIA – the risk and impact assessment – by incorporating the views of potentially affected persons, including those in vulnerable situations.

Stakeholder engagement, as set out in the HUDERIA Methodology, may take various forms. The level of participation of affected persons should be informed by the risks factors and potential and actual impacts identified as part of the COBRA stage. Involving stakeholders throughout the AI system's life cycle can also offer a variety of additional benefits, such as fostering transparency, building trust and improving usability and performance of the AI system.

Explanation

The SEP involves five key steps:[15] stakeholder analysis; positionality reflection; establishment of engagement objectives; determination of an engagement method; and implementation.

Stakeholder analysis

The stakeholder analysis identifies stakeholder groups that may be affected by, or may affect, the activities within the life cycle of the system. Such analysis[16] assesses the relative interests, rights, potential and existing vulnerabilities and advantages of identified stakeholders as well as the salience of identified stakeholder groups. At this step, consider meaningfully including the views of those who:

1. are disproportionately at risk from the use of the system;
2. are particularly vulnerable to potential harms;
3. have particularly limited ability to influence how the system is designed and used (such as currently or historically marginalised, disadvantaged or under-represented groups or persons in situations of vulnerability or presenting specific needs).

Positionality reflection

The next step involves reflection on the positional standpoint vis-à-vis affected stakeholders with a view to recognising the limitations of HUDERIA users' perspectives and identifying missing viewpoints that would strengthen the assessment of the system's potential and actual impacts.

Depending on the relevant risk factors and potential impacts identified at the COBRA element, this may include an assessment of HUDERIA users' self-identified demographics, education and training, socio-economic background and the institutional and team context.

The main questions on which HUDERIA users should reflect when undertaking this stage of the methodology are as follows.

- To what extent do my personal characteristics, group identifications, socio-economic status, educational, training and work background, team composition and institutional frame represent sources of power and advantage or sources of marginalisation and disadvantage?
- How does this positionality influence my and my team's ability to identify and understand affected stakeholders and the potential impacts of the AI system?

[15] The process described in this section is illustrative in nature with the final determination on the process of stakeholder engagement being up to the discretion of the authorities or, where applicable, the AI project teams responsible for the system.

[16] SEP Resources (List of questions to assess relative stakeholder salience) [to be developed and adopted by the CDNET in 2026] provides detailed questions and tools to guide the identification of relevant stakeholders.

Depending on the risk factors identified during the COBRA process, HUDERIA users should also consider engaging external stakeholders or consultants with specific expertise, such as human rights law expertise, related to the system's potential and actual human rights impacts.

Establishment of engagement objectives

Setting clear objectives for stakeholder engagement aims to create a clear understanding of how and why engagement activities are being conducted. These facilitate the inclusive, informed and meaningful involvement of potentially affected persons.[17]

Determination of engagement method

Determining the appropriate stakeholder engagement method(s)[18] necessitates evaluation and accommodation of the needs of potentially affected persons, taking into consideration, as appropriate, the outcomes of the COBRA process and other relevant factors such as resource constraints, difficulties in reaching isolated or socially excluded groups, capacity constraints such as challenges resulting from digital divides or information gaps, time frames, etc.

The following criteria may serve as guidance in the SEP element.

1. **Engagement** – meaningful involvement of affected or potentially affected persons is integrated during the relevant elements of the process.
2. **Equality and prohibition of discrimination** – engagement and consultation processes are inclusive and gender-sensitive and account for the needs of persons and groups with protected characteristics or who may be at risk of vulnerability or marginalisation.
3. **Empowerment** – consideration of age-appropriateness and accessibility needs, and capacity building of persons and groups with protected characteristics or who may be at risk of vulnerability or marginalisation is undertaken to ensure their meaningful involvement.
4. **Transparency** – provide for the sharing of meaningful and intelligible information between stakeholders at relevant and regular intervals, make available information about the understanding of potential implications and human rights impacts, where appropriate, and publicly communicate HUDERIA findings and impact management plans (action plans).
5. **Accountability** – responsibility for the implementation, monitoring and follow-up of mitigation measures is assigned to particular entities, individuals or functions within the organisation.

Implementation

Having completed the preceding four activities, the appropriate engagement process can be undertaken. It should be consistent with the results of the stakeholder analysis, positionality reflection and established engagement objectives and methods, and be appropriately documented.

[17] SEP Resources [to be developed and adopted by the CDNET in 2026] provides indicative detailed questions and the description of options for stakeholder engagement.

[18] SEP Resources (Examples of relevant engagement methods with relevant questions) [to be developed and adopted by the CDNET in 2026] provides possible examples of relevant engagement methods and a list of relevant questions that can aid determination of appropriate stakeholder groups.

3. Risk and impact assessment

Introduction

The purpose of the risk and impact assessment is to provide detailed evaluations of the potential and actual impacts that the activities within the life cycle of an AI system could have on human rights, democracy and the rule of law.

In accordance with the triage made in the COBRA step, carrying out the risk and impact assessment is particularly important for AI systems that may pose significant risks to human rights, democracy and the rule of law. Following the triage of the COBRA analysis, this step may be needed only for certain AI systems, in particular those assessed as posing significant risks to human rights, democracy and the rule of law.

The risk and impact assessment aims to:

- re-examine, contextualise and corroborate the potential and actual harm identified in the COBRA step;
- identify and analyse further potential and actual harm by engaging in extended reflection to pinpoint gaps in the completeness and comprehensiveness of the previously enumerated examples of harm;
- evaluate the risk variables of scale, scope, reversibility and probability of the potential adverse impacts, so that their risks can be better assessed to be subsequently prioritised, managed and mitigated.

The risk and impact assessment builds upon the initial identification of the context-based risk factors to human rights, democracy and the rule of law and the mapping of potential impact on human rights, democracy and the rule of law carried out in the COBRA step and the potential insights from the SEP to address the potential and actual impacts of the AI system.

This is done meaningfully through a two-step process that enables the formation of a mitigation plan and the establishment of access to remedies during the next step of the HUDERIA Mwethodology.

Explanations regarding the risk and impact assessment questions and prompts

Introduction

The risk and impact assessment in the context of HUDERIA is organised into two steps.

In the first step, the focus is on identifying potential impacts and, more specifically, "how" the potential and actual impacts identified at the COBRA and SEP steps could occur, enabling a more open-ended and exploratory approach that allows for deeper analysis of the specific contexts, scope, scale and reversibility of impacts, particularly concerning individuals in vulnerable situations or vulnerable groups.

In the second step, the assessment of the risk variables of scale, scope, reversibility and probability of potential or actual impacts identified takes place. A thorough context-responsive consideration of these variables helps prioritise mitigation actions by differentiating the severity of AI system impacts.

Scale

The scale of a potential and actual adverse impact refers to the seriousness of the potential harm's expected consequence.

Consideration of the gravity of any potential harm should include reflection on the different ways and different extents to which persons or groups (in particular those who possess characteristics that could make them more vulnerable to the adverse impact) could suffer that harm.

Deliberations on scale should consider the following additional questions.

1. For each potential and actual adverse impact identified, are there persons or groups who possess characteristics that could make them more vulnerable to the impact? If so, what are these characteristics and could those who possess them suffer the harm more acutely or seriously than others?
2. For each potential and actual adverse impact identified, which persons or groups could encounter the gravest impact from the harm under consideration?

Responses to these questions will subsequently serve an important function during the mitigation planning stage when the redress and prioritisation of potential harm are under consideration.

Scope

The scope of a potential and actual adverse impact refers to the estimation of both the number of affected persons and the time frame of the impacts.

The estimations of scope for identified potential and actual adverse impacts are analysed one by one with special consideration given to the exposure levels of particular groups of affected persons to harm and to cumulative or aggregate impacts of the system on present and future potentially affected persons and groups of persons.

Deliberations on scope may include consideration of these questions.

1. For the potential and actual adverse impact identified, are there groups who possess characteristics that could make them vulnerable to higher levels of exposure[19] [20] to the impact? If so, how much exposure could these groups face?
2. For the potential and actual adverse impact identified, consider the overall time scale of the AI system's impacts on the right or area of concern (in the case of democracy or the rule of law) under consideration. Are there cumulative or aggregate impacts of the system on affected persons and future affected persons that could expand the impacts of the system beyond the scope of impact already identified?
3. Some "big picture" questions to reflect on when assessing cumulative or aggregate impacts may include the following.
   - Could the provision and use of the system contribute to wider adverse human rights, democracy or rule of law impacts when its deployment is co-ordinated with (or occurs in tandem with) other systems that serve similar functions or purposes?
   - Could the provision and use of the system replicate, reinforce or augment socio-historically entrenched legacy harm or inherent characteristics in ways that could create knock-on effects for impacted persons and groups?
   - Could the provision and use of the system be understood to contribute to wider aggregate adverse impacts (on the environment or public health, for example) when its deployment is considered in combination with other systems that may have similar impacts?

Reversibility

As explained previously, reversibility refers to the information about the degree of reparability or restoration that is possible for potentially affected persons as a result of efforts to overcome the adverse impact under consideration and to restore those affected to a situation similar or equivalent to their situation before the impact. Much as with considerations surrounding the scale of a potential impact, gaining an understanding of how reversible a harm is will depend on knowledge both about the specific context of the harm and about the affected persons who are subjected to it. Establishing the degree of reversibility for a potential adverse impact involves considerations about the effort needed to overcome and (potentially) reverse the harm.

[19] SEP Resources A and B [to be developed and adopted by the CDNET in 2026] provide detailed questions (List of questions to assess relative stakeholder salience) and the description of formats of stakeholder engagement (List of factors determining the objectives and levels of stakeholder engagement) that can assist project teams in determining particularly relevant stakeholder groups and objectives for engagement.

[20] The term "level of exposure" here is understood as the proportion of a group that is adversely impacted by an AI system, where, in cases where a small fraction of the group is impacted, members have low levels of exposure, and in cases where a very large fraction of the group is impacted, members have high levels of exposure. As an example, members of a group that is characterised by low socio-economic status may have a high level of exposure to the potential adverse impacts of an AI model that is used to allocate public benefits.

Members of different groups may require different levels of effort to overcome adverse impacts, depending on their age, their positions in society and the circumstances of the harm (with vulnerable and marginalised groups often possessing less resilience than other dominant, privileged or majority groups).

Probability

Assessing the probability of a risk involves estimating the likelihood that a given adverse impact will occur, based as appropriate on qualitative judgment, quantitative analysis and contextual understanding.

Determining a risk's level of probability involves a broad analysis of contextual and operational conditions and generally determined by the level (kind, quantity and quality) of information that the risk is likely to materialise. This ensures that risk assessments are grounded in both data and expert insights, making it easier to prioritise and mitigate potential risks.

Outcome of the risk and impact assessment

After questions and prompts on identifying and assessing potential and actual adverse impacts have been completed, impact prevention and mitigation prioritisation and planning can be launched. The process of impact mitigation planning and setting up access to remedies is covered in the next step.

4. Mitigation plan

Introduction

Once potential and actual adverse impacts have been identified and assessed, a mitigation plan should be drawn up and a reflection regarding the provision of access to remedies to potentially affected persons should take place, as appropriate.

This part of the HUDERIA process specifies the actions and processes aimed at addressing potential and actual adverse impacts through:

- formulating mitigation measures;
- drawing up a mitigation plan based upon the severity and probability of the identified harm;
- where appropriate, setting up access to remedy for potentially affected persons and other relevant parties.

