HUDERIA
Méthodologie et modèle

Le Comité des Ministres
a approuvé le modèle HUDERIA
le 25 février 2026
et la méthodologie HUDERIA
le 26 février 2025

Conseil de l'Europe

Édition anglaise:
HUDERIA – Methodology and Model

Les points de vue exprimés dans cet ouvrage n'engagent que le ou les auteurs et ne reflètent pas nécessairement la ligne officielle du Conseil de l'Europe.

La reproduction d'extraits (jusqu'à 500 mots) est autorisée, sauf à des fins commerciales, tant que l'intégrité du texte est préservée, que l'extrait n'est pas utilisé hors contexte, ne donne pas d'informations incomplètes ou n'induit pas le lecteur en erreur quant à la nature, à la portée et au contenu de ce texte. Le texte source doit toujours être cité comme suit : « © Conseil de l'Europe, année de publication ». Pour toute autre demande relative à la reproduction ou à la traduction de tout ou partie de ce document, veuillez vous adresser à la Division publications et identité visuelle (DPIV), Conseil de l'Europe (F-67075 Strasbourg Cedex), ou publishing@coe.int.

Toute autre correspondance relative à ce document doit être adressée au Secrétariat du Comité directeur pour les technologies numériques nouvelles et émergentes (CDNET), Gouvernance numérique et sport, Conseil de l'Europe, F-67075 Strasbourg Cedex, courriel : secretariat.cdnet@coe.int

Photo : ©Shutterstock

Conception de la couverture et mise en page : Division publications et identité visuelle (DPIV), Conseil de l'Europe.

© Conseil de l'Europe, février 2026
Imprimé dans les ateliers du Conseil de l'Europe

HUDERIA a été créé par le **Comité sur l'intelligence artificielle (CAI)** du Conseil de l'Europe pour soutenir **une évaluation cohérente et pratique des risques et impacts des systèmes d'IA** du point de vue des **droits humains, de la démocratie et de l'État de droit**. S'appuyant sur des concepts établis de risques pour les droits humains, il traduit ces idées en directives adaptées aux réalités du cycle de vie de l'IA, aidant les acteurs publics et privés à identifier, évaluer, prévenir et atténuer les risques liés aux divers usages de l'intelligence artificielle. HUDERIA combine des directives de haut niveau et neutres sur le plan technologique (la **méthodologie HUDERIA**) avec des supports pratiques et adaptables (le **modèle HUDERIA : ressources d'analyse des risques fondées sur le contexte (COBRA)**), offrant des outils et des recommandations évolutives pouvant être ajustés en fonction de différents contextes, besoins et niveaux de capacité. **En tant qu'instrument non juridiquement contraignant**, pour les Parties à la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, HUDERIA offre une référence flexible qui peut être utilisée ou adaptée – en tout ou en partie – conjointement avec d'autres cadres ou outils, afin de soutenir les approches d'évaluation des risques et des impacts conformes à la Convention-cadre.

Table des matières

PARTIE I – MÉTHODOLOGIE POUR L'ÉVALUATION DES RISQUES ET DES IMPACTS DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE DU POINT DE VUE DES DROITS HUMAINS, DE LA DÉMOCRATIE ET DE L'ÉTAT DE DROIT **5**

INTRODUCTION **7**

Qu'est-ce que l'HUDERIA ? 7

Relation avec la Convention-cadre 7

Principaux objectifs de l'HUDERIA 7

QUELLE EST L'APPROCHE DE L'HUDERIA ? **9**

Approche socio-technique 9

Orientation générale et spécifique 9

Adaptabilité et flexibilité 9

Approche graduée et différenciée 9

CONTOURS DE L'HUDERIA **10**

1. L'ANALYSE DES RISQUES FONDÉE SUR LE CONTEXTE (COBRA) **13**

Introduction 13

Cadrage préliminaire 13

Analyse des facteurs de risque 14

Cartographie des impacts potentiels sur les droits humains, la démocratie et l'État de droit 15

Triage 16

2. PROCESSUS D'ENGAGEMENT DES PARTIES PRENANTES (SEP) **19**

Introduction 19

Explication 19

3. L'EVALUATION DES RISQUES ET DES IMPACTS **21**

Introduction 21

Explications concernant les questions et pistes de réflexion pour l'évaluation des risques et des impacts 21

Résultat de l'évaluation des risques et des impacts 23

4. PLAN D'ATTÉNUATION **25**

Introduction 25

Explications 25

RÉVISION ITÉRATIVE **29**

Introduction 29

Facteurs liés à la production, à la mise en œuvre et au déploiement 29

Facteurs liés à l'environnement réel 29

Mise en œuvre de la révision itérative 30

PARTIE II – MODÈLE DE RESSOURCES POUR L'ANALYSE DES RISQUES FONDÉE SUR LE CONTEXTE (COBRA)[1] **31**

PRÉFACE **33**

Qu'est-ce que l'HUDERIA ? 33

Relation avec la Convention-cadre 33

Orientation générale et spécifique 33

RESSOURCES COBRA **35**

Introduction 35

Aperçu des ressources COBRA A, B et C 35

Ressources COBRA A (Liste des facteurs de risque découlant du contexte d'application du système d'IA) 36

Ressources COBRA B (Liste des facteurs de risque découlant du contexte de conception et de développement du système d'IA) 38

Ressources COBRA C (Liste des facteurs de risque découlant du contexte de déploiement du système d'IA) 43

Ressources COBRA E (Domaines illustratifs de préoccupation potentielle du point de vue des droits humains, de la démocratie et de l'État de droit) 44

Ressources COBRA F (Liste illustrative des secteurs/domaines et domaines possibles de préoccupation potentiels) 57

1. Le Comité sur l'intelligence artificielle (CAI) a examiné la possibilité d'inclure une ressource COBRA D. Toutefois, sans préjudice de toute décision future concernant son inclusion, le modèle HUDERIA de ressources pour l'analyse des risques fondée sur le contexte, adopté le 5 novembre 2025, comprend les ressources COBRA A, B, C, E et F.

Partie I

MÉTHODOLOGIE POUR L'ÉVALUATION DES RISQUES ET DES IMPACTS DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE DU POINT DE VUE DES DROITS HUMAINS, DE LA DÉMOCRATIE ET DE L'ÉTAT DE DROIT

Introduction

Qu'est-ce que l'HUDERIA ?

L'Évaluation des risques et des impacts des systèmes d'intelligence artificielle (IA) du point de vue des droits humains, de la démocratie et de l'État de droit (« l'HUDERIA ») est un guide qui fournit une approche structurée d'évaluation des risques et des impacts des systèmes d'IA spécifiquement adaptée à la protection et à la promotion des droits humains, de la démocratie et de l'État de droit. Elle est destinée à jouer un rôle unique et essentiel à l'intersection des normes internationales en matière de droits humains et des cadres techniques existants sur la gestion des risques dans le contexte de l'IA.

L'HUDERIA peut être utilisée par les acteurs publics et privés pour aider à identifier les risques et les impacts sur les droits humains, la démocratie et l'État de droit tout au long du cycle de vie des systèmes d'IA, et à y apporter une réponse.

L'HUDERIA trouve son origine dans les travaux du Comité ad hoc sur l'intelligence artificielle (CAHAI) (2019-2021) et plus particulièrement de son Groupe d'élaboration des politiques, qui a chargé l'Alan Turing Institute, l'institut national du Royaume-Uni pour la science des données et l'IA, de préparer une proposition originale opérationnalisant les grandes lignes d'un modèle d'évaluation de l'impact sur les droits humains, la démocratie et l'État de droit. La méthodologie HUDERIA a été adoptée par le Comité sur l'intelligence artificielle (CAI) du Conseil de l'Europe le 28 novembre 2024.

Relation avec la Convention-cadre

L'HUDERIA est un outil autonome, non juridiquement contraignant, qui, en tant que tel, n'a pas d'effet juridique. Elle ne revêt pas de caractère obligatoire et n'a pas vocation à être une aide à l'interprétation de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, ci-après dénommée « la Convention-cadre ». De nombreux cadres, politiques, orientations, normes ou outils existants ou futurs peuvent être utilisés en soutien à la gestion des risques et des impacts de l'IA, y compris l'HUDERIA.

Les Parties à la Convention-cadre ont la possibilité d'utiliser ou d'adapter le guide, en tout ou en partie, de développer de nouvelles approches de l'évaluation des risques ou d'utiliser ou adapter les approches existantes conformément à leurs lois applicables, à condition que les Parties respectent pleinement leurs obligations au titre de la Convention-cadre, en particulier le socle en matière de gestion des risques et des impacts définis dans son chapitre V.

Principaux objectifs de l'HUDERIA

L'HUDERIA vise :

- à aider à déterminer dans quelle mesure des activités de gestion des risques pour les droits humains, la démocratie et l'État de droit pourraient être nécessaires, et à offrir une méthodologie pour l'identification, l'évaluation, la prévention et l'atténuation des risques et des impacts applicable à une vaste majorité de technologies et de contextes d'application de l'IA, et qui soit réactive aux innovations et nouveaux cas d'utilisation de l'IA ;
- à promouvoir la compatibilité et l'interopérabilité avec les guides, normes et cadres existants et futurs développés par les organisations ou organes compétents (tels que l'Organisation internationale de normalisation (ISO), la Commission électrotechnique internationale (IECEI), l'Union internationale des télécommunications (UIT), le Comité européen de normalisation (CEN), le Comité européen de normalisation en électronique et en électrotechnique (CENELEC), l'Institut des ingénieurs électriciens et électroniciens, l'Organisation de coopération et de développement économiques (OCDE), l'Institut national des normes et de la technologie (NIST), y compris le cadre de gestion des risques liés à l'IA (*AI Risk Management Framework*) du NIST ainsi que le cadre de gestion des risques et les évaluations d'impact sur les droits fondamentaux au titre de la loi sur l'IA de l'Union européenne.

Quelle est l'approche de l'HUDERIA ?

L'HUDERIA combine les connaissances contemporaines sur les processus et mécanismes de gouvernance technique et sociotechnique pouvant faciliter les activités responsables tout au long du cycle de vie des systèmes d'IA, et les procédures de diligence nécessaires à la protection et la promotion des droits humains, de la démocratie et de l'État de droit.

L'HUDERIA s'appuie sur des variables, des concepts et un langage bien connus pour l'évaluation des risques pour les droits humains (ampleur, portée, probabilité et réversibilité des effets négatifs potentiels sur les droits humains). Elle vise à faciliter leur examen en fournissant des orientations supplémentaires compte tenu de la complexité sociotechnique du cycle de vie de l'IA.

Approche sociotechnique

L'HUDERIA adopte une approche sociotechnique, qui considère que tous les aspects du cycle de vie du système d'IA sont affectés par la relation interconnectée de la technologie, des choix humains et des structures sociales. Dans cette approche, la gestion des risques et des impacts des systèmes d'IA tient compte à la fois des aspects techniques de ces derniers et des contextes juridique, social, politique, économique, culturel et technologique dans lesquels ils opèrent. Une telle approche promeut le développement d'une IA sûre, sécurisée et fiable, à la fois performante et favorisant le respect des droits humains, de la démocratie et de l'État de droit.

Orientation générale et spécifique

L'HUDERIA offre une structure combinant une orientation générale et spécifique et une certaine flexibilité en laissant une marge d'adaptation pour la mise en œuvre pratique.

Au niveau général, la méthodologie HUDERIA décrit des concepts, processus et éléments de haut niveau guidant les activités d'évaluation des risques et des impacts des systèmes d'IA qui sont susceptibles d'avoir des impacts sur les droits humains, la démocratie et l'État de droit.

Au niveau spécifique, le modèle HUDERIA fournit des supports et des ressources (tels que des outils flexibles adaptés aux différents éléments du processus HUDERIA et des recommandations modulables) pouvant contribuer à la mise en œuvre de la méthodologie HUDERIA. Ces ressources sont mentionnées tout au long du texte et constituent une bibliothèque de connaissances pouvant faciliter la prise en compte des risques et des impacts liés aux droits humains, à la démocratie et à l'État de droit, y compris dans d'autres approches de gestion des risques.

Adaptabilité et flexibilité

La méthodologie et le modèle HUDERIA permettent tous les deux une adaptation aux différents contextes, besoins et capacités en fixant des buts, principes et objectifs tout en laissant une marge d'appréciation pour décider de la manière de les atteindre et en offrant un éventail d'options politiques et de gouvernance pouvant être adaptées aux contextes.

Approche graduée et différenciée

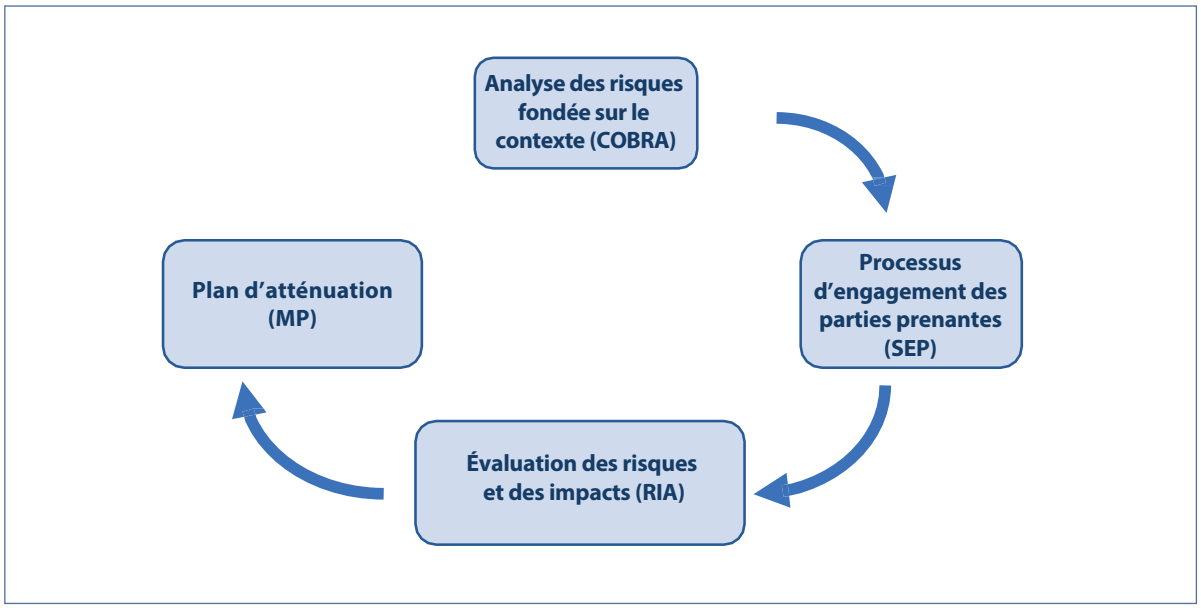
L'HUDERIA vise à établir une approche graduée et différenciée des mesures d'identification, d'évaluation, de prévention et d'atténuation des risques et des impacts, qui tienne compte de la gravité et de la probabilité d'occurrence des impacts négatifs sur les droits humains, la démocratie et l'État de droit, ainsi que des facteurs contextuels pertinents.

Structure de l'HUDERIA

La méthodologie HUDERIA comprend quatre éléments :

1. l'**analyse des risques fondée sur le contexte** (COBRA) fournit une approche structurée de la collecte et de la cartographie des informations nécessaires pour identifier et comprendre les risques qu'un système d'IA pourrait faire peser sur les droits humains, la démocratie et l'État de droit, compte tenu de son contexte sociotechnique. Elle permet également de déterminer si le système d'IA est une solution appropriée au problème envisagé ;
2. le **processus d'engagement des parties prenantes** (SEP) propose une approche pour permettre et rendre opérationnel l'engagement, le cas échéant, avec les parties prenantes pertinentes afin d'obtenir des informations sur les personnes potentiellement affectées et de contextualiser et corroborer les préjudices potentiels et les mesures d'atténuation ;
3. l'é**valuation des risques et des impacts** (RIA) propose des étapes possibles pour l'évaluation des risques et des impacts liés aux droits humains, à la démocratie et à l'État de droit ;
4. le **plan d'atténuation** (MP) propose des étapes possibles pour définir les mesures d'atténuation et de réparation, y compris l'accès aux recours et l'examen itératif.

S'il est logique d'exécuter l'élément COBRA en premier lieu, en fonction des besoins et des approches qu'il a été décidé de suivre, il peut être choisi de modifier la séquence des éléments et/ou d'appliquer ou d'utiliser seulement certaines parties de la méthodologie HUDERIA sur la base des approches existantes en matière de gouvernance de l'IA et des contextes, besoins et exigences spécifiques.



1. Analyse des risques fondée sur le contexte (COBRA)

Introduction

La COBRA aide à l'identification des différents facteurs de risque – caractéristiques ou propriétés d'un système d'IA et de son contexte qui affectent la probabilité d'impacts négatifs sur les droits humains, la démocratie et l'État de droit. Ces facteurs ne doivent pas nécessairement être traités comme des causes d'impacts négatifs, mais plutôt comme des conditions qui sont corrélées à un risque accru de préjudices et qui doivent donc être anticipées et prises en compte dans les efforts de gestion des risques et d'atténuation des impacts. Les facteurs de risque sont classés en trois grandes catégories : le contexte d'application du système, son contexte de conception et de développement, et son contexte de déploiement.

L'examen des facteurs de risque a pour but de faciliter la cartographie des potentiels impacts négatifs sur les droits humains, la démocratie et l'État de droit. Les résultats de cette analyse de la cartographie des facteurs de risque et des impacts visent à déterminer l'étendue de l'approche concernant les éléments ultérieurs de l'HUDERIA, y compris en établissant la proportionnalité des activités HUDERIA ultérieures.

En outre, le résultat de cette analyse de la cartographie des facteurs de risque et des impacts pourrait aussi aider à identifier les contextes sociotechniques spécifiques qui, tout au long du cycle de vie du système, nécessitent une attention particulière en matière de gouvernance.

L'élément COBRA comprend quatre étapes :

1. le cadrage préliminaire ;
2. l'analyse des facteurs de risque ;
3. la cartographie des impacts potentiels sur les droits humains, la démocratie et l'État de droit ;
4. le triage.

Cadrage préliminaire

Objectifs

Le but principal de cette étape est de mener une recherche de fond préliminaire, nécessaire pour guider les activités ultérieures d'identification des facteurs de risque et de cartographie des impacts.

Explications

Le processus COBRA commence par une recherche de cadrage préliminaire qui décrit l'objectif du système, les éléments principaux du système, les contextes dans lesquels il est destiné à être utilisé, le ou les domaines dans lesquels il fonctionnera, le degré d'intervention humaine, ainsi que la nature et la quantité des données qu'il traitera et sur lesquelles il sera formé, en notant toutes les vérifications qui ont déjà été effectuées pour évaluer les biais dans l'ensemble de données ou le modèle. Il identifie les personnes ou les groupes qui pourraient être affectés par le système ou l'affecter, en se concentrant sur les caractéristiques contextuelles pertinentes des personnes et des groupes identifiés, y compris les caractéristiques protégées et les facteurs de vulnérabilité. Il fournit un cadrage préliminaire des potentiels effets négatifs sur les droits humains, la démocratie et l'État de droit en explorant les domaines de préoccupation illustratifs[2] ; et il donne une première cartographie des rôles et des responsabilités tout au long du cycle de vie du système d'IA[3].

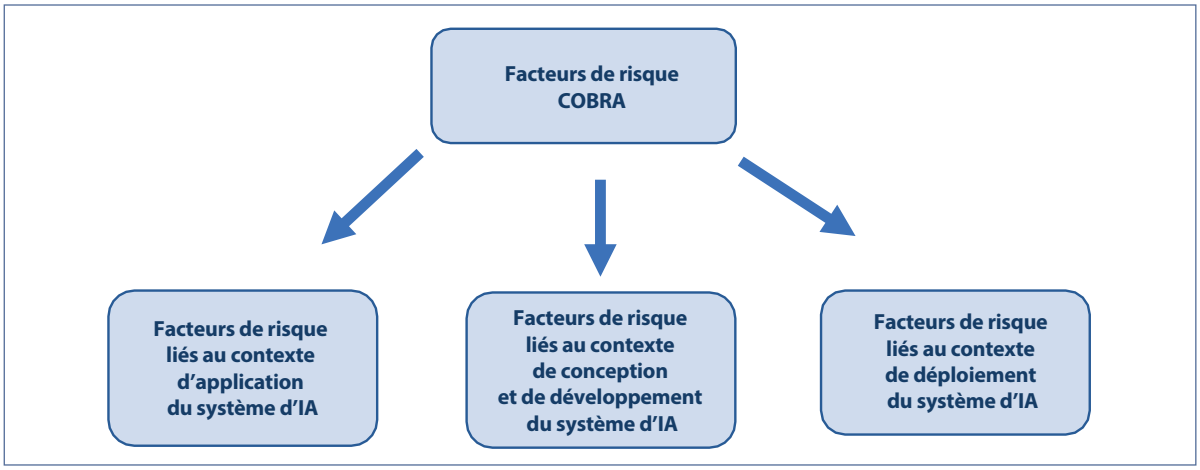
2. Les ressources COBRA E (Domaines illustratifs de préoccupation potentielle du point de vue des droits humains, de la démocratie et de l'État de droit) fournissent un outil qui pourrait être utilisé pour conduire ou informer cette évaluation.
3. La section Rôles et responsabilités du modèle HUDERIA [à élaborer et adopter par le CDNET en 2026] fournira des conseils relatifs à cet aspect de la méthodologie HUDERIA.

Cette activité préliminaire de cadrage peut s'appuyer sur des documents organisationnels (par exemple l'étude d'opportunité, la preuve de concept ou la charte du projet), sur la collaboration et sur des recherches documentaires (si nécessaire). Cette activité préliminaire de cadrage ainsi que les éléments ultérieurs du processus HUDERIA devraient se dérouler, le cas échéant, au sein d'une équipe multidisciplinaire, composée d'experts possédant un éventail de spécialisations complémentaires[4] et de formations techniques et non techniques.

Analyse des facteurs de risque

Objectifs

Le but principal de cette étape est de collecter les informations pertinentes concernant les facteurs de risque liés au contexte prévu d'application du système, à son contexte de conception et de développement, et à son contexte de déploiement. Ces facteurs faciliteront la cartographie des potentiels impacts négatifs sur les droits humains, la démocratie et l'État de droit, et l'évaluation ultérieure des variables de risque principales : gravité (ampleur, portée et réversibilité)[5] et probabilité.



Explications

Les systèmes d'IA sont conçus, développés et utilisés dans une grande variété de contextes et de nombreuses façons différentes, rendant importante l'évaluation holistique de divers facteurs liés au contexte d'application du système, à son contexte de conception et de développement, et à son contexte de déploiement.

Le **contexte d'application du système d'IA**[6] comprend des informations relatives au secteur et au domaine d'application du système, aux environnements juridique et réglementaire dans lequel le système est développé et utilisé, l'objectif visé par le système, et d'autres détails pertinents dans le contexte d'application du système, tels que tout héritage connu de biais de discrimination.

Le **contexte de conception et de développement du système d'IA**[7] comprend les caractéristiques techniques pertinentes du système. Il peut s'agir de limitations connues du système, de considérations liées à la collecte, à l'enrichissement, au stockage, à l'utilisation et au retrait des données, ainsi que de considérations liées à l'algorithme ou au modèle lui-même. Parmi les considérations particulièrement pertinentes, on peut citer les caractéristiques techniques liées à la vie privée et à la protection des données, au biais et à la discrimination, ainsi qu'à l'explicabilité et à l'interprétabilité.

4. L'expertise pertinente pourrait comprendre, le cas échéant, les questions de droits humains, de protection de la vie privée et des données personnelles, la science des données, la gestion des ensembles de données, la sécurité, les risques liés à l'IA, et les tests, l'évaluation, la vérification et la validation de l'IA.
5. Conformément à la Convention-cadre, au Haut-Commissariat des Nations Unies aux droits de l'homme et aux Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme, le terme « gravité » s'entend, aux fins de l'HUDERIA, comme une combinaison des variables que sont l'ampleur, la portée et la réversibilité.
6. Les ressources COBRA A (Liste des facteurs de risque découlant du contexte d'application du système d'IA) fournissent un outil qui pourrait être utilisé pour conduire ou informer cette évaluation.
7. Les ressources COBRA B (Liste des facteurs de risque découlant du contexte de conception et de développement du système d'IA) fournissent un outil qui pourrait être utilisé pour conduire ou informer cette évaluation.

Enfin, le **contexte de déploiement du système d'IA**[8] comprend des facteurs qui régissent la manière dont les risques potentiels peuvent se manifester et être gérés en pratique, comme les mesures qui seront prises pour protéger la vie privée et les données personnelles, atténuer les biais dommageables, assurer une formation adéquate, se prémunir contre les utilisations involontaires, et garantir la responsabilité et la conformité juridique.

Cartographie des impacts potentiels sur les droits humains, la démocratie et l'État de droit

Objectifs

L'étape de cartographie identifie les personnes ou groupes potentiellement affectés et conduit une évaluation initiale des variables de risque principales – gravité (ampleur, portée, réversibilité) et probabilité. La cartographie aide à informer les éléments ultérieurs de la méthodologie HUDERIA, et l'étendue de l'intervention de gouvernance et des mesures d'atténuation qui pourraient être appropriées (voir la section «Triage» ci-dessous). L'analyse des principales variables de risque est essentielle pour obtenir une vue d'ensemble claire et structurée des endroits où les menaces sont les plus susceptibles de se produire et de leur impact potentiel.

Explications

Les **ressources COBRA E**[10] **et F**[11] pourraient être utilisées pour identifier les domaines d'application potentiellement sensibles et les domaines de préoccupation potentiellement pertinents liés aux droits humains, à la démocratie et à l'État de droit[9].

À l'aide des informations collectées lors des étapes précédentes, il pourrait être possible :

a. de déterminer si le système fonctionnera à proximité d'activité(s) (prise de décision ou actions) pouvant produire des impacts sur des personnes affectées dans les secteurs/domaines[13] ;

b. d'identifier et d'énumérer les domaines de préoccupation pertinents[10] et, avec cela à l'esprit, de répondre à la question de savoir si le système pourrait avoir des impacts potentiels ou réels sur des droits humains spécifiques, la démocratie et l'État de droit[14] ;

c. de décrire, pour chaque impact potentiel ou réel identifié, la nature de l'impact potentiel ou réel[11], en tenant compte des impacts différentiels sur les personnes et les groupes affectés en fonction des caractéristiques contextuelles pertinentes, y compris les caractéristiques protégées et les facteurs de vulnérabilité ;

L'analyse de ces points fournira des informations pour l'évaluation initiale des principales variables de risque – la gravité (ampleur, portée, réversibilité) et la probabilité – qui aident à déterminer le risque et à choisir la bonne approche pour les éléments ultérieurs de la méthodologie HUDERIA, ce qui contribuera à garantir que les interventions de gouvernance et les mesures d'atténuation sont alignées sur les besoins tout au long du cycle de vie du système d'IA.