Explanations

Scoping and prioritisation

Diligent risk and impact prevention and mitigation planning begins with a scoping and prioritisation stage. With input from affected persons if and as appropriate, one should go through each identified potential and actual adverse impact and map out the interrelations and interdependencies between them as well as the surrounding social risk factors identified at the COBRA stage (such as, for instance, contextually specific vulnerabilities and precariousness).

Where prioritisation of prevention and mitigation actions is necessary (for instance, where delays in addressing a potential harm or the specific vulnerability of an affected individual or group could reduce its reversibility), decision making should be steered by considerations of the relative probability and severity of the impacts under consideration.

Legal obligations

An important consideration in the drawing up of a mitigation plan is that legal obligations in regard to the respect for human rights, democracy and the rule of law, as set forth in applicable international and domestic law, should be taken into account at this stage of the HUDERIA process when considering whether and, if so, how potential adverse impacts can be mitigated and actual adverse impacts can be addressed.

The availability and effectiveness of legal remedies, including restoration or compensation as legal remedies, are determined by applicable international and domestic law.

Mitigation hierarchy

When deciding upon the range of available actions that can be taken to prevent or mitigate potential adverse impacts, a structured approach called the "mitigation hierarchy" (avoid, mitigate, restore, compensate) may be used.

During the early stages of an AI system's life cycle, the impacts under consideration will not yet have occurred, so the mitigation options of "avoid" and "mitigate" will be more relevant. In later iterations of the monitoring, review and re-evaluation (that is, during the deployment stage) adverse impacts may have already occurred, making the mitigation options of "restore" and "compensate" relevant alongside "avoid" and "mitigate".

Descriptions of the options within the mitigation hierarchy are as follows.

AVOID	MITIGATE	RESTORE	COMPENSATE
Making changes to the design, development or deployment processes behind the production and use of the AI system, or to the AI system itself, at the outset, to avoid adverse impact. It is important to note that avoid does not equate to ignoring potential negative impacts.	Implementing actions in the design, development or deployment processes behind the production and use of the AI system, or making changes to the AI system itself, to minimise adverse impact.	Making changes to restore or rehabilitate affected persons to a situation similar to, or at least equivalent to, their situation before the adverse impact.	Compensation in kind or by other means, where feasible and when other mitigation approaches are neither possible nor effective.
Level 1	Level 2	Level 3	Level 4
Most preferred	…		Least preferred

The use of the term "mitigation hierarchy" suggests giving precedence to avoiding potential and actual adverse impacts altogether, in the first instance, and then to reducing and remediating them. It is also notable that in the later stages of an AI system's life cycle, where options of restoration and compensation become more relevant, more than one of these mitigation options may be relevant (where, for instance, an affected individual needs to be rehabilitated simultaneously as immediate actions to minimise further harms are also taken).

In all situations, decisions about which prevention and/or mitigation action(s) to take should be guided by considerations prioritising the protection of human rights, democracy and the rule of law, and choices made to avoid and mitigate adverse impacts should be preferred to choices to compensate or remunerate potentially impacted persons for any harm suffered.

In view of the entirety of the information obtained at this stage of the HUDERIA process, there is an opportunity to revisit the zero questions. This information may also be useful for informing the discussion on the question of whether the life cycle activities of the AI system at issue (under development or currently already in use in the case of an iterative review) align with human rights, democracy and the rule of law.

Access to remedies

Measures to address adverse impacts are not limited to legal remedies. Such impacts can be addressed using other mitigation measures such as those set out in policy, guidance or other instruments.

When putting in place such measures, the following points could be considered.

a. Whether there are in place existing accountability measures and mechanisms in relation to human rights, democracy and the rule of law. It is essential that these existing frameworks are applied to the context of artificial intelligence systems.

b. The technical complexity, opacity and data-driven nature of some AI systems can limit their transparency. This can create a significant imbalance in access to, understanding of or control over information between the various parties involved in the AI system's life cycle. Steps to document and provide information about the AI system and its impacts to affected persons can facilitate the provision and accessibility of effective remedies for adverse impacts on them.

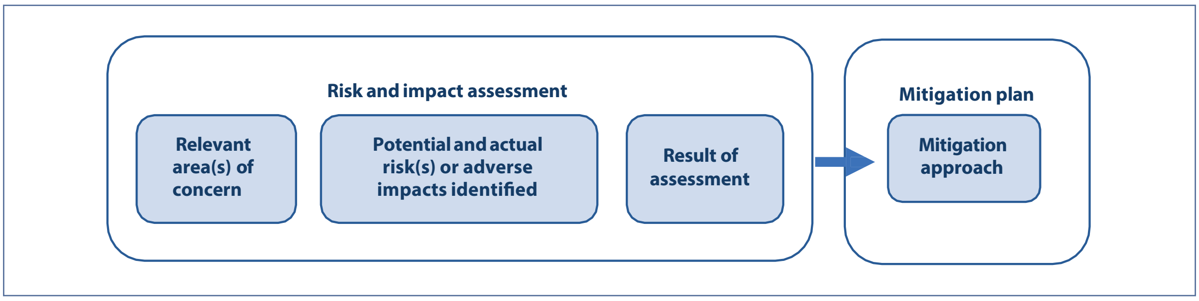
c. The information provided in these measures should be context-appropriate, clear and meaningful, ensuring that persons can effectively use it to exercise their rights in proceedings related to decisions impacting them.

d. If, and as appropriate, the provision of further effective procedural guarantees and safeguards to the affected persons, in line with applicable international and domestic law, may be required.

Outcome of this element of the HUDERIA process

This element should produce a clear description of the measures and actions to address the risk and impacts identified, along with a clarification of the roles and responsibilities of the various actors involved in mitigation, management and monitoring. Where appropriate, this element should also produce an accessible outline of the remedial mechanisms and measures available to impacted persons.

Additionally (see the "Iterative review" section below), a plan is established for monitoring mitigation efforts and for iteratively reassessing and re-evaluating these efforts throughout subsequent phases of the AI system's life cycle.


Risk and impact assessment
Relevant area(s) of concern
Potential and actual risk(s) or adverse impacts identified
Result of assessment
Mitigation plan
Mitigation approach

Iterative review

Introduction

Carrying out the HUDERIA process at the beginning of an AI system's life cycle is the first – albeit critical – step in a longer, iterative process of responsible monitoring and reassessment. The process of iterative review ensures that a risk and impact assessment remains effective throughout the whole AI system's life cycle. It is an ongoing process, offering regular opportunities to identify new impacts and to update the mitigation plan.

Over time, the impacts of the AI system are likely to evolve as a result of decisions made during its development and implementation, contextual applications or external changes in the real-world environment. These changes, which may include those regarding the data life cycle, AI system development and design, procurement processes, changes in AI techniques, system integration or operationalisation, security vulnerabilities, or significant events or occurrences leading to harmful or unintended consequences, can influence the AI system's performance and/or its impact on affected persons and groups.

Such changes necessitate a review procedure to ensure that human rights, democracy and the rule of law are continuously upheld throughout the AI system's life cycle. Particular attention should be paid to how these changes affect the system's performance and its impact on persons and communities.

Production, implementation and deployment factors

Choices made at any point during the life cycle of the system as well as events occurring during the system's deployment may require a review of prior decisions and assessments, particularly those made as a result of the HUDERIA process, creating the need for reassessment, reconsideration and amendment.

These changes, specifically those regarding the data life cycle, AI system development and design, changes in AI techniques, system integration or operationalisation, security vulnerabilities, or significant events or occurrences leading to harmful or unintended consequences, can influence the AI system's performance and/or its impact on affected persons and groups. The processes during an AI system's life cycle are iterative and often non-linear, frequently requiring revision and updates, as appropriate.

Real-world environment factors

Changes in social, regulatory, policy or legal environments in which the system is in production or use may have effects on how well the AI system works and on how it impacts the rights of affected persons or groups.

Likewise, regulatory and policy changes or changes in data-recording methods may take place in the population of concern in ways that affect whether the data used to train the model accurately portray phenomena, populations or related factors in an accurate manner.

In the same vein, cultural or behavioural shifts may occur within affected populations that alter the underlying data distribution and hamper the performance of a model, which has been trained on data collected prior to such shifts. All of these alterations of contextual conditions can have a significant effect on how an AI system performs and on the way it impacts affected persons, groups, communities and society in general.

Implementing the iterative review

While the HUDERIA Methodology provides flexibility on the exact modalities, thresholds, triggers, and monitoring and governance mechanisms for the iterative review process, the following principles could be considered.

a. A continual review of the HUDERIA process plays a pivotal role in its continued efficacy and reliability.

b. A plan is established for monitoring impacts and for reassessing and re-evaluating the HUDERIA process during each phase of the project's life cycle up to system retirement or decommissioning.

c. The processes used for iterative review should remain as responsive as possible to the way the AI system interacts with its operating environments and with impacted persons (possible application areas of the AI system, the emergence of new forms of system misuse, etc.).

d. In rapidly evolving or changing contexts, there may be a need for more frequent reassessment and re-evaluation interventions.

Part II

RESOURCE MODEL FOR CONTEXT-BASED RISK ANALYSIS (COBRA)

Preface

What is HUDERIA?

The risk and impact assessment of artificial intelligence (AI) systems from the point of view of human rights, democracy and the rule of law (HUDERIA) is guidance that provides a structured approach to risk and impact assessment for AI systems specifically tailored to the protection and promotion of human rights, democracy and the rule of law. It is intended to play a unique and critical role at the intersection of international human rights standards and existing technical frameworks on risk management in the AI context.

HUDERIA can be used by both public and private actors to help identify and address risks and impacts to human rights, democracy and the rule of law throughout the life cycle of AI systems.

Relationship to the Framework Convention

HUDERIA is a stand-alone, non-legally binding guidance that does not have legal effect. It is not mandatory, nor intended as an interpretative aid for the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, hereinafter referred to as "the Framework Convention". In addition, while HUDERIA has a facilitative role, it is not a means for implementing the Framework Convention. Many existing or future frameworks, policies, guidance, standards or tools may be used to assist in conducting AI risk and impact management, including HUDERIA.

Parties to the Framework Convention have the flexibility to use or adapt the guidance, in whole or in part, to develop new approaches to risk assessment or to use or adapt existing approaches in keeping with their applicable laws, provided that parties fully meet their obligations under the Framework Convention, including the baseline for risk and impact management set out in its Chapter V.

General and specific guidance

HUDERIA is comprised of the HUDERIA Methodology and the HUDERIA Model.

At the general level, the HUDERIA Methodology describes high-level concepts, processes and elements guiding risk and impact assessment activities of AI systems that could have impacts on human rights, democracy and the rule of law.

At the specific level, the HUDERIA Model provides supporting materials and resources (such as flexible tools relevant for different elements of the HUDERIA process, resources of an illustrative nature and scalable recommendations) that can aid in the implementation of the HUDERIA Methodology. The present document forms part of the HUDERIA Model and contains the COBRA (context-based risk analysis) resources, which help operationalise the COBRA process through structured questions, explanations and examples designed to support systematic risk identification and assessment.

These resources provide a library of knowledge that can facilitate considering and addressing risks and impacts related to human rights, democracy and the rule of law, including in other approaches to risk management. For greater certainty, the resources and guidelines in the HUDERIA Model are not provided as examples of best practices, nor do they set forth minimum standards; rather, they can be used to inspire customised approaches by both public and private actors to risk management that may be adopted by states.