Les résultats de cette analyse peuvent également aider à identifier les possibilités d'utiliser le système d'IA pour soutenir des actions positives qui font progresser les droits humains, y compris la promotion et la garantie de la non-discrimination.

8. Les ressources COBRA C (Liste des facteurs de risque découlant du contexte de déploiement du système d'IA) fournissent un outil qui pourrait être utilisé pour conduire ou informer cette évaluation.
9. Dans les ressources COBRA, les références aux droits humains tels qu'ils sont énoncés dans les divers instruments internationaux relatifs aux droits humains sont incluses à titre d'illustration. Ces références ne s'appliquent qu'aux États qui sont parties à ces instruments. Chaque État est censé appliquer ses propres lois et obligations internationales.
10. Les ressources COBRA E et F pourraient être utilisées pour identifier des domaines de préoccupation relatifs aux droits humains, à la démocratie et à l'État de droit ainsi que pour identifier des secteurs/domaines potentiellement sensibles.
11. Pour la clarté de l'évaluation, il convient de prendre en compte à la fois les impacts négatifs ou restrictifs et les impacts bénéfiques, améliorants ou autrement positifs produits par les systèmes d'IA, étant donné que divers problèmes de partialité et de discrimination peuvent se poser en ce qui concerne les systèmes qui produisent les deux types d'impacts.

Détermination du niveau de risque

Les variables suivantes peuvent être employées pour indexer le niveau de risque de chacun des impacts négatifs potentiels sur les droits humains, la démocratie et l'État de droit qui ont été identifiés à la suite de l'exercice de cartographie :

1. l'**ampleur**[12] des impacts négatifs potentiels (c'est-à-dire l'importance du préjudice potentiel);
2. la **portée** des effets négatifs potentiels (y compris le nombre de personnes affectées, les caractéristiques protégées ou la vulnérabilité des individus ou des groupes et la durée des impacts) ;
3. la **réversibilité**[13] des impacts négatifs potentiels (c'est-à-dire l'information sur la possibilité de remédier aux effects négatifs ou de rétablir les personnes affectées dans leur situation antérieure à l'impact ou dans une situation équivalente) ;
4. la **probabilité** des impacts négatifs potentiels.

Les équipes concernées devraient passer en revue chacun des impacts potentiels qui ont été identifiés et considérer, pour chaque domaine de préoccupation lié aux droits humains, à la démocratie et à l'État de droit, et chaque groupe affecté, l'ampleur, la portée, la réversibilité et la probabilité des effets négatifs potentiels ou réels. La législation ou la politique nationale peut fournir des définitions plus détaillées qui peuvent être utilisées pour éclairer cette détermination des risques et pour déterminer des approches appropriées et proportionnées pour les activités ultérieures d'HUDERIA (par exemple l'engagement des parties prenantes).

Il pourrait être envisagé d'établir une méthode pour combiner ces variables afin de permettre l'étalonnage du risque et la détermination d'approches appropriées et proportionnées pour les activités ultérieures de HUDERIA (par exemple l'engagement des parties prenantes), ainsi que celle de l'étendue et de la profondeur des interventions de gouvernance et des mesures de gestion et d'atténuation des risques. Cela peut impliquer la formulation de méthodes quantitatives ou semi-quantitatives de calcul des risques, de matrices de risques ou de procédures plus qualitatives ou fondées sur des règles. Tout mécanisme d'étalonnage des risques résultant de la combinaison de ces variables aux fins de l'évaluation d'HUDERIA peut prendre en compte les éléments suivants :

- en ce qui concerne les droits humains, les effets de faible portée et à fort impact ainsi que les effets de grande portée et à faible impact sur chaque personne affectée;
- en ce qui concerne la démocratie et l'État de droit, le mécanisme pourrait prendre en compte, en particulier, les effets durables et de grande portée sur les personnes, les institutions et la société en général.

Triage

Objectifs

L'objectif principal de cette étape est de s'appuyer sur les informations recueillies dans le cadre des activités COBRA précédentes, et ainsi :

- de faciliter l'identification et le triage des systèmes qui présentent un risque significatif, de sorte que la méthodologie HUDERIA ne soit pas onéreuse pour les systèmes d'IA à risque minime ou faible ;
- de procéder à une évaluation initiale visant à déterminer si le système d'IA devrait être développé ou déployé, en se fondant sur la question de savoir si les avantages du développement ou du déploiement du système d'IA l'emportent sur les risques qu'il comporte, compte tenu notamment de ses impacts potentiels sur les droits humains, la démocratie et l'État de droit, et si l'utilisation du système d'IA est incompatible avec le respect des droits humains, de la démocratie et de l'État de droit.

Approche adaptable du triage

Les activités préalables de cette étape permettent de dresser un premier tableau du profil de risque du système d'IA.

Les informations recueillies peuvent, par exemple, être suffisantes pour déterminer qu'il est peu probable que le système ait un quelque impact sur les droits humains, la démocratie ou l'État de droit, rendant ainsi superflus les éléments suivants d'HUDERIA. Une conclusion similaire pourrait être tirée si les impacts identifiés

12. Le terme « ampleur » est parfois appelé « sévérité » dans le contexte de l'évaluation des risques.
13. Le terme « réversibilité » est parfois appelé « rémédiabilité » dans le contexte de l'évaluation des risques.

sont insignifiants ou peu probables. Si les impacts identifiés conduisent à la décision de ne pas développer ou déployer un système d'IA parce qu'il est considéré comme incompatible avec le respect des droits humains, de la démocratie et de l'État de droit, les éléments suivants de l'HUDERIA sont également inutiles. Enfin, dans les cas où des impacts potentiels graves sont identifiés, un éventail de stratégies de gestion des risques et de réponses (y compris le processus d'engagement des parties prenantes et diverses autres activités de gouvernance du projet) peuvent être justifiées. Pour répondre à cette complexité, l'HUDERIA ne prescrit pas d'orientations détaillées pour ajuster les efforts de gestion des risques, mais présente simplement des éléments proposés qui pourraient être appliqués sur la base du risque d'impacts négatifs potentiels sur les droits humains, la démocratie et l'État de droit. Différentes approches de la détermination des étapes de gestion des risques fondées sur les impacts potentiels et réels identifiés – ou une combinaison d'entre elles – peuvent être appliquées en fonction du cadre ou de l'environnement réglementaire national spécifique, du secteur, du système et du contexte (par exemple des approches fondées sur des seuils, des scénarios, la proportionnalité, la dynamique, ou des approches spécifiques au contexte).

La décision finale d'utiliser une méthode qualitative, quantitative, mixte ou autre est laissée à la discrétion des autorités ou, le cas échéant, des équipes de projet d'IA responsables du système.

« Questions zéro »

Pour déterminer si les avantages de la construction ou du déploiement du système d'IA, y compris les avantages sociaux supplémentaires pouvant résulter de l'utilisation du système au-delà de son objectif premier, l'emportent sur les risques compte tenu des facteurs de risque et des impacts potentiels identifiés, il convient de considérer les points suivants :

- se demander si l'utilisation du système est appropriée compte tenu de la nature du problème que le projet d'IA tente de résoudre ;
- la mesure dans laquelle les technologies et les processus existants déjà en place pour résoudre le problème considéré sont mieux placés pour le faire, compte tenu du profil de risque et des impacts négatifs potentiels du système envisagé, en mettant l'accent, le cas échéant, sur tout risque marginal ajouté par l'introduction de l'IA dans le contexte actuel ;
- la mesure dans laquelle le système envisagé sera en mesure de répondre aux besoins et aux attentes des opérateurs ;
- la mesure dans laquelle les effets du système prospectif seront équitables pour tous les groupes affectés ;
- la mesure dans laquelle la qualité et la représentativité des données actuellement ou potentiellement disponibles sont suffisantes pour que le système prospectif soit efficace et sûr, et qu'il évite raisonnablement les biais potentiellement dommageables ;
- la mesure dans laquelle des ressources (humaines et matérielles) suffisantes sont disponibles et capables de répondre aux exigences techniques et de mener à bien des actions techniques et de gouvernance afin d'atténuer de manière adéquate les risques identifiés ; et
- les contextes d'utilisation potentiels du système et les risques de mauvaise utilisation ou d'abus, y compris à travers son déploiement au-delà de l'objectif prévu.

2. Processus d'engagement des parties prenantes (SEP)

Introduction

La possibilité de conduire cette étape peut être considérée afin d'améliorer la qualité des informations pour l'élément suivant de l'HUDERIA – l'évaluation des risques et des impacts – en intégrant les points de vue des personnes potentiellement affectées identifiées, y compris celles en situation de vulnérabilité.

L'engagement des parties prenantes, tel qu'il est défini dans la méthodologie HUDERIA, peut prendre diverses formes. Le niveau de participation des personnes affectées devrait être renseigné par les facteurs de risque et les impacts potentiels et réels identifiés lors de l'étape COBRA. L'implication des parties prenantes tout au long du cycle de vie du système d'IA peut également offrir une variété d'avantages supplémentaires tels que la promotion de la transparence, l'instauration de la confiance et l'amélioration de la facilité d'utilisation et des performances du système d'IA.

Explication

Le SEP comporte cinq étapes clés[14] : analyse des parties prenantes, réflexion sur la positionnalité, définition des objectifs de l'engagement, détermination de la méthode d'engagement, lancement et mise en œuvre.

Analyse des parties prenantes

L'analyse des parties prenantes identifie les groupes de parties prenantes susceptibles d'être affectés par les activités menées dans le cadre du cycle de vie de système, ou de les affecter. Cette analyse[15] doit permettre d'évaluer les intérêts relatifs, les droits, les vulnérabilités et avantages potentiels et existants des parties prenantes identifiées, ainsi que l'importance des groupes de parties prenantes identifiés. À cette étape, il faut envisager d'inclure de façon constructive les points de vue de ceux qui :

1. sont exposés de manière disproportionnée aux risques liés à l'utilisation du système ;
2. sont particulièrement vulnérables aux éventuels préjudices ; ou
3. ont une capacité particulièrement limitée à influer sur la manière dont le système est conçu (par exemple les groupes actuellement ou historiquement marginalisés, défavorisés ou sous-représentés, ou les personnes en situation de vulnérabilité ou ayant des besoins spécifiques).

Réflexion sur la positionnalité

L'étape suivante du SEP consiste à réfléchir au point de vue positionnel vis-à-vis des parties prenantes affectées, le but étant de reconnaître les limites des perspectives des utilisateurs d'HUDERIA et d'identifier les points de vue manquants qui permettraient d'améliorer l'évaluation des impacts potentiels et réels du système.

En fonction des facteurs de risque et des impacts potentiels identifiés lors de la phase COBRA, cela pourrait comprendre une évaluation autodéterminée des utilisateurs d'HUDERIA de leurs caractéristiques démographiques, de leur éducation et leur formation, de leur parcours socioéconomique, mais aussi du contexte à l'échelle de l'équipe et de l'organisation.

14. Le processus décrit dans cette section est de nature illustrative, la décision finale concernant le processus de participation des parties prenantes étant laissée à la discrétion des autorités ou, le cas échéant, des équipes de projet d'IA responsables du système.
15. Les ressources SEP (Liste de questions pour l'évaluation de l'importance des parties prenantes) [à élaborer et adopter par le CDNET en 2026] fournissent des questions et outils détaillés pour guider l'identification des parties prenantes pertinentes.

Les principales questions auxquelles les utilisateurs d'HUDERIA doivent réfléchir lorsqu'ils entreprennent cette étape de la méthodologie sont les suivantes :

- Dans quelle mesure mes caractéristiques personnelles, mes identifications de groupe, mon statut socio-économique, mes antécédents en matière d'éducation, de formation et de travail, la composition de mon équipe et mon cadre institutionnel représentent-ils des sources de pouvoir et d'avantages ou des sources de marginalisation et de désavantages ?
- Comment cette position influence-t-elle ma capacité et celle de mon équipe à identifier et à comprendre les parties prenantes affectées et les impacts potentiels du système d'IA ?

En fonction des facteurs de risque identifiés au cours du processus COBRA, les utilisateurs d'HUDERIA devraient également envisager de faire appel à des parties prenantes externes ou à des consultants disposant d'une expertise spécifique, par exemple en matière de droit des droits humains, en ce qui concerne les impacts potentiels et réels du système sur les droits humains.

Définition des objectifs de l'engagement

La définition d'objectifs clairs pour l'engagement des parties prenantes vise à créer une compréhension aisée de *comment* et *pourquoi* des activités d'engagement sont menées. Ces objectifs doivent permettre l'engagement inclusif, éclairé et constructif des personnes potentiellement affectées[16].

Détermination de la méthode d'engagement

Pour déterminer la ou les méthodes appropriées[17] à employer pour l'engagement des parties prenantes, il est nécessaire d'évaluer les besoins des personnes potentiellement affectées en prenant en considération, le cas échéant, les résultats du processus COBRA et d'autres facteurs pertinents tels que les contraintes en matière de ressources, les difficultés à aller au-devant des groupes isolés ou socialement exclus, les contraintes de moyens comme les difficultés dues à la fracture numérique ou au déficit d'information, les délais, etc.

Les critères suivants peuvent servir d'orientation pour l'élément SEP :

1. l'**engagement :** la participation significative des personnes affectées ou potentiellement affectées est intégrée dans les éléments pertinents du processus ;
2. l'**égalité et l'interdiction de la discrimination :** les processus d'engagement et de consultation sont inclusifs, sensibles au genre et tiennent compte des besoins des personnes
3. et des groupes présentant des caractéristiques protégées ou susceptibles d'être vulnérables ou marginalisés ;
4. l'**autonomisation :** la prise en compte de l'âge et des besoins en termes d'accessibilité ainsi que le renforcement des capacités des personnes et des groupes présentant des caractéristiques protégées ou risquant d'être vulnérables ou marginalisés sont entrepris afin de garantir la participation significative de ces personnes et de ces groupes ;
5. la **transparence :** assurer le partage d'informations pertinentes et intelligibles entre les parties prenantes à intervalles réguliers, mettre à la disposition des parties prenantes participantes des informations sur le système d'IA qui permettent une compréhension globale des implications et impacts potentiels sur les droits humains, le cas échéant, et communiquer publiquement les conclusions d'HUDERIA et les plans de gestion des impacts (plans d'action) ; et
6. la **responsabilité :** la responsabilité de la mise en œuvre, du contrôle et du suivi des mesures d'atténuation est attribuée à des entités, des personnes ou des fonctions particulières au sein de l'organisation.

Mise en œuvre

Une fois les quatre activités précédentes menées à terme, les processus d'engagement proportionné peuvent être menés. Ils devraient être en phase avec les résultats de l'analyse des parties prenantes, la réflexion sur la positionnalité et les objectifs et méthodes d'engagement, et être documentés de manière appropriée.

16. Les ressources SEP [à élaborer et adopter par le CDNET en 2026] fournissent des questions détaillées indicatives et des options d'engagement des parties prenantes.
17. Les ressources SEP (Exemples de méthodes d'engagement et de questions pertinentes) [à élaborer et adopter par le CAI en 2026] fournissent de possibles exemples de méthodes d'engagement et une liste de questions pertinentes qui pourront aider à la détermination des groupes de parties prenantes appropriés.

3. Évaluation des risques et des impacts

Introduction

L'objectif de l'évaluation des risques et des impacts est de fournir des évaluations détaillées des impacts potentiels et réels que les activités menées au cours du cycle de vie d'un système d'IA pourraient avoir sur les droits humains, la démocratie et l'État de droit.

Conformément au triage effectué lors de l'étape COBRA, réaliser l'évaluation des risques et des impacts est particulièrement important pour les systèmes d'IA qui pourraient faire peser un risque significatif sur les droits humains, la démocratie et l'État de droit. Après le triage de l'analyse COBRA, cette étape des processus peut n'être nécessaire que pour certains systèmes d'IA, en particulier ceux qui sont considérés comme présentant des risques importants pour les droits humains, la démocratie et l'État de droit.

L'évaluation des risques et des impacts vise à :

- réexaminer, contextualiser et corroborer les préjudices potentiels et réels identifiés lors de la COBRA ;
- identifier et analyser d'autres préjudices potentiels et réels en se livrant à une réflexion approfondie afin d'identifier des lacunes dans la complétude et l'exhaustivité des préjudices précédemment énumérés ;
- évaluer les variables de risque en matière d'ampleur, de portée, de réversibilité et de probabilité des impacts négatifs potentiels, afin de mieux estimer les risques pour qu'ils soient ensuite hiérarchisés, gérés et atténués.

L'évaluation des risques et des impacts s'inscrit dans le prolongement de l'identification des facteurs de risques contextuels pour les droits humains, la démocratie et l'État de droit et de la cartographie des impacts potentiels pour les droits humains, la démocratie et l'État de droit, menées dans la COBRA, et des potentiels éclairages apportés par le SEP, afin de traiter les impacts potentiels et réels du système d'IA.

Cette évaluation est conduite de manière constructive à travers un processus en deux étapes qui permettent l'élaboration d'un plan d'atténuation des impacts et la mise en place d'accès à des voies de recours lors de l'étape suivante de la méthodologie HUDERIA.

Explications concernant les questions et pistes de réflexion pour l'évaluation des risques et des impacts

Introduction

L'évaluation des risques et des impacts dans le contexte de l'HUDERIA est organisée en deux étapes.

Lors de la première étape, l'accent est mis sur l'identification des impacts potentiels et réels et, plus spécifiquement, sur *comment* les impacts potentiels et réels identifiés lors des étapes COBRA et SEP pourraient se produire, ce qui permet une approche plus ouverte et exploratoire et facilite une analyse plus approfondie des contextes spécifiques, de la portée, de l'ampleur et de la réversibilité des impacts, en particulier en ce qui concerne les individus en situation de vulnérabilité ou les groupes vulnérables.

Lors de la seconde étape, l'évaluation des variables de risque que sont l'ampleur, la portée, la réversibilité et la probabilité des impacts potentiels ou réels identifiés est conduite. Un examen approfondi de ces variables en fonction du contexte permet de hiérarchiser les mesures d'atténuation en différenciant la gravité des impacts du système d'IA.

Ampleur

L'ampleur d'un impact négatif potentiel et réel se réfère à la sévérité de la conséquence attendue du préjudice potentiel.

La considération de la gravité de tout préjudice potentiel devrait inclure une réflexion quant aux différentes manières dont les personnes ou les groupes (en particulier ceux qui possèdent des caractéristiques qui pourraient les rendre plus vulnérables à l'impact négatif) pourraient subir ce préjudice, ainsi qu'aux différentes mesures dans lesquelles ils pourraient le subir.

Les délibérations concernant l'ampleur prennent en considération les questions supplémentaires suivantes :

1. Pour chaque impact négatif potentiel et réel identifié, existe-t-il des personnes ou des groupes qui possèdent des caractéristiques susceptibles de les rendre plus vulnérables à l'impact ? Dans l'affirmative, quelles sont ces caractéristiques et ceux qui les possèdent pourraient-ils subir le préjudice de manière plus aiguë ou plus grave que les autres ?
2. Pour chaque impact négatif potentiel et réel identifié, quels sont les personnes ou les groupes qui pourraient subir l'impact le plus grave préjudice considéré ?

Les réponses à ces questions joueront par la suite un rôle important lors de la phase de planification des mesures d'atténuation, lorsque la réparation et la hiérarchisation des potentiels préjudices seront envisagées.

Portée

La portée d'un effet négatif potentiel et réel concerne l'estimation du nombre de personnes affectées et de la durée des effets.

Les estimations de la portée des impacts négatifs potentiels et réels identifiés sont examinées une à une, en accordant une attention particulière aux niveaux d'exposition au préjudice de certains groupes de personnes affectés ainsi qu'aux impacts cumulés ou agrégés du système sur les personnes et groupes de personnes affectés actuels et futurs.

Les délibérations sur la portée peuvent comprendre la considération des questions suivantes :

1. Pour chaque impact négatif potentiel et réel identifié, existe-t-il des groupes présentant des caractéristiques susceptibles de les rendre plus vulnérables à des niveaux plus élevés d'exposition[18][19] à l'impact en question ? Dans l'affirmative, à quel niveau d'exposition ces groupes pourraient-ils être confrontés ?
2. Pour chaque impact négatif potentiel et réel identifié, réfléchissez à la durée globale des impacts du système d'IA sur le droit ou le domaine de préoccupation (dans le cas de la démocratie ou de l'État de droit) considéré. Le système a-t-il des impacts cumulés ou agrégés sur les personnes affectées et les générations futures de personnes affectées qui auraient pour conséquence d'étendre les effets de l'impact au-delà de la portée déjà identifiée ?
3. Quelques questions générales à considérer lors de l'évaluation des impacts cumulés ou agrégés pourraient inclure :
   - Y a-t-il un risque que la mise à disposition et l'utilisation du système contribuent à produire, plus largement, des effets négatifs sur les droits humains, la démocratie ou l'Etat de droit lorsque le déploiement du système est coordonné avec (ou se produit en même temps que) d'autres systèmes qui remplissent des fonctions ou des objectifs analogues ?
   - Y a-t-il un risque que la mise à disposition et l'utilisation du système reproduisent, renforcent ou accroissent des préjudices hérités, enracinés dans l'histoire sociale, ou des caractéristiques inhérentes qui pourraient avoir des effets induits pour les individus et les groupes concernés ?

18. Les ressources SEP A et B [à élaborer et adopter par le CDNET en 2026] fournissent des questions détaillées (Liste de questions pour l'évaluation de l'importance des parties prenantes) et la description des formats d'engagement des parties prenantes (Liste des facteurs déterminant les objectifs et les niveaux d'engagement des parties prenantes) pouvant aider les équipes de projet à déterminer les groupes de parties prenantes particulièrement pertinents et les objectifs d'engagement.
19. L'expression « niveau d'exposition » est ici comprise comme la proportion d'un groupe qui est affectée négativement par un système d'IA. Lorsque seule une petite fraction du groupe est impactée, les membres ont un faible niveau d'exposition, et lorsqu'une très grande fraction du groupe est impactée, les membres ont un niveau d'exposition élevé. Par exemple, les membres d'un groupe caractérisé par un faible statut socio-économique peuvent avoir un niveau d'exposition élevé aux impacts négatifs potentiels d'un modèle d'IA utilisé pour l'attribution de prestations publiques.

- Peut-on considérer que la mise à disposition et l'utilisation du système contribuent, plus largement, à des effets négatifs agrégés (par exemple sur l'environnement et la santé publique) si l'on met en parallèle le déploiement du système et l'existence d'autres systèmes susceptibles d'avoir des impacts comparables ?

Réversibilité

Comme expliqué précédemment, la réversibilité se réfère à l'information sur le degré de réparabilité ou de restauration que les personnes potentiellement affectées concernées peuvent attendre comme suite aux efforts déployés pour surmonter l'impact négatif considéré et pour replacer les personnes affectées dans une situation au moins identique ou équivalente à celle dans laquelle elles se trouvaient avant l'impact considéré. À l'instar des considérations sur l'ampleur des impacts potentiels, la compréhension de la réversibilité d'un préjudice dépend de la connaissance que l'on a du contexte spécifique de ce préjudice et des personnes affectées qui y sont soumises. Pour déterminer le degré de réversibilité d'un impact négatif potentiel, il convient de réfléchir aux efforts à fournir pour surmonter le préjudice et (éventuellement) y mettre fin.

Les membres de groupes différents peuvent nécessiter des niveaux d'effort variés pour surmonter des impacts négatifs, en fonction de leur âge, de leur position dans la société et des circonstances du préjudice (les groupes vulnérables et marginalisés disposant souvent de moins de résilience que les groupes dominants, privilégiés ou majoritaires).

Probabilité

L'évaluation de la probabilité d'un risque consiste à estimer la vraisemblance qu'un risque donné se produise, en se fondant, le cas échéant, sur un jugement qualitatif, une analyse quantitative et une compréhension du contexte.

La détermination du niveau de probabilité d'un risque implique une large analyse des conditions contextuelles et opérationnelles et est généralement déterminée par le niveau (type, quantité et qualité) des informations indiquant que le risque est susceptible de se matérialiser. Cela garantit que les évaluations des risques sont fondées à la fois sur des données et sur des avis d'experts, ce qui facilite la hiérarchisation et l'atténuation des risques potentiels.

Résultat de l'évaluation des risques et des impacts

Après avoir répondu aux questions et réfléchi aux pistes de réflexion sur l'identification et l'évaluation des impacts négatifs potentiels et réels, la prévention des impacts ainsi que la hiérarchisation et la planification des mesures d'atténuation peuvent être amorcées. Ce processus de planification des mesures d'atténuation et de mise en place d'accès à des voies de recours est traité dans la section qui suit.

4. Plan d'atténuation

Introduction

Une fois que les impacts négatifs potentiels et réels ont été identifiés et évalués, un plan d'atténuation devrait être élaboré et une réflexion concernant la mise en place d'accès aux voies de recours pour les personnes potentiellement affectées devrait avoir lieu, le cas échéant.

Cette partie du processus HUDERIA spécifie les actions et les processus visant à traiter les impacts négatifs potentiels et réels en :

- formulant des mesures d'atténuation ;
- élaborant un plan d'atténuation en fonction de la gravité et de la probabilité des préjudices identifiés ;
- mettant en place, le cas échéant, l'accès à des voies de recours pour les personnes potentiellement affectées et les autres parties concernées.

Explications

Cadrage et priorisation

La planification diligente de la prévention et de l'atténuation des risques et des impacts débute par une étape de cadrage et de priorisation. Avec la contribution des personnes affectées associées, le cas échéant, il faudrait passer en revue chaque impact négatif potentiel et réel identifié et cartographier les relations mutuelles et les interdépendances entre ces impacts ainsi que les facteurs de risque sociaux environnants identifiés lors de l'étape COBRA (par exemple les vulnérabilités et la précarité spécifiques au contexte).

Lorsqu'il est nécessaire de prioriser les mesures de prévention et d'atténuation (par exemple lorsque des retards dans la prise en charge d'un préjudice potentiel ou la vulnérabilité spécifique d'un individu ou groupe affecté risquent de réduire sa réversibilité), les décisions devraient être dictées par la prise en compte de la probabilité et de la gravité relative des impacts examinés.

Obligations juridiques

Une considération importante dans l'élaboration d'un plan d'atténuation est que des obligations juridiques concernant le respect des droits humains, de la démocratie et de l'État de droit, telles que définies dans le droit international et national applicable, devraient être prises en compte à cette étape du processus HUDERIA en considérant si et, le cas échéant, comment les impacts négatifs potentiels peuvent être atténués et les impacts négatifs réels peuvent être traités.