COBRA Resources

Introduction

This document intends to provide illustrative guidance for the context-based risk analysis, the first element of the HUDERIA Methodology, and contains different elements, including material that can be used to guide reflections in the context of the assessment of potential risks and impact of AI systems. In line with the flexible approach of the Methodology, these elements can be adjusted as necessary. This material may also be used in future to build interactive tools (such as online platforms or interactive workflows) to facilitate the conduct of assessments of risks and impacts.

Elements in this section can be used to assist in identifying and assessing risk factors affecting the scope, scale, probability and reversibility of adverse impacts on human rights, democracy and the rule of law arising in the AI system's application context (COBRA Resources A), the AI system's design and development contexts (COBRA Resources B) and the AI system's deployment context (COBRA Resources C). These might need to be reviewed and considered iteratively throughout the life cycle of an AI system.

In order to facilitate the mapping of potential impacts on human rights, democracy and the rule of law, COBRA Resources E lists areas of potential concern from the point of view of human rights, democracy and the rule of law, provides an illustrative description of these areas and gives examples of potential AI-related risks for each of them. The list could be used in conjunction with COBRA Resources F, which lists and provides a brief description of potentially relevant sectors and domains and matches them with areas of potential concern in COBRA Resources E.

Overview of COBRA Resources A, B and C

This section provides illustrative examples of potential risk factors affecting the probability of adverse impacts on human rights, democracy and the rule of law arising in the following.

- An AI system's application context (COBRA Resources A), such as the legal and regulatory environments in which the system is being developed and used, the system's intended purpose, the categories of intended users and affected persons and other potential details pertaining to the system's application context, such as any known legacies of bias and discrimination.
- The AI system's design and development context (COBRA Resources B), in particular its technical characteristics, such as known limitations of the system, considerations related to data collection, enrichment, storage, use and retirement; and considerations related to the algorithm or model itself, technical characteristics related to privacy and personal data protection, bias and discrimination, and explainability and interpretability.
- The AI system's deployment context (COBRA Resources C), such as the characteristics of its intended users and affected groups, the measures taken to ensure overall security, prevent harm and uphold human dignity, protect privacy and personal data, mitigate harmful bias, ensure proper training, guard against unintended uses and ensure accountability and legal compliance.

The resources are illustrative, non-exhaustive and open-ended, meaning they will need to be adapted over time to account for developments, for example in the areas of technology and applicable legal and regulatory regimes.

COBRA resources A (List of risk factors arising in the AI system's application context)

	Examples of potential risk factors	
1. Sector or domain in which the AI system is being built	A.1.1	In light of its intended purpose, will the AI system play a significant role in a high-impact context – such as transport, social care, healthcare or education – where its output could influence important decisions or actions? What is the context in which the system is operating? *In such contexts, any malfunction or performance degradation throughout the system's life cycle can lead to harmful decisions, service interruptions or safety incidents. Consider whether the system might affect areas where human rights, democracy or the rule of law could be at stake.* ***COBRA Resources E*** *outlines examples of potential harm and protected interests that may be relevant and* ***COBRA Resources F*** *contains the list of sectors of potential relevance from the perspective of human rights, democracy and the rule of law.* ***Example:*** *a hybrid AI system used to control the kinematics of a surgical robot that assists doctors who are performing an emergency medical procedure.*
	A.1.2	Will the AI system perform a high-impact function independent of the sector in which it operates? What is the function of the system? *Even outside high-impact sectors, functions of this kind can lead to harmful decisions, service interruptions or other adverse outcomes that may affect the human rights of potentially affected persons or have effects on democracy and the rule of law – see* ***COBRA Resources E*** *for reference, even if it is not deployed in a traditionally high-impact context.* ***Example:*** *the system being developed is an AI-assisted human resources recruitment tool for a civil service function. The tool is not situated within a safety-critical sector; however, the system could have a negative impact on groups of individuals in situations of vulnerability.*
2. Existing law and regulatory environment	A.2.1	Is the sector or domain in which the AI system will be deployed subject to extensive legal or regulatory oversight? If so, what are the specific obligations or standards (compliance, certification, supervision, ethical) that may apply to the AI system or its outputs? *Legal or regulatory oversight in itself may be a signal of high-impact contexts. When such oversight exists, it is important to identify relevant obligations and standards that may interact with, or inform, HUDERIA considerations. The HUDERIA assessment focuses specifically on potential impacts on human rights, democracy and the rule of law. Other sectoral obligations (concerning financial stability, medical safety or cybersecurity, for example) remain subject to their respective regulatory frameworks. Where relevant, existing assessments or certifications (compulsory or voluntary) may be referenced as supporting documentation to strengthen the evidentiary basis of the HUDERIA analysis – but they do not replace or merge with it.* ***Example:*** *a bank using a risk-assessment tool to predict borrower default operates within the financial sector where extensive regulation pertaining to market abuse, risk management, equity law and competition law has historically been in place.*
3. Scope of the deployment	A.3.1	In a scenario where the AI system is fully deployed as designed, at what scale will the AI system directly and/or indirectly affect persons and groups (local populations, national populations, global populations)? *This question focuses on the spatial and jurisdictional reach of the AI system's potential impact. Scaling can materially alter the risk profile of the system with potential effects on local populations and on the communities and groups that comprise them as well as social and political processes. These impacts may be significant in relation to human rights, democracy and the rule of law (see* ***COBRA Resources E****, specifically areas (16) Democracy and (17) Rule of law). Where systems are introduced gradually or tested in pilot or trial stages on a limited group, such phased deployment can also serve as a mitigation measure, allowing potential impacts to be identified, assessed and managed before broader roll-out.* ***Example (local):*** *a predictive policing system used by a municipal police department and operating only within the boundaries of the municipality.* ***Example (national):*** *a classification system used by a governmental organisation to determine individuals' eligibility for social benefits affects individuals and groups within national populations.* ***Example (global):*** *a recommender system used by an international social media platform to personalise news delivery affects individuals and groups within the global population.*

	Examples of potential risk factors	
3. Scope of the deployment	A.3.2	In a scenario where the AI system scales optimally, how large a number of potentially affected persons will the AI system directly and/or indirectly affect? *This question focuses on the numerical and demographic magnitude of potential impact. The use of the AI system at scale or mass level may generate significant effects on social and political processes and institutions, with particular implications for human rights, democracy and the rule of law (see* ***COBRA Resources E****, specifically areas (16) Democracy and (17) Rule of law).* ***Example (small):*** *An AI chatbot is used by a local NGO with 60 employees to support survivors of domestic violence by providing information on legal rights and available shelters.* ***Example (moderate):*** *A regional employment agency uses a machine learning classifier to filter and prioritise job applicants for short-term public-sector contracts, serving around 10 000 applicants per year.* ***Example (large):*** *A national health insurer deploys a predictive system to flag "at-risk" patients for early intervention programmes affecting approximately 1.2 million policy holders.* ***Example (very large):*** *A government-run digital identity platform integrates AI-powered identity verification (such as facial recognition and document scanning) to control access to social protection schemes, affecting over 50 million people.*
	A.3.3	Considering the potential direct and indirect impacts of the AI system on persons, communities and the environment, what is the widest time scale within which the AI system could affect persons and groups? *The use of the AI system may give rise to longer-term effects extending beyond the period of its application, impacting the human rights of potentially affected persons, as well as social and political processes and institutions, particularly in relation to democracy and the rule of law (see* ***COBRA Resources E****, specifically areas (16) Democracy and (17) Rule of law).* ***Example (short-term):*** *A conversational AI system is used by a municipal housing office to respond to tenant queries about rent arrears, eviction notices and complaint procedures.* ***Example (medium-term):*** *A regional education authority uses an AI-driven profiling system to predict students' likelihood of academic failure, triggering targeted interventions or redirection to vocational tracks.* ***Example (long-term):*** *An AI classification system is deployed by a national migration agency to assess immigration, long-term visa or citizenship applications as low, medium or high risk.*
4. Existing legacies of bias	A.4.1	Do 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 patterns of discrimination, inequality, bias or inaccuracies that systematically lead to the unfair treatment of groups of individuals in situations of vulnerability (or groups of individuals with protected characteristics)? If so, how likely are these patterns to be replicated or augmented in the functioning of the system or in its output and short, medium and long-term impacts? Consider focusing upon equality and non-discrimination considerations surrounding the potential impacts of the AI system on the affected persons. *If such patterns exist, the system may replicate or amplify them in development and deployment, resulting in group-differentiated outcomes and indirect discrimination. At scale and over short, medium and long-term horizons, this can entrench disparities and undermine equality and non-discrimination.* ***Example:*** *the use of a job application screening system in a science and technology industry based on historic hiring data replicates patterns of hiring discrimination, delivering unfavourable outcomes for groups of individuals in situations of vulnerability.*
	A.4.2	In view of the sector or domain in which the AI system operates, will the potentially affected persons include those who may be significantly or disproportionately impacted, particularly groups of individuals in situations of vulnerability (or groups of individuals with protected characteristics) by the design and use of the system? *Vulnerability can be situational, arising from economic, social, institutional, technological or other contextual conditions. Consider focusing upon equality and non-discrimination considerations surrounding the potential impacts of the AI system on the potentially affected persons. Pay close attention to risk factors referred to in* ***risk factors B.5.1 to B.5.4*** *in* ***COBRA Resources B****.* ***Example:*** *a risk-assessment tool that is used by social workers to determine access to social benefits operates within the welfare administration domain, where persons in situations of economic or social vulnerability are significantly impacted.*

	Examples of potential risk factors	
4. Existing legacies of bias	A.4.3	*Is the AI system likely to be accessed by or impact upon children?* *Special attention should be paid to children because their evolving capacities, dependency relationships and limited ability to provide informed consent or seek redress create specific risk profiles that general equality or vulnerability assessments may not adequately capture. Impacts on children can be disproportionate and long-lasting due to developmental vulnerabilities, power asymmetries and heightened data sensitivity. These risks raise particular concerns regarding mental well-being, equality and non-discrimination, as well as the rights of the child (see **COBRA Resources E**, specifically area (14) Children). Additional risks may also arise from inadequate stakeholder engagement with children or with those responsible for their welfare (such as parents, guardians, civil society organisations or child-protection advocates) throughout the AI system's life cycle (see the SEP element in the HUDERIA Methodology).*
5. Environment context	A.5.1	What are the AI system's potential direct and indirect impacts on the environment (concerning for example energy consumption, carbon emissions, resource use or electronic waste)? To what extent are its development, deployment and use subject to existing environmental standards, regulations or sustainability commitments, and are these frameworks adequate to address the system's potential impacts? *The AI system may generate direct and indirect impacts on the environment, with potential consequences for energy efficiency, carbon emissions, resource use and waste management. These environmental impacts may in turn affect the enjoyment of human rights, particularly for individuals in situations of vulnerability or groups with protected characteristics (see **COBRA Resources E**).* ***Example:** An AI model used to generate daily analytic dashboards continues to reprocess an unchanging dataset of over 100 TB every night, consuming significant computing power and energy without producing any new or meaningful output. This design choice results in unnecessary emissions and resource consumption with no corresponding public or user benefit.*

COBRA Resources B (List of risk factors arising in the AI system's design and development context)