La disponibilité et l'efficacité des recours juridiques, y compris la restauration ou l'indemnisation en tant que recours juridiques, sont déterminées par le droit international et national applicable.

Hiérarchie des mesures d'atténuation

Pour déterminer l'éventail des mesures qui peuvent être prises pour prévenir ou atténuer les impacts négatifs potentiels, une approche structurée appelée « hiérarchie des mesures d'atténuation » (éviter, atténuer, restaurer, compenser) peut être utilisée.

Au cours des phases initiales de conception, au début du cycle de vie du système d'IA, les impacts à examiner n'ont pas encore eu lieu. À ce stade, les options d'atténuation « éviter » et « atténuer » sont donc plus pertinentes. En revanche, lors des itérations ultérieures du suivi, du réexamen et de la réévaluation (c'est-à-dire en phase de déploiement), des impacts négatifs pourraient s'être produits. Les options « restaurer » et « compenser » auront donc toute leur place, aux côtés des options « éviter » et « atténuer ».

Les différentes options de la hiérarchie des mesures d'atténuation sont les suivantes :

ÉVITER	ATTÉNUE	RESTAURER	COMPENSER
Procéder à des changements dans les processus de conception, de développement et de déploiement qui précèdent la production et l'utilisation du système d'IA, ou dans le système lui-même, dès le début, pour éviter les impacts négatifs. Attention : éviter ne signifie pas ignorer.	Mettre en œuvre des mesures au niveau des processus de conception, de développement ou de déploiement qui précèdent la production et l'utilisation du système d'IA, ou procéder à des changements dans le système lui-même, afin de réduire l'impact négatif le plus possible.	Procéder à des changements pour replacer les personnes affectées dans une situation au moins identique ou équivalente à celle dans laquelle elles se trouvaient avant le préjudice.	Compensation en nature ou par d'autres moyens, lorsque faisable et que d'autres mesures d'atténuation ne sont pas envisageables ou pas efficaces.
Niveau 1	Niveau 2	Niveau 3	Niveau 4
Mesure la plus recommandée	…		Mesure la moins recommandée

L'utilisation de l'expression « hiérarchie des mesures d'atténuation » suggère que la priorité soit donnée, dans un premier temps, à l'évitement total des impacts négatifs potentiels et réels, puis à leur réduction et à leur réparation. Il convient également de noter qu'aux derniers stades du cycle de vie du système d'IA, lorsque les options de restauration et de compensation sont plus pertinentes, plusieurs options d'atténuation peuvent être appropriées (c'est le cas par exemple d'une personne affectée qui doit être restaurée dans ses droits en même temps que des mesures immédiates sont aussi prises pour réduire le plus possible les préjudices ultérieurs).

En tout état de cause, les décisions concernant la ou les mesures de prévention et/ou d'atténuation à prendre devraient être guidées par des considérations priorisant la protection des droits humains, de la démocratie et de l'État de droit, et les choix visant à éviter et à atténuer les impacts négatifs devraient être favorisés par rapport aux choix visant à compenser ou à indemniser les personnes potentiellement affectées pour les préjudices qu'elles ont subis.

Compte tenu de l'ensemble des informations obtenues à ce stade du processus HUDERIA, il existe une possibilité de réexaminer les questions zéro. Ces informations pourraient également être utiles pour éclairer les discussions visant à savoir si les activités du cycle de vie du système d'IA en question (en cours de développement ou déjà utilisé en cas de révision itérative) sont conformes aux droits humains, à la démocratie et à l'État de droit.

Accès à des voies de recours

Les mesures visant à traiter des impacts négatifs ne se limitent pas aux recours juridiques. Ces impacts peuvent être traités à l'aide d'autres mesures d'atténuation telles que celles définies dans les politiques, les orientations ou d'autres instruments.

Lors de la mise en place de telles mesures, les sujets suivants devraient être abordés :

a. il convient de déterminer si des mesures et mécanismes de responsabilité existants sont en place en ce qui concerne les droits humains, la démocratie et l'État de droit. Il est essentiel que ces cadres existants soient appliqués au contexte des activités du cycle de vie des systèmes d'IA ;

b. la complexité technique, la nature axée sur les données et l'opacité relative de certains systèmes d'IA peuvent limiter leur transparence. Il peut en résulter une importante asymétrie dans l'accès, la compréhension ou le contrôle de l'information entre les différentes parties impliquées dans le cycle de vie du système d'IA. Des mesures visant à fournir des informations sur le système d'IA et sur ses impacts sur les personnes affectées peuvent faciliter la mise en place et l'accessibilité de voies de recours effectives en cas d'impacts négatifs sur ces personnes ;

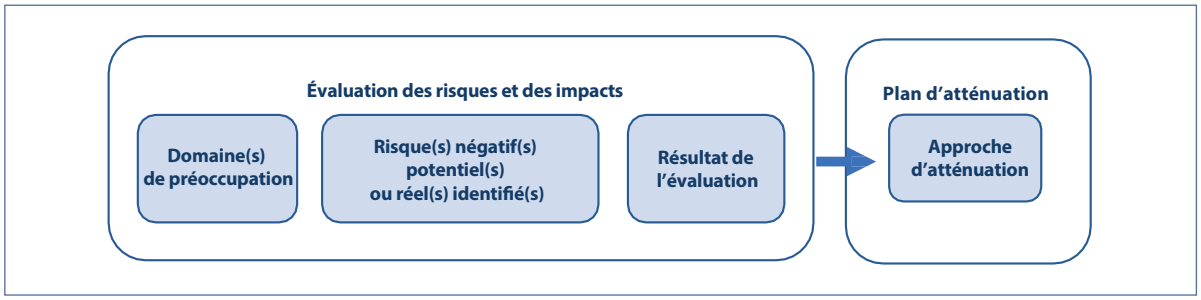
c. les informations fournies dans le cadre de ces mesures doivent être adaptées au contexte, claires et utiles, afin que les personnes puissent les utiliser efficacement pour exercer leurs droits dans les procédures liées aux décisions qui les concernent ;

d. le cas échéant, il peut être nécessaire de fournir aux personnes affectées d'autres garanties et sauvegardes procédurales effectives, conformément au droit international et national applicable.

Résultat de cet élément du processus HUDERIA

Cet élément devrait permettre de décrire de manière claire les mesures et les actions visant à traiter les risques et les impacts identifiés, ainsi que les rôles et les responsabilités des différents acteurs impliqués dans l'atténuation, la gestion et le suivi. Le cas échéant, cet élément devrait également fournir un aperçu accessible des mécanismes et mesures de réparation à la disposition des personnes impactées.

En outre (voir la section sur la révision itérative ci-dessous), un plan est établi pour le suivi des efforts d'atténuation et pour la réévaluation itérative de ces efforts au cours des phases ultérieures du cycle de vie du système d'IA.



Révision itérative

Introduction

La réalisation de l'HUDERIA au début du cycle de vie d'un système d'IA n'est qu'une première étape, bien qu'essentielle, d'un processus d'évaluation et de réévaluation responsables plus long. Le processus de révision itérative garantit que l'évaluation des risques et des impacts demeure effective tout au long du cycle de vie du système d'IA. Il s'agit d'un processus continu qui offre, à intervalles réguliers, l'occasion d'identifier de nouveaux impacts et de mettre à jour le plan d'atténuation.

Il est probable que les impacts du système d'IA évolueront au fil du temps, soit en raison de décisions prises en phase de production et de mise en œuvre du système d'IA, ou de ses applications contextuelles, soit en raison de changements externes dans l'environnement réel. Ces changements, qui peuvent concerner le cycle de vie des données, le développement et la conception du système d'IA, les processus de passation de marché, l'évolution des techniques d'IA, l'intégration ou l'opérationnalisation du système, les failles de sécurité, ainsi que les événements ou occurrences significatifs entraînant des conséquences préjudiciables ou involontaires, peuvent influencer les performances du système d'IA et/ou son impact sur les personnes et les groupes affectés.

De telles modifications appellent une révision destinée à s'assurer que les droits humains, la démocratie et l'État de droit sont respectés en permanence, tout au long du cycle de vie du projet d'IA. Une attention particulière devrait être accordée à la manière dont ces changements affectent la performance du système et son impact sur les personnes et les communautés.

Facteurs liés à la production, à la mise en œuvre et au déploiement

Les choix faits à tout moment pendant le cycle de vie du système ainsi que les événements se déroulant lors du déploiement du système peuvent nécessiter la révision des décisions et évaluations antérieures, en particulier celles résultant du processus HUDERIA, d'où la nécessité d'une réévaluation, d'un réexamen et d'une révision.

Ces changements, en particulier ceux concernant le cycle de vie des données, le développement et la conception du système d'IA, les changements dans les techniques d'IA, l'intégration ou l'opérationnalisation du système, les vulnérabilités en matière de sécurité, ainsi que des événements ou occurrences significatifs comme ceux ayant des conséquences préjudiciables ou inattendues, peuvent influencer la performance du système d'IA et/ou son impact sur les personnes et les groupes affectés. Les processus d'un modèle d'IA sont itératifs et fréquemment non linéaires, nécessitant des révisions et mises à jour fréquentes, le cas échéant.

Facteurs liés à l'environnement réel

Les changements qui surviennent dans les contextes sociaux, réglementaires, politiques ou juridiques durant lesquels le système est en cours de production ou d'utilisation peuvent avoir des effets sur la qualité de fonctionnement du système d'IA et sur la manière dont il influe sur les droits des personnes et des groupes affectés.

De même, les réformes réglementaires et politiques, ainsi que les changements dans les méthodes d'enregistrement des données, peuvent survenir au sein de la population concernée d'une manière qui fait se demander si les données utilisées pour entraîner le modèle décrivent de manière correcte les phénomènes, les populations ou les facteurs connexes.

Dans le même ordre d'idées, des changements culturels ou comportementaux peuvent survenir au sein des populations affectées, ce qui modifie la distribution des données et entrave la performance d'un modèle, lequel a été entraîné sur des données recueillies avant ces changements. Toutes ces modifications des conditions contextuelles peuvent avoir un effet significatif sur la performance du système d'IA et sur la manière dont celui-ci agit sur les personnes, les groupes, les communautés affectées et la société en général.

Mise en œuvre de la révision itérative

Si la méthodologie HUDERIA offre une certaine flexibilité en ce qui concerne les modalités exactes, les seuils, les déclencheurs et les mécanismes de suivi et de gouvernance du processus de révision itérative, les principes suivants pourraient être pris en compte :

a. la révision continue de l'HUDERIA joue un rôle essentiel dans le maintien de son efficacité et de sa fiabilité ;

b. un plan doit être établi pour suivre les impacts et pour réévaluer l'HUDERIA à chaque phase du cycle de vie du projet jusqu'au retrait ou à la mise hors service du système ;

c. les processus utilisés pour la révision itérative devraient rester aussi réactifs que possible à la manière dont le système d'IA interagit avec ses environnements opérationnels et avec les personnes touchées (par exemple les domaines d'application possibles du système d'IA, l'émergence de nouvelles formes d'utilisation abusive du système) ;

d. dans des contextes qui évoluent rapidement ou qui changent, il peut être nécessaire de procéder à des réévaluations plus fréquentes.

Partie II

MODÈLE DE RESSOURCES POUR L'ANALYSE DES RISQUES FONDÉE SUR LE CONTEXTE (COBRA)

Préface

Qu'est-ce que l'HUDERIA ?

L'Évaluation des risques et des impacts des systèmes d'intelligence artificielle (IA) du point de vue des droits humains, de la démocratie et de l'État de droit (« l'HUDERIA ») est un guide qui fournit une approche structurée d'évaluation des risques et des impacts des systèmes d'IA spécifiquement adaptée à la protection et à la promotion des droits humains, de la démocratie et de l'État de droit. Elle est destinée à jouer un rôle unique et essentiel à l'intersection des normes internationales en matière de droits humains et des cadres techniques existants sur la gestion des risques dans le contexte de l'IA.

L'HUDERIA peut être utilisée par les acteurs publics et privés pour aider à identifier les risques et les impacts sur les droits humains, la démocratie et l'État de droit tout au long du cycle de vie des systèmes d'IA, et à y apporter une réponse.

Relation avec la Convention-cadre

L'HUDERIA est un guide autonome de nature non contraignante, sans effet juridique. Elle n'est pas obligatoire et ne constitue pas non plus un instrument d'interprétation de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit, ci-après dénommée « la Convention-cadre ». Par ailleurs, bien que l'HUDERIA ait une fonction de facilitation, elle ne constitue pas un moyen de mise en œuvre de la Convention-cadre. De nombreux cadres, politiques, lignes directrices, normes ou outils existants ou futurs peuvent être utilisés pour contribuer à la gestion des risques et des impacts de l'IA, y compris l'HUDERIA.

Les Parties à la Convention-cadre disposent de la souplesse nécessaire pour utiliser ou adapter ces orientations, en tout ou en partie, afin d'élaborer de nouvelles approches d'évaluation des risques ou d'utiliser ou adapter les approches existantes, conformément à leurs législations nationales applicables, à condition que les Parties respectent pleinement leurs obligations au titre de la Convention-cadre, y compris le socle en matière de gestion des risques et des impacts défini dans son chapitre V.

Orientation générale et spécifique

L'HUDERIA comprend à la fois la méthodologie HUDERIA et le modèle HUDERIA.

Au niveau général, la méthodologie HUDERIA décrit les concepts, processus et éléments de haut niveau guidant les activités d'évaluation des risques et des impacts des systèmes d'IA susceptibles d'avoir des effets sur les droits humains, la démocratie et l'État de droit.