	Examples of potential risk factors	
1. Decision to design and definition of problem and outcome	B.1.1	What other approaches besides building the AI system could feasibly address the intended need(s)? When compared against the AI system, what benefits and risks would the other approaches present – particularly in terms of impacts on human rights, democracy and the rule of law? *The choice to build and deploy an AI system should be critically assessed against alternative approaches that could better serve the intended needs with fewer negative impacts. Risks include marginal or new harm associated with introducing AI into an existing context, especially regarding human rights, democracy and the rule of law. Assessing alternatives requires considering (a) potential adverse effects on human rights, democracy and the rule of law (including marginal risks of AI introduction); (b) whether existing technologies and processes, including mitigation and governance techniques, could address the need; (c) the sufficiency of available resources and data; (d) the complexity of the use contexts; (e) the potential benefits of using an AI-enabled solution, including opportunity costs of not proceeding; (f) the nature of the problem being solved, in particular if it is solving a policy or social problem. Failure to evaluate these alternatives may result in adopting a solution that unnecessarily increases risk or complexity.*
2. Technological maturity	B.2.1	Is the AI system's design based on well-understood and widely recognised validation techniques for a similar intended purpose in the same sector? If not, what risks may arise from limited technological maturity on the quality of the system's performance and resulting effects on human rights of potentially affected persons or democracy and the rule of law? *If the AI system's design and development is not based on well-understood techniques that have been previously validated for similar purposes in the same sector, including any applicable industry standards and best practices and taking into account the system's intended use and acceptable error threshold, there is a risk that its technological immaturity will undermine the quality, reliability and robustness of its performance. Such shortcomings can increase the likelihood of errors, reduce trustworthiness and heighten the risk of adverse impacts on potentially affected persons. However, innovation often entails novel or experimental approaches that diverge from established practice, which may increase uncertainty and the risk of performance failures or unforeseen impacts. In such cases, the absence of sectoral precedent should be recognised as a heightened risk factor requiring compensating control measures – such as enhanced human oversight, independent expert review or carefully supervised pilot deployments. These mechanisms are essential to manage uncertainty, verify the system's reliability and ensure that technological advancement proceeds in a safe and responsible manner.*

		Examples of potential risk factors
3. Existing system	B.3.1	If the AI system is to replace an existing system, have the flaws and risks of the system being replaced been identified and mitigated? *When an AI system replaces a human, technical or hybrid system serving a similar function, or if its deployment otherwise leads to human operators over-relying on AI output, risks may arise if the limitations, flaws or documented harm of the existing system are not properly understood and addressed. Failure to take these into account may lead to replication or amplification of existing risks, rather than improvement. The quality and impact of the AI system's performance will depend in part on how well it learns from, improves upon or augments the shortcomings of the replaced system.*
	B.3.2	Is the AI system replacing a human, technical or hybrid system that is critical infrastructure or serves a high-impact function? If so, what risks (like outages or disruptions) may arise from the process of updating or replacing the system, and how could these impact the human rights of potentially affected persons? *If the AI system replaces a human, technical or hybrid system that constitutes critical infrastructure or serves a high-impact function, the process of updating or replacing the system may pose significant risks. These may be risks arising from outages, disruptions of essential services or failures that could directly affect the human rights of potentially affected persons, democracy or rule of law. The critical nature of the replaced system magnifies the potential scale of such impacts.*
4. Cybersecurity context	B.4.1	What motivations and opportunities does the AI system present for malicious actors to breach, corrupt or misuse it? What risks are posed by attempts to compromise its safety, security and robustness (for example through adversarial attacks, data poisoning, model inversion or data breaches)? What risks could such malicious exploitation pose for the human rights of potentially affected persons, democracy or the rule of law? *The AI system may create motivations and opportunities for malicious actors to breach, corrupt or otherwise manipulate it. Such actions could be driven by financial gain, political objectives or other perceived benefits, and may result in the system being misused to facilitate human rights abuses, undermine trust or cause harm to potentially affected persons.*
5. Data quality and personal data protection	B.5.1	Are the data used in designing and developing the AI system sufficiently representative of the persons and groups affected, sufficiently accurate, complete, reliable, relevant, appropriate, up to date and of adequate quantity and quality for its intended use case, domain, function and purpose? If not, what risks may arise for affected persons from any shortcomings identified? Which techniques were used to perform this assessment? *If the data used in designing and developing the AI system are not sufficiently representative, sufficiently accurate, complete, reliable, relevant, appropriate, up to date and of adequate quantity and quality for the intended use case, domain, function and purpose, significant risks may arise. There may be risks arising from (a) the consequences of using inaccurate, inconsistent or incomplete data; (b) a lack of proper recording, traceability and auditability of data provenance and lineage across the system's life cycle; (c) measurement errors or biases introduced during data collection (through human involvement or otherwise); (d) missing or unusable data in collected or procured datasets; (e) a lack of transparency and accessibility of labelling/annotation processes for audit, oversight and review; (f) biases introduced by human labellers and annotators, including via proxies, without adequate safeguards; (g) biases introduced by automated labelling or annotation, if not subject to sufficient human oversight and transparent processes. Such shortcomings can undermine the reliability, trustworthiness and fairness of the AI system and may negatively affect the human rights of potentially affected persons.*
	B.5.2	Does the design and development of the AI system involve the use of personal data? If so, what risks to privacy and personal data protection may arise? *If personal data are used in the design and development of the AI system, there is a risk that personal data-protection requirements may not be respected. As applicable, there may be risks associated, among other things, with (a) insufficient information provided to affected persons and stakeholders about the consent or other legitimate basis for processing; (b) reliance on implied consent or unclear legal bases without consultation of affected persons and stakeholders regarding acceptability of data use; (c) insufficient safeguards for individuals regarding the use of their data once shared or used for training, including where applicable, the ability to retract or limit further use; (d) situations with potential divergence between applicable personal data-protection rules such as purpose limitation and the AI practices of large-scale or iterative data reuse; (e) the possibility of deanonymisation or re-identification through data linkage with existing, publicly available, or easily obtainable datasets, if not properly managed. Such shortcomings can undermine, among other things, privacy and personal data protection.*

		Examples of potential risk factors
5. Data quality and personal data protection	B.5.3	Will the AI system use dynamic data collected and processed in real time (or near real time) for continuous learning, adaptation or performance optimisation? If so, what risks may arise concerning safety, security, reliability, robustness, data quality, data integrity and – where appropriate – non-discrimination and bias mitigation? *If the AI system uses dynamic data collected and processed in real time (or near real time) for continuous learning, adaptation or performance optimisation, new risks may arise. There may be risks associated with threats to safety, security, reliability, robustness, data quality and data integrity. Real-time adaptation may also introduce or amplify risks of bias and discrimination if not properly managed. Such risks can undermine trust and may negatively impact the human rights of affected persons.*
	B.5.4	To what extent do the domain and type of data collected or procured pose risks of rapid or unexpected distributional shifts or drifts that could adversely impact the accuracy and performance of the AI system? *The domain in which data are collected or procured, and the type of data used, may pose risks of rapid or unexpected distributional shifts or drifts. Such shifts can undermine the accuracy, reliability and performance of the AI system, leading to errors or harmful outcomes for potentially affected persons if not anticipated and addressed. The absence of dynamic assessment, reassessment, validation and monitoring increases these risks.*
6. Model development and model implementation context	B.6.1	What potential risks of bias and indirect discrimination may arise in the context of model design, development and implementation? *Model design, development and implementation may introduce risks of bias and indirect discrimination. These risks may arise (a) during feature engineering (manual or automated), through grouping, disaggregation or exclusion of input features related to protected or sensitive characteristics – or their proxies – resulting in emergent bias; (b) through inferences generated by the model's learning mechanisms that are unreasonable, unfair, disparate in impact or influenced by hidden proxies for discriminatory features, thereby shaping output in discriminatory ways. If unaddressed, such risks can lead to unequal treatment and systemic discrimination, undermining the human rights of potentially affected persons.*
	B.6.2	Is the AI system built on techniques that are inherently hard to fully explain, predict or verify, such as non-deterministic systems, probabilistic models or evolving/dynamic models? If so, what are the potential risks that it could negatively impact the human rights of potentially affected persons, especially when the AI system is interacting with individuals? *If the AI system is built on techniques that are inherently difficult to fully predict, verify or explain – such as non-deterministic systems (different output for the same input), probabilistic models (likelihood-based outputs, such as Bayesian models or LLMs) or evolving/dynamic models (continuously learning or adapting) – specific risks may arise. There may be risks associated with: (a) reduced transparency, limiting the ability of affected persons to understand, challenge or seek redress for output; (b) unpredictable or unstable behaviour, where output may vary in ways that are difficult to foresee or audit; (c) misalignment between the model's intelligibility/accessibility and the sector-specific requirements, legal or otherwise, or expectations for its intended function. These risks may be amplified where appropriate mitigation measures are lacking – such as when organisations have insufficient capacity to provide complementary explanation mechanisms (such as surrogate models or feature-importance analyses) or when no safeguards exist to ensure reversibility or human intervention. Such risks can negatively impact human rights, democracy and the rule of law by reducing accountability, transparency and trust in the system's output.*
	B.6.3	To what extent has appropriate evaluation, verification and validation of the AI model been ensured throughout its life cycle, and how might any gaps identified affect the level of risk associated with its deployment or use? *While evaluation, verification and validation mechanisms are themselves mitigation and control measures, their absence or inadequacy constitutes a key risk factor affecting the reliability and safety of the AI model. If the AI model is not subject to sufficient monitoring, evaluation, verification and validation, there is a risk that errors, flaws or biases in its design and functioning will go undetected. Without transparent processes, including external peer review and independent expert evaluation, the reliability, fairness and safety of the system may be compromised, leading to harmful impacts on potentially affected persons and their human rights in addition to impacts on democracy and rule of law.*

		Examples of potential risk factors
6. Model development and model implementation context	B.6.4	To what extent does the AI system include processes of monitoring and regular re-evaluation to keep pace with real-world changes that may cause concept drifts or shifts in underlying data distributions? What risks may arise for affected persons if such processes are absent or insufficient? *Monitoring and re-evaluation operate as control measures designed to maintain system performance over time; it is their absence or insufficiency that creates the risk of degradation or unfairness in practice. If the AI system is not monitored and regularly re-evaluated to keep pace with real-world changes, there is a risk that it may gradually diverge from its intended purpose or no longer reflect the conditions under which it was designed and validated (concept drift or shifts in underlying data distributions). These changes can degrade the accuracy, fairness and reliability of the system over time, leading to harmful impacts on potentially affected persons.*
	B.6.5	To what extent are performance metrics for the AI system – beyond accuracy (sensitivity, precision, specificity) –appropriately selected, monitored and communicated in ways that reflect the specific context of the use case, minimise the risk of misinterpretation or misuse by users and stakeholders, and help ensure transparency and accountability throughout the system's life cycle? What risks may arise for affected persons if they are not? *The selection and communication of performance metrics function as safeguards that support transparency and accountability; where these mechanisms are weak or incomplete, they become risk points in themselves. If the performance metrics of the AI system are not carefully selected, contextually defined, monitored and transparently reported, several risks may arise. These may be risks associated with: (a) prioritising certain error types (such as false positives or false negatives) without considering the context of use and the potential disproportionate impacts of differential error rates on individuals in situations of vulnerability or on groups with protected characteristics; (b) limiting performance assessment to accuracy while omitting other relevant measures (like sensitivity, precision and specificity), which can obscure biases or weaknesses affecting reliability and fairness; (c) presenting metrics in formats that are overly technical, inaccessible or lacking contextual explanation, preventing users and stakeholders from properly interpreting system performance; and (d) irregular reporting or insufficient monitoring that hinder identification of performance variations across population groups or over time.*
	B.6.6	If the system substantially informs or takes decisions that may impact human rights in ways that require mechanisms for human oversight, review or intervention, have such features been integrated at the model design and implementation stages to ensure that the system can support procedural and substantive fairness in its future use? *Where an AI system is capable of substantially informing or taking decisions affecting individuals' rights or access to services, its design and implementation should include features that support procedural and substantive fairness. Such features may include: (a) the capacity for human review or override (both* ex ante *and* ex post*); (b) built-in operational constraints preventing the system from autonomously exceeding its intended scope; (c) traceability, auditability and explainability mechanisms; and (d) clear assignment of oversight responsibilities to individuals with the necessary competence, training and authority. These elements help ensure that, once deployed, the AI system functions within a framework that enables fairness, accountability and effective human control.*
	B.6.7	What potential is there for the AI system to be repurposed or used in ways that could raise additional ethical, legal or regulatory concerns? *Repurposing (scope creep, dual-use, capability escalation) can materially change the system's risk profile (including from the personal data-protection perspective) by shifting context, data uses or decision pathways. The potential for such repurposing often originates from design and development decisions (modular architectures, transferable models or open interfaces). HUDERIA remains relevant when a system is repurposed or its function evolves, as such changes may generate new or amplified impacts on human rights, democracy and the rule of law. In these cases, the assessment should be updated or repeated to account for the new context.* ***Example:*** *a computer vision system used to recognise individuals' faces as a form of identity authentication is repurposed for real-time remote biometric identification of petty criminals at a concert venue.*