Au niveau spécifique, le modèle HUDERIA fournit des supports et ressources complémentaires (tels que des outils flexibles pertinents pour différents éléments du processus HUDERIA, des ressources à titre illustratif et des recommandations évolutives) qui peuvent aider à la mise en œuvre de la méthodologie HUDERIA. Le présent document fait partie du modèle HUDERIA et contient les ressources COBRA, qui contribuent à opérationnaliser le processus COBRA grâce à des questions structurées, des explications et des exemples destinés à soutenir l'identification et l'évaluation systématiques des risques.

Ces ressources constituent une bibliothèque de connaissances permettant de faciliter la prise en compte et la gestion des risques et impacts liés aux droits humains, à la démocratie et à l'État de droit, y compris dans d'autres approches de gestion des risques.

Pour plus de clarté, les ressources et orientations contenues dans le modèle HUDERIA ne sont pas présentées comme des exemples de bonnes pratiques et ne fixent pas de normes minimales ; elles peuvent en revanche servir de source d'inspiration pour des approches adaptées par les acteurs publics et privés en matière de gestion des risques, susceptibles d'être reprises par les États.

Ressources COBRA

Introduction

Le présent document vise à fournir des orientations illustratives pour l'analyse contextuelle des risques, premier élément de la méthodologie HUDERIA. Il contient différents éléments, y compris du matériel pouvant être utilisé pour guider les réflexions dans le cadre de l'évaluation des risques et impacts potentiels des systèmes d'IA. Conformément à l'approche souple de la méthodologie HUDERIA, ces éléments peuvent être ajustés quand nécessaire. Ce matériel pourrait également être utilisé à l'avenir pour développer des outils interactifs (par exemple une plateforme en ligne, des flux de travail interactifs) afin de faciliter la conduite des évaluations des risques et impacts.

Les éléments de la présente section peuvent servir à identifier et évaluer les facteurs de risque influençant l'ampleur, la portée, la probabilité et la réversibilité d'impacts négatifs sur les droits humains, la démocratie et l'État de droit, liés au contexte d'application d'un système d'IA (**ressources COBRA A**), au contexte de conception et de développement du système d'IA (**ressources COBRA B**), ainsi qu'au contexte de déploiement du système d'IA (**ressources COBRA C**). Ces éléments devraient être examinés et pris en compte de manière itérative tout au long du cycle de vie d'un système d'IA.

Pour faciliter la cartographie des impacts potentiels sur les droits humains, la démocratie et l'État de droit, les **ressources COBRA E** recensent des domaines de préoccupation potentiels du point de vue des droits humains, de la démocratie et de l'État de droit, en fournissant une description illustrative, et donnent des exemples de risques potentiels liés à l'IA pour chacun de ces domaines. Cette liste pourrait être utilisée conjointement avec les **ressources COBRA F**, qui recensent et décrivent brièvement les secteurs et domaines potentiellement concernés et les mettent en correspondance avec les domaines de préoccupation potentiels identifiés dans les **ressources COBRA E**.

Aperçu des ressources COBRA A, B et C

Cette section fournit des exemples illustratifs des facteurs de risque potentiels affectant la probabilité d'impacts négatifs sur les droits humains, la démocratie et l'État de droit, découlant :

- du contexte d'application d'un système d'IA (**ressources COBRA A**), par exemple les environnements juridiques et réglementaires dans lesquels le système est développé et utilisé, la finalité prévue du système, les catégories d'utilisateurs et de personnes concernées, ainsi que d'autres éléments relatifs à son contexte d'application, tels que d'éventuels héritages connus de biais et de discrimination ;
- du contexte de conception et de développement du système d'IA (**ressources COBRA B**), en particulier ses caractéristiques techniques, telles que les limites connues du système, les considérations relatives à la collecte, l'enrichissement, le stockage, l'utilisation et la suppression des données, ainsi que les considérations liées à l'algorithme ou au modèle lui-même, les caractéristiques techniques concernant la protection de la vie privée et des données à caractère personnel, les biais et discriminations, ainsi que l'explicabilité et l'interprétabilité ;
- du contexte de déploiement du système d'IA (**ressources COBRA C**), telles que les caractéristiques des utilisateurs visés et des groupes affectés, les mesures mises en place pour garantir la sécurité globale, prévenir les préjudices et préserver la dignité humaine, les actions prévues pour protéger la vie privée et les données à caractère personnel, atténuer les biais préjudiciables, assurer une formation adéquate, prévenir les utilisations non intentionnelles, et garantir la responsabilité et la conformité juridique.

Les ressources sont illustratives, non exhaustives et évolutives, ce qui signifie qu'elles devront être adaptées au fil du temps afin de tenir compte des évolutions, par exemple dans les domaines de la technologie et les régimes juridiques et réglementaires applicables.

Ressources COBRA A (Liste des facteurs de risque découlant du contexte d'application du système d'IA)

		Exemples de facteurs de risque potentiels
1. Secteur ou domaine dans lequel le système est conçu	A.1.1	À la lumière de sa finalité prévue, le système d'IA jouera-t-il un rôle important dans un contexte à fort impact – tel que le transport, les services sociaux, la santé ou l'éducation – où ses résultats pourraient influencer des décisions ou des actions importantes ? Quel est le contexte dans lequel le système fonctionne ? *Dans de tels contextes, toute défaillance ou dégradation des performances au cours du cycle de vie du système peut entraîner des décisions préjudiciables, des interruptions de service ou des incidents de sécurité. Il convient d'examiner si le système est susceptible d'affecter des domaines où les droits humains, la démocratie ou l'État de droit pourraient être en jeu. Les **ressources COBRA E** présentent des exemples de préjudices potentiels et d'intérêts protégés pouvant être pertinents, tandis que les **ressources COBRA F** contiennent la liste des secteurs présentant un intérêt potentiel du point de vue des droits humains, de la démocratie et de l'État de droit.* Exemple : *un système hybride d'IA utilisé pour contrôler la cinématique d'un robot chirurgical assistant les médecins lors d'une intervention médicale d'urgence.*
	A.1.2	Le système d'IA exercera-t-il une fonction à fort impact, indépendamment du secteur dans lequel il opère ? Quelle est la fonction du système ? *Même en dehors des secteurs à fort impact, de telles fonctions peuvent conduire à des décisions préjudiciables, des interruptions de service ou d'autres conséquences négatives susceptibles d'affecter les droits humains des personnes concernées ou d'avoir des effets sur la démocratie et l'État de droit – voir les **ressources COBRA E** à cet égard –, même si le système n'est pas déployé dans un contexte traditionnellement à fort impact.* ***Exemple :** le système en cours de développement est un outil de recrutement assisté par IA pour une fonction de la fonction publique. L'outil n'est pas situé dans un secteur critique pour la sécurité ; cependant, le système pourrait avoir un impact négatif sur des groupes d'individus en situation de vulnérabilité.*
2. Cadre juridique et réglementaire existant	A.2.1	Le secteur ou domaine dans lequel le système d'IA sera déployé est-il soumis à une surveillance juridique ou réglementaire étendue ? Le cas échéant, quelles sont les obligations ou normes spécifiques (par exemple en matière de conformité, de certification, de supervision ou d'éthique) susceptibles de s'appliquer au système d'IA ou à ses résultats ? *Une surveillance juridique ou réglementaire étendue peut en soi constituer un indicateur de contexte à fort impact. Lorsqu'un tel encadrement existe, il est important d'identifier les obligations et normes pertinentes susceptibles d'interagir avec les considérations de l'HUDERIA ou de les éclairer. L'évaluation HUDERIA se concentre spécifiquement sur les impacts potentiels en matière de droits humains, de démocratie et d'État de droit. Les autres obligations sectorielles (par exemple en matière de stabilité financière, de sécurité médicale ou de cybersécurité) demeurent régies par leurs cadres réglementaires respectifs. Le cas échéant, les évaluations ou certifications existantes (obligatoires ou volontaires) peuvent être mentionnées comme documents de référence afin de renforcer la base factuelle de l'analyse HUDERIA — mais elles ne la remplacent pas et ne se confondent pas avec elle.* ***Exemple :** une banque utilisant un outil d'évaluation des risques pour prédire le défaut de remboursement des emprunteurs opère dans le secteur financier, où une réglementation étendue relative à l'abus de marché, à la gestion des risques, au droit des sociétés et au droit de la concurrence est historiquement en place.*
3. Portée du déploiement	A.3.1	Dans un scénario où le système d'IA est déployé conformément à sa conception, à quelle échelle affectera-t-il directement et/ou indirectement des personnes et des groupes (populations locales, populations nationales, populations mondiales) ? *Cette question vise à examiner la portée spatiale et juridictionnelle des impacts potentiels du système d'IA. Le changement d'échelle peut modifier de manière substantielle le profil de risque du système, avec des effets possibles sur les populations locales, sur les communautés et groupes qui les composent, ainsi que sur les processus sociaux et politiques. Ces impacts peuvent être significatifs du point de vue des droits humains, de la démocratie et de l'État de droit (voir les **ressources COBRA E**, en particulier les domaines (16) Démocratie et (17) État de droit). Lorsque les systèmes sont introduits progressivement ou testés dans le cadre de projets pilotes ou d'expérimentations sur un groupe limité, ce déploiement par étapes peut également constituer une mesure d'atténuation, permettant d'identifier, d'évaluer et de gérer les impacts potentiels avant une mise en œuvre à plus grande échelle.* ***Exemple (local) :** un système prédictif de police utilisé par un service de police municipale et opérant uniquement dans les limites de la municipalité.* ***Exemple (national) :** un système de classification utilisé par un organisme public pour déterminer l'éligibilité des individus à des prestations sociales affecte des personnes et groupes au sein de la population nationale.* ***Exemple (global) :** un système de recommandation utilisé par une plateforme internationale de médias sociaux pour personnaliser la diffusion de nouvelles affecte des personnes et groupes au sein de la population mondiale.*

		Exemples de facteurs de risque potentiels
3. Portée du déploiement	A.3.2	Dans un scénario où le système d'IA se généralise de manière optimale, combien de personnes potentiellement affectées le système d'IA affectera-t-il directement et/ou indirectement ? *Cette question porte sur l'ampleur numérique et démographique de l'impact potentiel. L'utilisation du système d'IA à grande échelle ou à l'échelle de masse peut générer des effets significatifs sur les processus et institutions sociaux et politiques, avec des implications particulières pour les droits humains, la démocratie et l'État de droit (voir **ressources COBRA E**, notamment les domaines (16) Démocratie et (17) État de droit).* ***Exemple (petite échelle) :** un robot conversationnel d'IA est utilisé par une organisation non gouvernementale locale comptant 60 employés pour soutenir des survivantes et survivants de violences domestiques en fournissant des informations sur les droits juridiques et les hébergements disponibles.* ***Exemple (moyenne échelle) :** une agence régionale pour l'emploi utilise un classificateur fondé sur l'apprentissage automatique pour filtrer et hiérarchiser les candidatures aux contrats de courte durée dans le secteur public, traitant environ 10 000 candidatures par an.* ***Exemple (grande échelle) :** un assureur maladie national déploie un système prédictif pour signaler les patients « à risque » aux fins de programmes d'intervention précoce, touchant environ 1,2 million d'assurés.* ***Exemple (très grande échelle)** : une plateforme nationale d'identité numérique gérée par l'État intègre une vérification d'identité fondée sur l'IA (par exemple la reconnaissance faciale et la lecture automatisée de documents) pour contrôler l'accès aux régimes de protection sociale, touchant plus de 50 millions de personnes.*
	A.3.3	En tenant compte des impacts directs et indirects potentiels du système d'IA sur les personnes, les communautés et l'environnement, quel est l'horizon temporel le plus large dans lequel le système d'IA pourrait affecter des personnes et des groupes ? *L'utilisation du système d'IA peut entraîner des effets à long terme qui se prolongent au-delà de la période de son application, touchant les droits humains des personnes potentiellement concernées, ainsi que les processus et institutions sociaux et politiques, en particulier en ce qui concerne la démocratie et l'État de droit (voir les **ressources COBRA E**, notamment les domaines (16) Démocratie et (17) État de droit).* ***Exemple (court terme) :** un système conversationnel d'IA est utilisé par un bureau municipal du logement pour répondre aux questions des locataires concernant les arriérés de loyer, les avis d'expulsion et les procédures de plainte.* ***Exemple (moyen terme) :** une autorité régionale de l'éducation utilise un système de profilage fondé sur l'IA pour prédire les risques d'échec scolaire des élèves, entraînant des interventions ciblées ou une réorientation vers des filières professionnelles.* ***Exemple (long terme) :** un système de classification par IA est déployé par une agence nationale de migration pour évaluer les demandes d'immigration, de visa longue durée ou de citoyenneté comme présentant un risque faible, moyen ou élevé.*
4. Héritages existants de biais	A.4.1	Le ou les secteurs dans lesquels le système d'IA sera utilisé, et à partir desquels sont tirées les données servant à son entraînement, comportent-ils des schémas de discrimination, d'inégalités, de biais ou d'inexactitudes entraînant systématiquement un traitement inéquitable de groupes d'individus en situation de vulnérabilité (ou de groupes d'individus ayant des caractéristiques protégées) ? Dans l'affirmative, quelle est la probabilité que ces schémas soient reproduits ou accentués dans le fonctionnement du système ou dans ses résultats et impacts à court, moyen et long terme ? Il convient de se concentrer sur les considérations d'égalité et de non-discrimination entourant les impacts potentiels du système d'IA sur les personnes concernées. *Si de tels schémas existent, le système peut les reproduire ou les amplifier lors de son développement et de son déploiement, entraînant des résultats différenciés selon les groupes et une discrimination indirecte. À grande échelle et sur des horizons à court, moyen et long terme, cela peut renforcer les disparités et compromettre l'égalité et la non-discrimination.* ***Exemple :** l'utilisation d'un système de présélection des candidatures dans le secteur scientifique et technologique, fondé sur des données d'embauche historiques, reproduit des schémas de discrimination à l'embauche et délivre des résultats défavorables pour des groupes d'individus en situation de vulnérabilité*