COBRA Resources C (List of risk factors arising in the AI system's deployment context)

	Examples of potential risk factors	
1. Privacy and personal data protection	C.1.1	Does the AI system process personal data, including the sensitive category of data (as, for instance, set out in Article 6 of Convention 108+ or other relevant frameworks)? If so, what risks may arise concerning the privacy and data-protection rights of potentially affected persons and the responsibilities of those developing, deploying or using the system in relation to data protection and information governance? *If the AI system processes personal data, there is a risk that such processing may not comply with applicable personal data-protection and privacy laws or standards. Risks include unlawful or unfair collection and use, inadequate safeguards or insufficient accountability in handling personal data and disclosure of personal data in output.*
	C.1.2	Is the AI system designed for individual-targeted curation, profiling, prediction or behavioural steering? If so, what risks may arise, including if potentially affected persons cannot access sufficiently accessible information in this connection? *If the AI system is designed for individual-targeted curation, profiling, prediction or behavioural steering, there is a risk that affected persons may not benefit from the highest level of transparency and information rights available under the applicable legal framework (particularly where sensitive data are involved, as defined in instruments such as Convention 108+). In particular, they may lack (a) clear information on the collection and use of their personal data; (b) the rationale behind the system's outputs, explained in plain, non-technical language; (c) the purpose of the curation, profiling, prediction or behavioural steering; (d) the categories of persons or bodies to whom their data, profile or the results of processing may be communicated.*
2. Non-discrimination and bias	C.2.1	How might the modalities of deployment – including where, how and by whom the AI system is implemented – influence the emergence or amplification of bias or discriminatory effects identified earlier in the life cycle? *Bias and discriminatory patterns identified during design or development may be reinforced or mitigated depending on how the AI system is deployed.*
3. System operators	C.3.1	What potential risks to the human rights of those operating the system may arise from the deployment of the AI system? *The deployment of the AI system may adversely affect the human rights of those operating it (for example by creating excessive surveillance, undue stress, unsafe working conditions or erosion of professional autonomy). Such impacts could compromise their human rights but also the integrity and accountability of the system's operation.*
4. Training	C.4.1	Are system operators sufficiently trained fully to understand the system's limitations and to intervene effectively in situations of complexity, uncertainty, anomaly or failure? If not, what risks to human rights, democracy and the rule of law and system accountability may arise in this connection? *If operators of the AI system are not sufficiently trained, they may fail to understand the system's intended uses and limitations or to recognise conditions of complexity, uncertainty, anomalies or failures. This could prevent them from exercising appropriate human judgment and intervention, increasing the risk of harm to potentially affected persons and undermining the accountability, transparency and reliability of the system.*
5. Human in the loop	C.5.1	Does the AI system have a high level of automation or operational "autonomy"? If so, what risks might arise from interactions with affected persons that are not fully mediated by human control? *AI systems with a high level of automation or operational "autonomy" may interact with potentially affected persons in ways not fully mediated by human control. This creates risks of undermining human dignity and individual autonomy, as well as eroding transparency, oversight, accountability and responsibility. While full replacement of human operators may occur only in certain contexts, there are many scenarios in which human operators remain involved but increasingly over-rely on AI output. This can result in a false or "illusory" sense of meaningful human oversight, where the task is largely executed by the system even though nominal human supervision exists. Without appropriate safeguards, such interactions could adversely affect the human rights of potentially affected persons.*

		Examples of potential risk factors
6. Out-of-the-scope uses	C.6.1	How could malfunctions, misuse or malicious application of the AI system affect the human rights of potentially affected persons, democracy and the rule of law? To what extent are these risks identified? *There is a risk that the AI system may malfunction, be misused, or abused – for example, through breakdown or technical failure, use beyond its intended scope or deliberate malicious misapplication. Such events could have an adverse effect on the human rights of potentially impacted persons, in particular by undermining transparency, oversight, accountability and responsibility of the system. This risk category could also impact individual autonomy and therefore impact democracy and the rule of law when applied at scale, such as disseminating misinformation and disinformation.*
7. Proximity to decision making or action	C.7.1	To what extent does the AI system directly take, review or substantially inform decisions or actions that may affect human rights, democracy or the rule of law? How does this degree of involvement influence the potential level of risk? *The degree of proximity of the AI system to the relevant activity (decision making or action) can amplify or reduce its potential impact on human rights. Where the system directly takes or reviews decisions, or substantially informs them, the risk of adverse human rights impacts may be higher compared to systems that only provide peripheral support or no meaningful influence.*

COBRA Resources E (Illustrative[21] areas of potential concern from the point of view of human rights, democracy and the rule of law)

This section provides a tool that could be used to perform and/or inform the COBRA assessment. It outlines areas of potential concern (left column) with, for each of them, an indicative description, examples of potential AI-related risks, potentially relevant domains and references to potentially relevant human rights instruments and, as appropriate, other legal instruments.

The list may be understood as an illustrative array of elements that can be used to assist the assessment of AI-related risks in different contexts. The resource is indicative, non-exhaustive and open-ended, meaning it is bound to evolve in the light of developments in the areas of technology and relevant legal regimes. Seventeen areas of potential concern have been selected. These areas represent fundamental aspects enshrined in various international human rights instruments, underlining the protection of human dignity and individual autonomy.

[21] References to international human rights instruments in this table are included for illustrative purposes. Those references only apply to states that are parties to those instruments. Each state is expected to apply its own applicable laws in accordance with its international legal obligations, which could include encouraging the private sector to respect and support human rights, including as set out in the United Nations Guiding Principles on Business and Human Rights.

(1) Physical and mental integrity and human dignity	**Illustrative description**
	Human rights provisions provide a number of protections regarding the physical and mental integrity of individuals, reinforcing the importance of personal and individual autonomy and self-ownership. Many of these protections may apply, for example, in situations involving the use of force by authorities exercising law-enforcement powers and where individuals are held in custody. In some countries these protections may include investigations into the conduct of public authorities in health or life-threatening situations. Human rights protections may extend to various other situations, where the physical and mental integrity of individuals may be at stake.
	Examples of potential AI-related risks
	AI risks to personal, physical and mental integrity, as well as human dignity, may stem from issues such as lack of transparency in decision making, faults in the system design (including where such design is not age-appropriate), development or use of AI systems and the potential for AI-driven systems to dehumanise and objectify individuals. Risks may also arise from AI systems that manipulate individuals.
	Potentially relevant human rights obligations
	The following legal provisions may be of relevance: **ICCPR:**[22] Articles 6, 7, 9 and 10 **ICESCR:**[23] Articles 7 (decent work) and 12 (health) **ECHR**[24] Articles 2 (right to life), 3 (prohibition of torture) and 13 (right to an effective remedy) **ESC:**[25] Articles 7 (protection of children and young persons), 8 (protection of maternity), 11 (protection of health), 13 (social and medical assistance), 14 (social welfare systems), 16 (protection of families), 17 (protection of children), 18 (protection of migrant works and their families), 23 (social protection of the elderly), 26 (dignity at work), 30 (protection against poverty and social exclusion) and 31 (housing) **EU Charter:**[26] Articles 1 (human dignity), 3 (right to the integrity of the person) and 4 (prohibition of torture or degrading treatment or punishment) **Pact of San Jose:**[27] Articles 3 (right to juridical personality), 4 (right to life) and 5 (right to humane treatment), and Article 12 (right to food) of the Protocol of San Salvador
(2) Physical liberty and security, movement and residence	**Illustrative description**
	Human rights and rule of law concerns may arise in regard to the physical treatment of individuals in situations of arrest, detention, punishment or human trafficking. Protections exist against arbitrary arrest or detention, torture, cruel, inhuman or degrading treatment or punishment, unjustified restrictions on freedom of movement and human trafficking and to ensure access to remedy and justice.
	Examples of potential AI-related risks
	The collection of personal data and/or use of AI tools may enable or result in the inference of behaviours or activities that are used to justify or facilitate activities such as arrest or detention, or restrictions of movement and residence, even if these inferences are incorrect; such incorrect decisions made or informed by AI tools that are acted upon without sufficient further examination can lead to wrongful treatment or undermine the ability of an accused person to challenge the lawfulness of their detention. AI systems may be used to facilitate or enable human trafficking, for instance by generating content that is used to groom or lure individuals into situations of exploitation. As is the case with human trafficking in general, women and children are at greatest risk. AI may also be used to perform mass surveillance and to identify individuals for no justifiable or reasonably expected purpose.

[22] The United Nations (UN) International Covenant on Civil and Political Rights and its optional protocols.

[23] The UN International Covenant on Economic, Social and Cultural Rights and its optional protocol.

[24] The Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5) and its additional protocols.

[25] The European Social Charter (ETS No. 35) and its protocols and the Revised European Charter (ETS No. 163).

[26] The Charter of Fundamental Rights of the European Union.

[27] The American Convention on Human Rights and its additional protocol.

2) Physical liberty and security, movement and residence	**Potentially relevant human rights obligations**
	The following legal provisions may be of relevance: **ICCPR:** Articles 7, 8, 9, 10, 11, 12, 13 and 14 **ILO Convention 105** (prohibition of forced labour) **ECHR:** Articles 3 (prohibition of torture), 4 (prohibition of slavery and forced labour), 5 (right to liberty and security) and 13 (right to an effective remedy), and Article 2 (freedom of movement) of Protocol No. 4 **ESC:** Article 1(2) (the right to work) **EU Charter:** Articles 3 (right to the integrity of the person), 4 (prohibition of torture or degrading treatment or punishment), 5 (prohibition of slavery and forced labour), 6 (right to liberty and security) and 45 (right to freedom of movement and residence) **Pact of San Jose:** Articles 5 (right to humane treatment), 6 (freedom from slavery), 7 (right to personal liberty) and 22 (freedom of movement and residence)
(3) Justice and administration of justice	**Illustrative description**
	Access to justice and a complex set of minimal rules regulating access to justice in the broader field of justice and public administration. Human rights protections in this area cover various rights and guarantees with respect to access to remedies, the quality of examination of cases by courts (such as providing for a public pronouncement of a judgment), the quality of participation in court proceedings (such as providing for a fair and public hearing) and the requirements in respect of the composition of courts (such as requiring a competent, independent and impartial tribunal). However, in many countries a wide scope of matters is ruled by public administration bodies with citizens having access to courts at a later stage of proceedings with respect to certain matters. Public administration is equipped with decision-making powers that allow them to significantly influence individuals and their important life interests of this connection.
	Examples of potential AI-related risks
	The use of AI in the justice system introduces risks such as harmful bias, lack of transparency or insufficient human oversight, the use of flawed data (which is particularly problematic in the context of the administration of justice) and the erosion of judicial independence and the loss of public confidence in justice institutions. It also raises concerns about due process infringements that could result in violations of human rights.
	Potentially relevant human rights obligations
	The following legal provisions may be of relevance: **ICCPR:** Articles 2.3, 9, 14 and 26 **ECHR:** Articles 5 (right to liberty and security) and 6 (right to a fair trial), Article 8 (right to respect for private and family life), Article 13 (right to an effective remedy) and Article 3 of Protocol No. 7 (compensation for wrongful conviction) **EU Charter:** Articles 41 (right to good administration), 42 (right of access to documents), 47 (right to an effective remedy and to a fair trial) and 48 (presumption of innocence and right of defence) **Pact of San Jose:** Articles 8 (right to a fair trial), 10 (right to compensation), 24 (right to equal protection) and 25 (right to judicial protection)

(4) Privacy and data protection

Illustrative description

Privacy and data-protection rights ensure limits to outside influence over private affairs and that people can keep their personal data, behaviours and decisions from being disclosed or monitored without their consent, safeguarding personal autonomy and dignity.