		Exemples de facteurs de risque potentiels
4. Héritages existants de biais	A.4.2	Compte tenu du secteur ou du domaine dans lequel le système d'IA est utilisé, les personnes potentiellement affectées incluront-elles celles susceptibles d'être touchées de manière significative ou disproportionnée, notamment des groupes d'individus en situation de vulnérabilité (ou des groupes d'individus présentant des caractéristiques protégées), en raison de la conception et de l'utilisation du système ? *La vulnérabilité peut être situationnelle, résultant de conditions économiques, sociales, institutionnelles, technologiques ou d'autres facteurs contextuels. Il convient de se concentrer sur les considérations d'égalité et de non-discrimination entourant les impacts potentiels du système d'IA sur les personnes concernées, et de porter une attention particulière aux facteurs de risque B.5.1 à B.5.4 des* ***ressources COBRA B****.* ***Exemple :*** *un outil d'évaluation des risques utilisé par des travailleurs sociaux pour déterminer l'accès aux prestations sociales opère dans le domaine de l'administration sociale, où les personnes en situation de vulnérabilité économique ou sociale sont fortement impactées.*
	A.4.3	Le système d'IA est-il susceptible d'être accessible aux enfants ou d'avoir un impact sur eux ? *Une attention particulière doit être portée aux enfants. Leur prise en considération distincte est nécessaire, car leurs capacités évolutives, leurs relations de dépendance et leur aptitude limitée à donner un consentement éclairé ou à demander réparation créent des profils de risque spécifiques que les évaluations générales d'égalité ou de vulnérabilité ne permettent pas toujours de saisir adéquatement. Les impacts sur les enfants peuvent être disproportionnés et durables en raison de leurs vulnérabilités développementales, des asymétries de pouvoir et de la sensibilité accrue des données les concernant. Ces risques soulèvent des préoccupations particulières en matière de bien-être mental, d'égalité et de non-discrimination, ainsi que de droits de l'enfant (voir les* ***ressources COBRA E****, en particulier le domaine (14) Enfants). Des risques supplémentaires peuvent également découler d'un engagement insuffisant des parties prenantes envers les enfants ou envers les personnes responsables de leur bien-être (par exemple les parents, les tuteurs, les organisations de la société civile, les défenseurs de la protection de l'enfance) tout au long du cycle de vie du système d'IA (voir l'élément SEP de la méthodologie HUDERIA).*
5. Contexte environnemental	A.5.1	Quels seront les impacts directs et indirects du système d'IA sur l'environnement ? Dans quelle mesure son développement, son déploiement et son utilisation respectent-ils les normes, réglementations et engagements environnementaux applicables (par exemple en matière d'efficacité énergétique, d'émissions de carbone, d'utilisation des ressources, de gestion des déchets) ? *Le système d'IA peut générer des impacts directs et indirects sur l'environnement, avec des conséquences potentielles en matière d'efficacité énergétique, d'émissions de carbone, d'utilisation des ressources et de gestion des déchets. Ces impacts environnementaux peuvent, à leur tour, affecter la jouissance des droits humains, en particulier pour les personnes en situation de vulnérabilité ou les groupes présentant des caractéristiques protégées (voir les* ***ressources COBRA E****).* ***Exemple :*** *un modèle d'IA utilisé pour produire des tableaux de bord analytiques quotidiens continue de retraiter chaque nuit un jeu de données inchangé de plus de 100 To, consommant une quantité importante de puissance de calcul et d'énergie sans générer de résultats nouveaux ou pertinents. Ce choix de conception entraîne des émissions et une consommation de ressources inutiles, sans bénéfice correspondant pour le public ou les utilisateurs.*

Ressources COBRA B (Liste des facteurs de risque découlant du contexte de conception et de développement du système d'IA)

		Exemples de facteurs de risque potentiels
1. Décision de conception et définition du problème et du résultat	B.1.1	Quelles autres approches, en dehors de la création du système d'IA, pourraient raisonnablement répondre aux besoins identifiés et, comparées au système d'IA, quels bénéfices et risques présenteraient-elles – en particulier en termes d'impacts sur les droits humains, la démocratie et l'État de droit ? *Le choix de concevoir et de déployer un système d'IA doit faire l'objet d'une évaluation critique au regard d'approches alternatives susceptibles de mieux répondre aux besoins visés avec moins d'effets négatifs. Les risques incluent des préjudices marginaux ou nouveaux liés à l'introduction de l'IA dans un contexte existant, notamment en ce qui concerne les droits humains, la démocratie et l'État de droit. L'évaluation des alternatives nécessite de considérer : a) les effets négatifs potentiels sur les droits humains, la démocratie et l'État de droit (y compris les risques marginaux liés à l'introduction de l'IA) ; b) si des technologies et processus existants, y compris des techniques de gouvernance et d'atténuation, pourraient répondre aux besoins ; c) la suffisance des ressources et des données disponibles ; d) la complexité des contextes d'utilisation ; e) les bénéfices potentiels liés à l'utilisation d'une solution fondée sur l'IA, y compris le coût d'opportunité de ne pas la mettre en œuvre ; f) la nature du problème à résoudre, en particulier s'il s'agit d'un problème politique ou social. Ne pas évaluer ces alternatives peut aboutir à l'adoption d'une solution qui augmente inutilement les risques ou la complexité.*

		Exemples de facteurs de risque potentiels
2. Maturité technologique	B.2.1	La conception du système d'IA repose-t-elle sur des techniques de validation bien comprises et largement reconnues pour une finalité similaire dans le même secteur ? Si ce n'est pas le cas, quels risques peuvent découler d'une maturité technologique limitée en termes de qualité des performances du système et d'effets résultants sur les droits humains des personnes potentiellement affectées, ainsi que sur la démocratie ou l'État de droit ? *Si la conception et le développement du système d'IA ne reposent pas sur des techniques bien comprises et préalablement validées pour des finalités similaires dans le même secteur – y compris les normes industrielles et les bonnes pratiques applicables – et sans tenir compte de la finalité prévue du système ni du seuil d'erreur acceptable, il existe un risque que son immaturité technologique compromette la qualité, la fiabilité et la robustesse de ses performances. De telles lacunes peuvent accroître la probabilité d'erreurs, réduire la confiance et augmenter le risque d'impacts négatifs sur les personnes potentiellement affectées. Cependant, l'innovation implique souvent des approches nouvelles ou expérimentales qui s'écartent des pratiques établies, ce qui peut accroître l'incertitude ainsi que le risque de défaillances de performance ou d'impacts imprévus. Dans de tels cas, l'absence de précédent sectoriel doit être reconnue comme un facteur de risque accru, nécessitant la mise en place de mesures de contrôle compensatoires – telles qu'un renforcement de la supervision humaine, une évaluation indépendante par des experts, ou des déploiements pilotes sous étroite supervision. Ces mécanismes sont essentiels pour gérer l'incertitude, vérifier la fiabilité du système et garantir que le progrès technologique se déroule de manière sûre et responsable.*
3. Système existant	B.3.1	Si le système d'IA est destiné à remplacer un système existant, les lacunes et risques de ce dernier ont-ils été identifiés et atténués ? *Lorsqu'un système d'IA remplace un système humain, technique ou hybride remplissant une fonction similaire, ou lorsque son déploiement conduit les opérateurs humains à se fier excessivement aux résultats produits par l'IA, des risques peuvent survenir si les limites, défaillances ou préjudices documentés du système précédent ne sont pas correctement compris et pris en compte. Le fait de ne pas intégrer ces éléments peut conduire à reproduire ou amplifier les risques existants, plutôt qu'à les réduire. La qualité et l'impact des performances du système d'IA dépendront en partie de sa capacité à tirer les leçons du système remplacé, à corriger ses faiblesses et à améliorer ou renforcer ses fonctionnalités.*
	B.3.2	Le système d'IA remplace-t-il un système humain, technique ou hybride constituant une infrastructure critique ou remplissant une fonction à fort impact ? Le cas échéant, quels risques (par exemple des pannes ou des interruptions de service) pourraient découler du processus de mise à jour ou de remplacement du système, et comment ces risques pourraient-ils affecter les droits humains des personnes potentiellement concernées ? *Lorsqu'un système d'IA remplace un système humain, technique ou hybride constituant une infrastructure critique ou exerçant une fonction essentielle pour la sécurité ou la mission, le processus de mise à jour ou de remplacement peut engendrer des risques importants. Ces risques peuvent résulter de pannes, d'interruptions de services essentiels ou de défaillances susceptibles d'avoir des effets directs sur les droits humains des personnes concernées, ainsi que sur la démocratie ou l'État de droit. La nature critique du système remplacé amplifie l'ampleur potentielle de ces impacts.*
4. Contexte de cybersécurité	B.4.1	Quelles motivations et opportunités le système d'IA présente-t-il pour des acteurs malveillants cherchant à le pirater, le corrompre ou le détourner ? Quels risques découlent de tentatives visant à compromettre sa sûreté, sa sécurité et sa robustesse (par exemple par des attaques antagonistes, l'empoisonnement de données, l'inversion de modèle ou les violations de données) ? Quels risques une telle exploitation malveillante pourrait-elle faire peser sur les droits humains des personnes potentiellement affectées, ainsi que sur la démocratie ou l'État de droit ? *Les systèmes d'IA peuvent susciter chez certains acteurs malveillants des motivations et des occasions de les pirater, de les corrompre ou de les manipuler de quelque manière que ce soit. De telles actions peuvent être motivées par un gain financier, des objectifs politiques ou d'autres avantages perçus, et peuvent conduire au détournement du système afin de faciliter des atteintes aux droits humains, de miner la confiance publique ou de causer un préjudice aux personnes concernées.*

		Exemples de facteurs de risque potentiels
5. Qualité des données et protection des données personnelles	B.5.1	Les données utilisées pour la conception et le développement du système d'IA sont-elles suffisamment représentatives des personnes et groupes concernés, exactes, complètes, fiables, pertinentes, appropriées, à jour et d'une quantité et qualité adéquates pour le cas d'usage, le domaine, la fonction et la finalité prévus ? Dans le cas contraire, quels risques pourraient découler pour les personnes concernées des insuffisances identifiées ? Quelles techniques ont été utilisées pour réaliser cette évaluation ? *Si les données utilisées pour la production du système d'IA ne sont pas suffisamment représentatives, exactes, complètes, fiables, pertinentes, appropriées, à jour et d'une qualité et quantité suffisantes pour l'usage, le domaine, la fonction et la finalité prévus, des risques significatifs peuvent en résulter. Ces risques peuvent provenir : a) des conséquences de l'utilisation de données inexactes, incohérentes ou incomplètes ; b) de l'absence d'enregistrement, de traçabilité ou d'auditabilité suffisante de la provenance et de la lignée des données tout au long du cycle de vie du système ; c) d'erreurs de mesure ou de biais introduits lors de la collecte des données (qu'ils proviennent d'une intervention humaine ou non) ; d) des données manquantes ou inutilisables dans les ensembles de données collectés ou achetés ; e) du manque de transparence et d'accessibilité des processus d'étiquetage ou d'annotation pour l'audit, la supervision et l'examen ; f) de biais introduits par les étiqueteurs ou annotateurs humains, y compris par l'usage de variables de substitution, en l'absence de garanties suffisantes ; g) de biais introduits par des processus d'étiquetage ou d'annotation automatisés, s'ils ne font pas l'objet d'une supervision humaine adéquate ni de procédures transparentes. De telles insuffisances peuvent compromettre la fiabilité, la crédibilité et l'équité du système d'IA et porter atteinte aux droits humains des personnes potentiellement affectées.*
	B.5.2	La conception et le développement du système d'IA impliquent-ils l'utilisation de données à caractère personnel ? Le cas échéant, quels risques pour la vie privée et la protection des données personnelles peuvent en découler ? *Si des données à caractère personnel sont utilisées dans la conception et le développement du système d'IA, il existe un risque de non-respect des exigences en matière de protection des données personnelles. Le cas échéant, des risques peuvent notamment être associés : a) à une information insuffisante des personnes concernées et des parties prenantes quant au consentement ou à tout autre fondement juridique du traitement ; b) au recours à un consentement implicite ou à des bases juridiques floues, sans consultation des personnes concernées et des parties prenantes sur l'acceptabilité de l'utilisation des données ; c) à des garanties insuffisantes pour les individus concernant l'utilisation de leurs données une fois celles-ci partagées ou utilisées à des fins d'entraînement, y compris, le cas échéant, la possibilité de retirer ou de limiter leur utilisation ultérieure ; d) à des situations de divergence potentielle entre les règles applicables en matière de protection des données – telles que la limitation de la finalité – et les pratiques de l'IA fondées sur la réutilisation de données à grande échelle ou de manière itérative ; e) à la possibilité de désanonymisation ou de réidentification par recoupement avec des ensembles de données existants, accessibles au public ou facilement disponibles, si ces risques ne sont pas correctement gérés. De telles insuffisances peuvent porter atteinte à la vie privée et à la protection des données personnelles.*
	B.5.3	Le système d'IA utilisera-t-il des données dynamiques collectées et traitées en temps réel (ou quasi réel) pour assurer un apprentissage continu, une adaptation ou une optimisation des performances ? Le cas échéant, quels risques peuvent survenir en matière de sécurité, de sûreté, de fiabilité, de robustesse, de qualité et d'intégrité des données, et – le cas échéant – de non-discrimination et d'atténuation des biais ? *Si le système d'IA utilise des données dynamiques collectées et traitées en temps réel (ou quasi réel) à des fins d'apprentissage continu, d'adaptation ou d'optimisation des performances, de nouveaux risques peuvent apparaître. Ces risques peuvent être liés à des menaces pesant sur la sécurité, la sûreté, la fiabilité, la robustesse, la qualité et l'intégrité des données. L'adaptation en temps réel peut également introduire ou amplifier des biais et des discriminations si elle n'est pas correctement gérée. De tels risques peuvent miner la confiance et affecter négativement les droits humains des personnes concernées.*
	B.5.4	Dans quelle mesure le domaine des données collectées ou acquises et leur type présentent-ils des risques de décalages ou dérives distributionnels rapides ou inattendus, susceptibles d'avoir un impact négatif sur la précision et les performances du système d'IA ? *Le domaine dans lequel les données sont collectées ou acquises ainsi que le type de données utilisées peuvent en effet présenter des risques de dérives ou décalages distributionnels rapides ou imprévus. De tels changements peuvent compromettre la précision, la fiabilité et les performances du système d'IA, entraînant des erreurs ou des effets préjudiciables pour les personnes potentiellement affectées s'ils ne sont pas anticipés et pris en compte. L'absence d'évaluation dynamique, de réévaluation, de validation et de surveillance continue accroît ces risques.*