Specifically in the public-sector context, privacy and data-protection laws protect individuals' personal information by ensuring that public authorities or entities acting on their behalf only process personal information that they are authorised to.

Privacy protections also allow individuals to maintain control over how they grow, define and present their personal identity. These protections protect against unauthorised use or misrepresentation of an individual's personal attributes, such as name, likeness or other distinguishing characteristics.

There are also aspects of privacy and individual autonomy that relate to interaction with others. Human rights protections in this area cover various rights and guarantees with respect to family life, marriage and, more generally, relationships with others, both privately and in a work setting.

Examples of potential AI-related risks

The collection of vast amounts of personal data (even when such data are collected transparently) as well as the use of AI techniques to infer personal information from it can create serious risks to both privacy and identity. From invasive surveillance and re-identification of anonymised data to misuse of biometrics, creation of deepfakes, identity theft and behavioural manipulation, AI technologies have the ability to undermine personal autonomy and expose sensitive information.

Potentially relevant human rights obligations

The following legal provisions may be of relevance:

ICCPR: Article 17

UNCRC[28]: Article 16

ECHR: Articles 8 right to respect for private and family life), 10 (freedom of expression) and 12 (right to marry)

EU Charter: Articles 7 (respect for private and family life), 8 (protection of personal data) and 9 (right to marry and right to found a family)

Convention 108+[29]: Articles 3, 5, 6 and 9

Pact of San Jose: Article 11 (right to privacy), 17 (rights of the family), 18 (right to a name) and 20 (right to nationality)

[28] The United Nations Convention on the Rights of the Child and its optional protocols.

[29] The Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, as amended (ETS No.108, CETS No 223) and its protocols.

(5) Equality and non-discrimination[30]

Illustrative description

Rights to equal protection under the law and to not be discriminated against in the enjoyment of rights and freedoms complement the other substantive provisions of various international human rights instruments.

Discrimination may be direct or indirect, the result of association, based on one or more grounds, or based on an action or a failure to act.

For general reference, protected characteristics referred to in the body of the existing human rights law consisting of international (at both global and regional levels) and domestic legal instruments, as applicable in each state, may include but are not limited to:

(1) Sex

(2) "Race"[31] and colour

(3) Language

(4) Religion

(5) Political or other opinion

(6) National or social origin

(7) Association with a national minority

(8) Property

(9) Birth

(10) Age

(11) Gender identity and expression

(12) Sexual orientation

(13) Sex characteristics

(14) Health and disability

(15) Parental and marital status

(16) Immigration status

(17) Status related to employment.

Respect for equality may go beyond prohibiting less favourable treatment without justification based on one or more protected characteristics.

Examples of potential AI-related risks

Discriminatory trends can be created or exacerbated by: the purpose of a system if that purpose is itself discriminatory;

datasets that are not sufficiently representative or that encode historical inequalities; algorithmic systems that do not sufficiently account for patterns of bias in datasets or that otherwise generate biased outputs; inadequate testing and evaluation related to bias; a lack of representation or consideration (irrespective of whether this takes place negligently or intentionally) of the perspectives of diverse groups during AI development or other AI activities; insufficient training and other governance measures (such as inadequate oversight processes) for developers or users of AI systems in relation to potential discrimination.

At the same time, a system trained with theoretically completely unbiased data may still result in unfair downstream impacts for reasons relating to how they are deployed.

[30] Equality and non-discrimination provisions complement other substantive provisions of various international human rights instruments and these issues are therefore relevant in respect of all areas of potential concerns in this table.

[31] Since all human beings belong to the same species, theories based on the existence of different "races" are rejected. However, the term "race" is used in order to ensure that those persons who are generally and erroneously perceived as "belonging to another race" are not excluded from the protection provided.

(5) Equality and non-discrimination	**Potentially relevant human rights obligations**
	The following legal provisions may be of relevance: **ICCPR**: Articles 2, 3, 23, 24 and 26 **ICESCR**: Articles 2.2 and 3 **ILO Conventions No. 111** (discrimination), **No. 190** (violence and harassment in the workplace) **ECHR**: Article 14 (prohibition of discrimination), Protocol No. 12, Article 1 **ESC**: Articles 1(2), 4(3), 19(4-5) and 20 **EU Charter**: various provisions within Chapter III on equality (Articles 20-26), Article 14 (right to education), Article 15 (freedom to choose an occupation and right to engage in work) and Article 16 (freedom to conduct a business) **Pact of San Jose**: Article 24 (right to equal protection) and dedicated legal instruments such as **ICERD**[32], **CEDAW**[33], **UNCRC** and **UNCRPD**[34]
(6) Thought, conscience, religion and belief	**Illustrative description**
	Human rights provisions protect an individual's ability to hold, practice and express their religious beliefs or the lack thereof without interference or discrimination. These include protections for the ability to worship, to change one's religion or beliefs, and to observe religious practices (such as dress, dietary restrictions or religious holidays) publicly or privately. They also protect against coercion to adopt or renounce any religion and ensure equal treatment regardless of religious affiliation, whether it is real or supposed. Freedom of conscience protects the ability to hold and act upon closely held beliefs of right and wrong, not necessarily based in religious systems. Examples could include atheism, vegetarianism or conscientious objection to military service. Although there is some debate over the content of freedom of thought, it could include not having to reveal one's thoughts, not having punishment for one's thoughts and protection from impermissible alteration of thought.
	Examples of potential AI-related risks
	AI-related risks to religious rights can stem from algorithms that create or perpetuate harmful bias that affects people on the basis of their actual or perceived religious beliefs, through privacy violations or by facilitating religious profiling, suppression of religious expression and manipulation of beliefs. These risks threaten individuals' ability to practise and express their religion without fear of discrimination, surveillance or coercion. Additionally, AI systems, such as emotion recognition systems and systems that perpetuate disinformation, could interfere with freedom of thought or conscience and could impose a chilling effect that could interfere with freedom of thought or conscience. The use of AI systems can compromise individual autonomy in matters of belief by subtly shaping perceptions, influencing thought processes and potentially restricting the freedom to form and express religious or spiritual convictions without undue external interference.
	Potentially relevant human rights obligations
	The following legal provisions may be of relevance:
	ICCPR: Articles 18, 19 and 20
	ECHR: Article 9 (freedom of thought, conscience and religion), Article 10 (freedom of expression) and Article 14 (prohibition of discrimination)
	EU Charter: Article 10 (freedom of thought, conscience and religion), Article 11 (freedom of expression and information) and Chapter III on equality
	Pact of San Jose: Article 12 (freedom of conscience and religion) and Article 13 (freedom of thought and expression)
	UNCRC: Article 14

[32] The United Nations International Convention on the Elimination of All Forms of Racial Discrimination.

[33] The United Nations Convention on the Elimination of All Forms of Discrimination Against Women and its optional protocol.

[34] The United Nations Convention on the Rights of Persons with Disabilities and its optional protocol.

(7) Opinions, expression and information	**Illustrative description** Protections for opinions and free expression apply to holding, seeking, imparting and receiving information and ideas through any media. Forms of expression may be political in nature, artistic, including the production of plays and performances, personal or commercial and may include news reporting, photographs and forms of conduct, such as boycotts or campaigns, clothing, symbols, etc. These protections may also apply in certain relations governed by the rule of private law (such as labour relations) and statements made in private correspondence or meetings behind closed doors.
	Examples of potential AI-related risks AI-related risks to opinions, expression and access to information may include potential censorship, bias in algorithmic systems used for content selection and the creation and promulgation of inaccurate information. Additionally, AI-facilitated surveillance can stifle free speech, while the use of algorithms can result in reduced access to all viewpoints. The use of AI systems can undermine individual autonomy in expression by subtly influencing opinions, shaping discourse and limiting the ability to freely form, articulate and share ideas without covert manipulation or constraint. Based on how an individual expresses themselves publicly, an AI system can make inferences that are incorrect and this could alter the information or access that the individual may otherwise benefit from (for example, an individual wearing certain clothes could be categorised as belonging to a group they do not).
	Potentially relevant human rights obligations The following legal provisions may be of relevance: **ICCPR:** Articles 19 and 20 **UNCRC:** Article 13 **ECHR:** Article 10 (freedom of expression) **EU Charter:** Articles 10 (freedom of thought, of conscience and of religion), 11 (freedom of expression and information) and 42 (right of access to documents) **Pact of San Jose:** Articles 13 (freedom of thought and expression) and 14 (right to reply)
8) Peaceful assembly and association	**Illustrative description** Human rights protections exist in respect of non-violent gatherings and walkabouts, including meetings in private and public places. There are protections for voluntary groupings for a common goal and the possibility of forming or being affiliated with a group or organisation pursuing particular aims. Prominent examples of such associations are political parties, minority and religious associations and trades unions.
	Examples of potential AI-related risks Potential AI risks in this area may include arbitrary or discriminatory surveillance of members of an association and meeting participants, predictive policing, social media censorship, misinformation and targeted harassment, all of which can deter individuals from exercising their rights to peacefully gather and associate freely and participate in government.
	Potentially relevant human rights obligations The following legal provisions may be of relevance: **ICCPR:** Article 21 **UNCRC:** Article 15 **ECHR:** Article 11 (freedom of assembly and association) **ESC:** Articles 5, 6.1-2, 21, 28 and 29 **EU Charter:** Articles 12 (freedom of assembly and of association) and 28 (right to collective bargaining and action) **Pact of San Jose:** Articles 15 (right of assembly) and 16 (freedom of association)

(9) Property	**Illustrative description**
	Human rights protections in this area are generally not limited to covering the ownership of physical goods, but may extend other rights and interests, including intellectual property rights, constituting assets that can also be regarded as "property rights".
	Examples of potential AI-related risks
	Property rights may be adversely affected by AI systems through such issues as: ▸ the use of flawed algorithms in finance and real estate to make automated decisions regarding property values, loans or credit; ▸ AI-generated content, such as artwork, music or text, and the development of AI models may raise complex issues related to intellectual property rights; ▸ without human oversight, rapid advancements in AI technology could outpace existing legal frameworks, leaving gaps in property rights protection; ▸ the use of intellectual property to train models where laws are unclear or inconsistently applied regarding the question of whether such use is permitted. These risks can undermine individuals' control over their possessions and intellectual creations, leading to potential exploitation and discrimination.
	Potentially relevant human rights obligations
	The following legal provisions may be of relevance: **ECHR:** Article 1 of Protocol No. 1 (protection of property) **EU Charter:** Article 17 (right to property) **Pact of San Jose:** Article 21 (right to property)
(10) Education	**Illustrative description**
	Human rights protections in this area relate to individuals' access to and the provision of quality education without discrimination. Protections may encompass not only access to primary education but also secondary, higher, vocational and lifelong learning opportunities.
	Examples of potential AI-related risks
	AI-related risks include harmful bias produced by algorithms in areas such as admissions, grading and personalised learning. Other AI-related risks (including those resulting from the lack of digital literacy) can undermine the effectiveness and fairness of educational, vocational and training systems, impacting the overall learning experience for students or labour market opportunities.
	Potentially relevant human rights obligations
	The following legal provisions may be of relevance: **ICESCR:** Article 13 **UNCRC:** Article 28 **ECHR:** Article 2 of Protocol No. 1 (right to education) **ESC:** Articles 1(4), 7(6), 9, 10, 15, 17, 19 (11-12), 20(b) and 30 **EU Charter:** Article 14 (right to education) **Pact of San Jose:** Article 13 (right to education) of the Protocol of San Salvador

(11) Arts, sciences, culture and language

Illustrative description

Human rights protections in this sphere allow access for all to a variety of cultural resources, the enjoyment and the benefits of scientific progress and its applications, and the protection of moral and material interests resulting from any scientific, literary or artistic production of which they are the author. Likewise, these protections cover the freedom of scientific research and creative activity. They also protect the ability of individuals, including members of ethnic, religious or linguistic minorities, to enjoy their culture, practise their religion and use their language.

Examples of potential AI-related risks

AI-related considerations for culture, arts and sciences may include complex issues related to intellectual property issues, data accuracy and quality control issues, unreliable sources of data and risks to critical thinking and ethical research practices. These risks can undermine the richness and diversity of human expression, creativity and knowledge advancement.

Potentially relevant human rights obligations

The following legal provisions may be of relevance:

ICCPR: Article 27

ICESCR: Article 15

ESC: Articles 15(3), 22(c), 23 and 30.

EU Charter: Articles 13 (freedom of the arts and sciences) and 22 (cultural, religious and linguistic diversity)

Pact of San Jose: Article 14 (right to the benefits of culture) of the Protocol of San Salvador

(12) Labour and employment

Illustrative description

Labour rights protect workers in the workplace in relation to fair treatment, safe and healthy working environments, healthy, safe and decent working conditions, reasonable wages, privacy in the workplace, decent standards of living, freedom from discrimination, fair and equitable treatment and the ability to form and join trades unions.

Labour rights may include, for example, minimum standards of wages and decent standards of living, limits on working hours, protection from forced labour, issues related to child labour, freedom of association and the effective recognition of the right to collective bargaining and elimination of discrimination.

Examples of potential AI-related risks

AI may have a variety of transformative impacts on labour and employment.

AI-related risks include the use of AI to monitor and manage employees, leading to concerns over privacy and creating stressful environments due to constant tracking of performance, behaviour, productivity and potentially AI-based job displacement.

AI systems used to inform or make decisions regarding hiring, firing or promotions risk biased, opaque or unfair labour practices, including those affecting people working on AI.

AI can be used to automate and optimise supply chains, which might inadvertently increase the demand for cheap labour, including child labour in certain sectors. The use of AI in monitoring and managing labour can reduce human oversight, potentially allowing exploitative practices to go unnoticed, particularly regarding child labour.

(12) Labour and employment

Potentially relevant human rights obligations

The following legal provisions may be of relevance:

ICESCR: Articles 6, 7, 8 and 10

UNCRC: Article 32

ESC: Articles 1-10, 21, 22, 26, 28 and 29

EU Charter: Articles 8 (personal data protection), 12 (freedom of assembly and association), 15 (freedom to choose an occupation and right to engage in work), 27 (workers' right to information and consultation within the undertaking), 28 (right of collective bargaining and action), 29 (right to access to placement services), 30 (protection in the event of unjustified dismissal), 31 (fair and just working conditions), 32 (prohibition of child labour and protection of young people at work) and 33 (family and professional life)

Pact of San Jose: Articles 7 (just, equitable and satisfactory conditions of work) and 8 (trades union rights) of the Protocol of San Salvador

Additionally, the following International Labour Organization conventions are of potential relevance in this context:

1. Convention No. 87 – Freedom of Association and Protection of the Right to Organise (1948)

2. Convention No. 98 – Right to Organise and Collective Bargaining (1949)

3. Convention No. 111 – Discrimination (Employment and Occupation) (1958)

4. Convention No. 138 – Minimum Age Convention (1973)

4. Convention No. 155 – Occupational Safety and Health (1981)

5. Convention No. 190 – Violence and Harassment (2019)

(13) Health, healthcare and social security/social protection

Illustrative description

The health sector encompasses access to medical services, public health initiatives and healthcare infrastructure to ensure the physical and mental well-being of individuals. Protections in this area generally relate to the ability of individuals to access healthcare in order to enjoy the highest attainable standard of physical and mental health. The social security sector provides access to financial support and protection to individuals facing unemployment, disability, old age or economic hardship through benefits and welfare programmes.

Protections in this area generally relate to the ability of individuals to receive such assistance.

Social protection, social assistance and protection against poverty and social exclusion aim to exclude or reduce poverty and inequality and to promote human dignity.

Examples of potential AI-related risks

The use of AI systems may result in harmful biased data leading to unequal access to healthcare or social benefits, particularly for marginalised groups.

The use of AI in processing sensitive personal health and social data raises significant privacy and security risks (including issues relating to consent and individual autonomy), including data breaches or misuse.

Over-reliance on AI for medical diagnoses or social benefit decisions could lead to errors, reducing human oversight and potentially harming patient outcomes or wrongly denying benefits.

Regarding social protection and social security, the use of AI may present risks when it is used to determine the eligibility of persons for social benefits or to detect cases where social benefits were paid out incorrectly and the authority has repayment claims: consequences of mistakes (possibly caused by biases) can be severe and often affect persons in vulnerable situations in particular.

Potentially relevant human rights obligations

The following legal provisions may be of relevance:

ICESCR: Articles 9, 10 and 12

ESC: Articles 2, 3, 7, 8, 11, 12, 13, 14, 15, 17(1), 19(2), 22(b), 23, 27(1)(b) and 30

European Code of Social Security (ECSS): Articles 7-12 (medical care), 13-18 (sickness benefit), 19-24 (unemployment benefit), 25-30 (old-age benefit), 31-38 (unemployment injury benefit), 46-52 (maternity benefit), 53-58 (invalidity benefit) and 59-64 (survivor's benefit)

UNCRC: Article 24

EU Charter: Articles 8 (personal data protection), 34 (social security and social assistance) and 35 (healthcare)

Pact of San Jose: Articles 9 (right to social security) and 10 (right to health) of the Protocol of San Salvador

(14) Children

Illustrative description

Both general human rights and children-specific protections exist in respect of this group. They are aimed at ensuring that children can grow, learn, play, develop and flourish with dignity and that children's specific vulnerabilities and needs are properly considered. In all actions concerning children, their best interests must be a primary consideration. With the above considerations in mind, many of the areas of potential concern presented in this table will be relevant for children.

Examples of potential AI-related risks

AI systems could pose risks to children if used in making decisions that may not take into account their needs or rights, and by creating fake but realistic images or videos that can confuse, mislead or harm children. Content that is generated, selected or recommended by AI systems (for instance) on social media or video platforms can expose children to content that is illegal, harmful or not age-appropriate. Targeted advertising and AI-based content can exploit children's vulnerabilities, influencing their behaviour and preferences, including for commercial gain.

AI-driven social media platforms can contribute to anxiety, depression and negative self-esteem in children through cyberbullying, addiction and harmful comparisons. Children are especially susceptible to harm linked to generative AI. Children are less capable of discerning synthetic content from genuine content, identifying inaccurate information and when they are being manipulated by dark patterns, and understanding that they are interacting with a machine rather than with a human being. AI is increasingly used to generate child sexual abuse material (CSAM) and exploit children. Generative AI can enable the creation of synthetic CSAM and cause harm to children and families, as well as to society. Predators also can use generative AI to groom, extort and exploit victims.

Potentially relevant human rights obligations

ICESCR: Articles 10(3) and 12

ESC: Articles 7 and 17

ECSS: Articles 39-45 (family benefit)

EU Charter: Articles 7, 8, 24 and 32

Pact of San Jose: Article 19 (rights of the Child)

In addition to human rights provisions contained in international human rights treaties, the following specialised human rights instruments may be of relevance:

1) the 1989 United Nations Convention on the Rights of the Child (UNCRC) and its optional protocols;

2) the General comment No. 25 on children's rights in relation to the digital environment;

2) the 2007 Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS No. 201, the Lanzarote Convention).

(15) Environment

Illustrative description

In some jurisdictions, human rights provisions may recognise some level of responsibility of public authorities for protecting individuals from environmental harm.

Examples of potential AI-related risks

AI systems, particularly those that use significant computational resources for training and inference, consumes large amounts of energy. This contributes to carbon emissions, especially when powered by non-renewable energy sources. Likewise, the production of hardware required for AI system's training and running requires many rare earth materials, the extraction of some of which has wider pollution effects.

AI applications in agriculture, the economy, resource management or wildlife monitoring, while offering many potential environmental benefits, may also have unforeseen environmental impacts if not carefully managed, such as disrupting ecosystems, over-exploiting natural resources, pollution or harming biodiversity.

(15) Environment	**Potentially relevant human rights obligations**
	If seen through a human rights angle, issues regarding environmental protection could be addressed under the following provisions: **ICESCR** Article 12 **ECHR:** Articles 2 (right to life), 3 (prohibition of torture), 6 (right to fair trial), 8 (right to private and family life), 10 (freedom of expression) and 11 (freedom of assembly and association) **EU Charter:** Article 37 (environmental protection) **Pact of San Jose:** Article 11 (right to a healthy environment) of the Protocol of San Salvador **Aarhus Convention**[35]
(16) Democracy	**Illustrative description**
	Human rights provisions protect the ability of individuals to participate in the conduct of public affairs, to vote in free and fair elections, to run for public office and to have access to government services. An important feature of democratic systems of government is political pluralism, which is ensured in large part by the protection of human rights, the respect of which is essential for a thriving democracy, such as freedom of expression, freedom of association and freedom of peaceful assembly; and by the existence of pluralist and independent media and a range of political parties representing different interests and views, fair access to and meaningful participation in public debate and public decision making and access to accurate and trustworthy information. Rights governing meaningful participation in public decision-making processes and in government ensure that individuals have the ability to engage in the political process, including to vote, run for office and express opinions on public policies. Fostering democratic governance, accountability and representation allows citizens to influence decisions that affect their lives and communities.
	Examples of potential AI-related risks
	AI systems playing a role in influencing or informing the democratic processes could adversely affect the fair access of individuals to and participation in public debate and free and fair elections, as well as their ability to freely form opinion through the following, which is a non-exhaustive list: (a) deception, misinformation or disinformation at local, national or global levels caused by the deployment of the system; (b) manipulation at local, national or global levels enabled by the deployment of the system; (c) intimidation or behavioural control, at local, national or global levels enabled by the deployment of the system. AI could be used in ways that interfere with participation in government, for example through use of algorithms to screen political candidates or participants in public processes or to suppress speech on matters relating to public policy.
	Potentially relevant human rights obligations
	The following legal provisions may be of relevance: **ICCPR:** Articles 19, 20, 21 and 25 **ECHR:** Articles 10 (freedom of expression) and 11 (freedom of assembly and association) and Article 3 of Protocol No. 1 (right to free elections) **ESC:** Articles 5 and 6.1-2 **EU Charter:** Articles 11 (freedom of expression and information), 42 (right of access to documents), 12 (freedom of assembly and of association), 39 (right to vote and to stand as a candidate in elections to the European Parliament) and 40 (right to vote and to stand as a candidate in municipal elections) **Pact of San Jose:** Articles 13 (freedom of thought and expression), 14 (right to reply), 15 (right of assembly), 16 (freedom of association) and 23 (right to participate in government)

[35] The UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters.

(17) Rule of law

Illustrative description

Rule of law is a principle of governance in which all persons, institutions and entities, public and private, including the state itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with respect for international human rights. Judicial independence is foundational to democracy and ensuring public confidence in the administration of justice.

Rule of law institutions are organisations (the judiciary, legislature and executive, legal profession and semi-independent bodies like anti-corruption bodies, data-protection authorities, ombudsmen, etc.) and systems (legal systems) that ensure laws are applied fairly, consistently and transparently, protecting individuals' rights and maintaining order in society. Together, these institutions maintain the rule of law by ensuring that laws govern society, not arbitrary decisions, and that everyone is treated equally under the law.

Examples of potential AI-related risks

The use of AI presents risks (through possible lack of transparency and sufficient human oversight, the use of systems resulting in unintended consequences, harmful bias or cybersecurity threats, among other things) to the rule of law, procedural fairness, transparency and accountability in contexts such as:

(a) the integrity of democratic institutions and processes, including the principle of the separation of powers, and respect for judicial independence;

(b) access to justice;

(c) accountability mechanisms (oversight of the executive branch) and anti-corruption bodies and policies.

Potentially relevant human rights obligations

The effective exercise and protection of these rights are essential for ensuring respect for the rule of law.

ICCPR: Articles 2.3, 4 and 14

ECHR: Articles 6 (right to a fair trial) and 13 (right to an effective remedy)

EU Charter: Articles 41 (right to good administration), 42 (right of access to documents), 47 (right to an effective remedy and to a fair trial) and 48 (presumption of innocence and right of defence)

Pact of San Jose: Articles 8 (right to a fair trial), 10 (right to compensation) and 25 (right to judicial protection)

COBRA Resources F (Illustrative[36] list of sectors/domains and potential areas of concern)

This resource lists (not in order of priority) the sectors and domains of potential concern from the perspective of human rights, democracy and the rule of law (first and second column) and is intended to be used in conjunction with **COBRA Resources E**, in order to carry out the mapping of impacts as part of COBRA. Similarly, it is non-exhaustive and should be regularly reviewed.

	Domains	Areas of potential concern based on COBRA resources E
1. Public administration	**(a) Healthcare**, including, but not limited to, issues such 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.	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (13) Health, healthcare and social security/social protection; (14) Children; (15) Environment.
	(b) Family life and social care, including, but not limited to, issues such as mutual enjoyment of parents with children, custody, access, contact rights, state care, foster families, adoption and reproductive services, access to and provision of public benefits.	(1) Physical and mental integrity and human dignity; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (9) Property; (13) Health, healthcare and social security/social protection; (14) Children.
	(c) Migration and border control, including, but not limited to, issues such 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 methods of entry to and removal from the territory of the state.	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (12) Labour and employment; (13) Health, healthcare and social security/social protection; (14) Children; (15) Environment; (17) Rule of law.
	(d) Infrastructure development and maintenance, including, but not limited to, issues such as health security and enjoyment of public space, public transportation and mobility, management of environmental hazards, land and urban planning, housing, digital infrastructure, energy management and energy consumption.	(1) Physical and mental integrity and human dignity; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (9) Property; (13) Health, healthcare and social security/social protection; (14) Children; (15) Environment.
	(e) Emergency services, including, but not limited to, issues such as management of rescue operations, emergency communications infrastructures and management of the aftermaths of disasters.	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (14) Children; (15) Environment.
	(f) Public education, including, but not limited to, issues such as access to educational institutions and educational assessments, and official recognition of studies.	(5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (10) Education; (14) Children; (11) Arts, sciences, culture and language.
	(g) Employment, including but not limited to, issues such as recruitment, access to employment, performance management and worker policies, accessibility and reasonable accommodation of persons with disabilities.	(1) Physical and mental integrity and human dignity; (4) Privacy and data protection; (5) Equality and non-discrimination; (8) Peaceful assembly and association; (12) Labour and employment; (13) Health, healthcare and social security/social protection.

[36] References to international human rights instruments in this table are included for illustrative purposes. Those references only apply to states that are parties to those instruments. Each state is expected to apply its own applicable laws in accordance with its international legal obligations, which could include encouraging the private sector to respect and support human rights, including as set out in the United Nations Guiding Principles on Business and Human Rights.

	Domains	Areas of potential concern based on COBRA resources E
2. Law enforcement and security	**(a) Police and assimilated services**, including, but not limited to, issues such 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 (such as 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, surveillance and restrictions by police and other law-enforcement agencies.	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (7) Opinions, expression and information; (8) Peaceful assembly and association; (9) Property; (14) Children; (15) Environment; (17) Rule of law.
	(b) Prosecutions, including, but not limited to, issues such as collection and assessment of evidence.	(1) Physical and mental integrity and human dignity; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (7) Opinions, expression and information; (14) Children; (15) Environment; (17) Rule of law.
3. Administration of justice	**(a) Courts and justice**, including, but not limited to, issues such as arrests, detentions, decisions regarding bail, release on parole, conditional release and wearing of electronic bracelets, notification of rights and decisions, 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.	(2) Physical liberty and security, movement and residence; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (7) Opinions, expression and information; (9) Property; (14) Children; (17) Rule of law.
	(b) Institutional aspects of organisation of the judiciary, namely respect for judicial independence and including, but not limited to, issues such 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.	(3) Justice and administration of justice; (5) Equality and non-discrimination; (9) Property; (12) Labour and employment; (17) Rule of law.
4. Democratic processes	**(a) Electoral system**, including, but not limited to, issues such 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.	(4) Privacy and data protection; (5) Equality and non-discrimination; (7) Opinions, expression and information; (8) Peaceful assembly and association; (16) Democracy.
	(b) Institutions and political processes, including, but not limited to, issues such as the supremacy of the constitution, the role of the judiciary in the balancing of powers, judicial independence, delegation of the legislative function.	(3) Justice and administration of justice; (5) Equality and non-discrimination; (7) Opinions, expression and information; (16) Democracy; (17) Rule of law.

	Domains	Areas of potential concern based on COBRA resources E
4. Democratic processes	**(c) Opinions and public discourse**, including, but not limited to, issues such as expression of protected speech in various forms, protection of journalistic sources, information gathering activities, access collection and automated processing of data, property, research and investigation activities, disclosure regime concerning information received in confidence, protection of whistle-blowers, participatory democracy and public consultations, including issues concerning the organisation of committee meetings.	(3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (8) Peaceful assembly and association; (10) Education; (11) Arts, sciences, culture and language; (14) Children; (16) Democracy; (17) Rule of law.
	(d) Peaceful assembly and association, including, but not limited to, issues such as time, place and manner of conduct of assemblies, conditions and modalities of forming or being affiliated with a group or organisation pursuing particular aims, surveillance of assemblies and identification of participants, and participation of citizens in public life.	(5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (8) Peaceful assembly and association; (9) Property; (11) Arts, sciences, culture and language; (14) Children; (16) Democracy.
	(e) Access to information, including, but not limited to, issues such as access to personal information, 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.	(3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (9) Property; (10) Education; (11) Arts, sciences, culture and language; (14) Children; (16) Democracy; (17) Rule of law.
	(f) Media, including but not limited to, issues such 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, media platforms, mis/disinformation, online content moderation.	(5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (8)Peaceful assembly and association; (10) Education; (11) Arts, sciences, culture and language; (14) Children; (16) Democracy.
5. Prison and probation	**(a) Management of prisons and detention facilities**, including, but not limited to, issues such as prisoner profiling, psychological screening of potentially vulnerable inmates, management of dangerous prisoners, management of the prison population, searches of visitors and inmates, and surveillance of communications.	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (13) Health, healthcare and social security/social protection; (15) Environment.
	(b) Parole and probation services, including, but not limited to, issues such as release on parole, conditional release, monitoring of individuals and any electronic wearable devices.	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination.
6. Essential services offered by the private sector	(a) Communication services	(4) Privacy and data protection; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (9) Property; (10) Education; (11) Arts, sciences, culture and language; (14) Children.
	(b) Education and vocational training	(5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (7) Opinions, expression and information; (10) Education; (11) Arts, sciences, culture and language; (12) Labour and employment; (14) Children.

	Domains	Areas of potential concern based on COBRA resources E
6. Essential services offered by the private sector	(c) Biomedical applications, life sciences, epidemiology and healthcare	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (4) Privacy and data protection; (5) Equality and non-discrimination; (9) Property; (13) Health, healthcare and social security/social protection; (14) Children; (15) Environment.
	(d) Environmental and waste management	(1) Physical and mental integrity and human dignity; (4) Privacy and data protection; (5) Equality and non-discrimination; (13) Health, healthcare and social security/social protection; (15) Environment.
	(e) Energy management	(5) Equality and non-discrimination.
	(f) Urban infrastructure and planning	(4) Privacy and data protection; (5) Equality and non-discrimination; (9) Property; (14) Children; (15) Environment.
	(g) Manufacturing and industrial automation	(1) Physical and mental integrity and human dignity; (12) Labour and employment; (15) Environment.
	(h) Construction and building	(4) Privacy and data protection; (5) Equality and non-discrimination; (12) Labour and employment; (15) Environment.
	(i) Security and public safety	(1) Physical and mental integrity and human dignity; (2) Physical liberty and security, movement and residence; (3) Justice and administration of justice; (4) Privacy and data protection; (5) Equality and non-discrimination; (8) Peaceful assembly and association; (15) Environment.
	(j) Domotics (smart home technologies)	(4) Privacy and data protection; (5) Equality and non-discrimination; (14) Children.
	(k) Housing and social accommodation provision	(1) Physical and mental integrity and human dignity; (5) Equality and non-discrimination; (13) Health, healthcare and social security/social protection.
	(l) Employment, human resources and labour management, including, but not limited to, issues such as recruitment, access to employment, performance management and worker policies, accessibility and reasonable accommodation of persons with disabilities.	(2) Physical liberty and security, movement and residence; (4) Privacy and data protection; (5) Equality and non-discrimination; (10) Education; (12) Labour and employment.
	(m) Financial services	(4) Privacy and data protection; (5) Equality and non-discrimination; (9) Property.
	(n) Information technology and networks	(4) Privacy and data protection; (5) Equality and non-discrimination; (7) Opinions, expression and information; (9) Property; (10) Education; (11) Arts, sciences, culture and language; (14) Children.
	(o) Vehicle manufacturing and transportation infrastructure	(2) Physical liberty and security, movement and residence; (5) Equality and non-discrimination; (15) Environment.
	(p) Agriculture and food supply	(1) Physical and mental integrity and human dignity; (5) Equality and non-discrimination; (6) Thought, conscience, religion and belief; (12) Labour and employment; (13) Health, healthcare and social security/social protection; (15) Environment.