		Exemples de facteurs de risque potentiels
6. Contexte de développement et de mise en œuvre du modèle	B.6.1	Quels risques potentiels de biais et de discrimination indirecte peuvent survenir dans le cadre de la conception, du développement et de la mise en œuvre du modèle ? *La conception, le développement et la mise en œuvre du modèle peuvent introduire des risques de biais et de discrimination indirecte. Ces risques peuvent apparaître : a) lors de l'ingénierie des variables (manuelle ou automatisée), par regroupement, désagrégation ou exclusion de caractéristiques d'entrée liées à des caractéristiques protégées ou sensibles – ou à leurs proxys –, entraînant un biais émergent ; b) par le biais d'inférences générées par des mécanismes d'apprentissage du modèle qui sont déraisonnables, inéquitables, qui produisent un impact différencié ou qui sont influencés par des proxys cachés aux caractéristiques discriminatoires, façonnant ainsi les résultats de manière discriminatoire. Si ces risques ne sont pas traités, ils peuvent conduire à un traitement inégal et à une discrimination systémique, compromettant les droits humains des personnes potentiellement affectées.*
	B.6.2	Le système d'IA est-il construit sur des techniques intrinsèquement difficiles à expliquer, prédire ou vérifier pleinement, telles que les systèmes non déterministes, les modèles probabilistes ou les modèles évolutifs/dynamiques ? Dans l'affirmative, quels risques potentiels cela pourrait-il poser pour les droits humains des personnes potentiellement affectées, notamment lorsque le système d'IA interagit directement avec des individus ? *Si le système d'IA repose sur des techniques difficiles à prédire, à vérifier ou à expliquer pleinement – telles que les systèmes non déterministes (résultats variables pour une même entrée), les modèles probabilistes (résultats fondés sur des probabilités, comme les modèles bayésiens ou les grands modèles de langage) ou les modèles évolutifs/dynamiques (systèmes qui continuent à apprendre ou à s'adapter dans le temps) –, des risques spécifiques peuvent apparaître. Ces risques peuvent être liés : a) à une réduction de la transparence, limitant la capacité des personnes concernées à comprendre, à contester ou à demander réparation pour les résultats produits ; b) à un comportement imprévisible ou instable, lorsque les résultats varient de manière difficile à anticiper ou à auditer ; c) à un désalignement entre l'intelligibilité du modèle (ou son accessibilité) et les exigences sectorielles, juridiques ou fonctionnelles liées à son usage prévu. Ces risques peuvent être amplifiés en l'absence de mesures d'atténuation appropriées, notamment lorsque les organisations ne disposent pas des capacités nécessaires pour fournir des mécanismes explicatifs complémentaires (par exemple des modèles de substitution ou des analyses d'importance des variables), ou lorsque aucune garantie n'existe pour assurer la réversibilité ou l'intervention humaine. De tels risques peuvent porter atteinte aux droits humains, à la démocratie et à l'État de droit, en réduisant la responsabilité, la transparence et la confiance dans les résultats produits par le système.*
	B.6.3	Dans quelle mesure une évaluation, une vérification et une validation appropriées du modèle d'IA ont-elles été assurées tout au long de son cycle de vie, et comment les lacunes éventuelles pourraient-elles influencer le niveau de risque associé à son déploiement ou à son utilisation ? *Bien que les mécanismes d'évaluation, de vérification et de validation constituent en eux-mêmes des mesures d'atténuation et de contrôle, leur absence ou leur insuffisance représente un facteur de risque majeur affectant la fiabilité et la sécurité du modèle d'IA. Si le modèle d'IA n'est pas soumis à une surveillance, une évaluation, une vérification et une validation suffisantes, il existe un risque que des erreurs, lacunes ou biais dans sa conception et son fonctionnement ne soient pas détectés. En l'absence de processus transparents, incluant notamment une évaluation indépendante par des experts et une revue par les pairs externe, la fiabilité, l'équité et la sécurité du système peuvent être compromises, entraînant des impacts préjudiciables pour les personnes potentiellement affectées et leurs droits humains, ainsi que des effets négatifs sur la démocratie et l'État de droit.*
	B.6.4	Dans quelle mesure le système d'IA comprend-il des processus de suivi et de réévaluation régulière afin de tenir compte des évolutions du monde réel susceptibles d'entraîner des dérives de concept ou des variations dans les distributions de données sous-jacentes ? Quels risques peuvent survenir pour les personnes concernées si de tels processus sont absents ou insuffisants ? *Le suivi et la réévaluation fonctionnent comme des mesures de contrôle visant à maintenir les performances du système dans le temps ; c'est leur absence ou leur insuffisance qui crée le risque de dégradation ou d'iniquité dans la pratique. Si le système d'IA n'est pas surveillé et réévalué régulièrement pour suivre l'évolution des conditions réelles, il existe un risque qu'il s'écarte progressivement de sa finalité prévue ou qu'il ne reflète plus les conditions dans lesquelles il a été conçu et validé (dérive conceptuelle ou modifications des distributions de données sous-jacentes). Ces changements peuvent dégrader l'exactitude, l'équité et la fiabilité du système au fil du temps, entraînant des effets préjudiciables pour les personnes potentiellement affectées.*

		Exemples de facteurs de risque potentiels
6. Contexte de développement et de mise en œuvre du modèle	B.6.5	Dans quelle mesure les indicateurs de performance du système d'IA – au-delà de la simple précision (par exemple la sensibilité, la précision ou la spécificité) – sont-ils pertinemment choisis, suivis et communiqués de manière à refléter le contexte spécifique du cas d'utilisation, à minimiser les risques de mauvaise interprétation ou de mésusage par les utilisateurs et les parties prenantes, et à aider à garantir la transparence et la responsabilité tout au long du cycle de vie du système ? Quels risques peuvent en découler pour les personnes concernées si cela n'est pas le cas ? La sélection et la communication des indicateurs de performance agissent comme des garanties soutenant la transparence et la responsabilité ; lorsque ces mécanismes sont faibles ou incomplets, ils deviennent eux-mêmes des facteurs de risque. Si les indicateurs de performance du système d'IA ne sont pas soigneusement choisis, définis en fonction du contexte, suivis et communiqués de manière transparente, plusieurs risques peuvent survenir, notamment : a) la priorisation de certains types d'erreurs (par exemple les faux positifs ou les faux négatifs) sans tenir compte du contexte d'utilisation ni des impacts potentiellement disproportionnés des taux d'erreur différenciés sur les personnes en situation de vulnérabilité ou sur les groupes présentant des caractéristiques protégées ; b) la limitation de l'évaluation de la performance à la précision, en négligeant d'autres mesures pertinentes (par exemple la sensibilité, la précision, la spécificité), ce qui peut masquer des biais ou des faiblesses affectant la fiabilité et l'équité ; c) la présentation des indicateurs dans des formats trop techniques, inaccessibles ou dépourvus d'explications contextuelles, empêchant les utilisateurs et les parties prenantes de bien interpréter la performance du système ; et d) un suivi ou un rapport irrégulier qui empêche d'identifier les variations de performance entre les groupes de population ou dans le temps.
	B.6.6	Si le système contribue de manière substantielle à l'information ou à la prise de décisions susceptibles d'avoir un impact sur les droits humains d'une manière nécessitant des mécanismes de supervision, de réexamen ou d'intervention humaine, de telles fonctionnalités ont-elles été intégrées dès les phases de conception et de mise en œuvre du modèle afin de garantir que le système puisse soutenir l'équité procédurale et substantielle lors de son utilisation future ? *Lorsqu'un système d'IA est capable de contribuer de manière substantielle à l'information ou à la prise de décisions ayant une incidence sur les droits des individus ou sur leur accès à des services, sa conception et sa mise en œuvre devraient intégrer des fonctionnalités aidant à rendre l'équité procédurale et substantielle effectivement opérationnelle. Ces fonctionnalités peuvent inclure : a) la possibilité de révision ou d'intervention humaine, tant en amont (*ex ante*) qu'en aval (*ex post*) ; b) des contraintes opérationnelles intégrées empêchant le système de dépasser de manière autonome son champ d'application prévu ; c) des mécanismes de traçabilité, d'auditabilité et d'explicabilité ; et d) une attribution claire des responsabilités de supervision à des personnes disposant des compétences, de la formation et de l'autorité nécessaires. Ces éléments contribuent à garantir que, une fois déployé, le système d'IA fonctionne dans un cadre permettant l'équité, la responsabilité et un contrôle humain effectif.*
	B.6.7	Quel est le potentiel de réutilisation ou de détournement du système d'IA d'une manière susceptible de soulever des préoccupations éthiques, juridiques ou réglementaires supplémentaires ? *La réutilisation (élargissement du champ d'application, double usage, montée en capacité) peut modifier de manière significative le profil de risque du système – notamment du point de vue de la protection des données à caractère personnel – en modifiant le contexte, les usages des données ou les chemins décisionnels. Le potentiel d'une telle réutilisation provient souvent des choix effectués lors de la conception et du développement (par exemple, architectures modulaires, modèles transférables ou interfaces ouvertes). L'HUDERIA reste pertinente lorsqu'un système est réutilisé ou que sa fonction évolue, car ces changements peuvent générer de nouveaux impacts – ou en amplifier certains – sur les droits humains, la démocratie et l'État de droit. Dans de tels cas, l'évaluation devrait être mise à jour ou répétée afin de tenir compte du nouveau contexte.* ***Exemple :** un système de vision par ordinateur utilisé pour reconnaître le visage des individus à des fins d'authentification d'identité est réutilisé pour l'identification biométrique à distance, en temps réel, de petits délinquants lors d'un concert.*

Ressources COBRA C (Liste des facteurs de risque découlant du contexte de déploiement du système d'IA)

		Exemples de facteurs de risque potentiels
1. Vie privée et protection des données personnelles	C.1.1	Le système d'IA traite-t-il des données à caractère personnel, y compris des catégories particulières de données (telles que définies, par exemple, à l'article 6 de la Convention 108+ ou dans d'autres cadres pertinents) ? Le cas échéant, quels risques peuvent survenir en ce qui concerne le droit au respect de la vie privée et la protection des données des personnes potentiellement concernées, ainsi que les responsabilités des acteurs chargés du développement, du déploiement ou de l'utilisation du système en matière de gouvernance et de protection des données ? *Si le système d'IA traite des données à caractère personnel, il existe un risque que ce traitement ne soit pas conforme aux lois ou aux normes applicables en matière de protection des données personnelles et de respect de la vie privée. Les risques incluent notamment : la collecte ou l'utilisation illicite ou déloyale de données, des garanties insuffisantes pour protéger les données traitées, ainsi qu'un manque de responsabilité ou de traçabilité dans la gestion et la divulgation éventuelle des données à caractère personnel dans les résultats du système.*
	C.1.2	Le système d'IA est-il conçu pour effectuer une curation, un profilage, une prédiction ou une orientation comportementale ciblés individuellement ? Le cas échéant, quels risques peuvent survenir, y compris lorsque les personnes potentiellement concernées n'ont pas accès à une information suffisamment claire et accessible à cet égard ? *Si le système d'IA est conçu pour une curation, un profilage, une prédiction ou une orientation comportementale ciblant des individus, il existe un risque que les personnes concernées ne bénéficient pas du niveau le plus élevé de transparence et de droits à l'information prévu par le cadre juridique applicable (en particulier lorsque des données sensibles sont en jeu, telles que définies dans des instruments comme la Convention 108+). En particulier, elles peuvent ne pas disposer : a) d'informations claires sur la collecte et l'utilisation de leurs données personnelles ; b) d'une explication compréhensible et non technique de la logique sous-tendant les résultats produits par le système ; c) d'une indication explicite de la finalité de la curation, du profilage, de la prédiction ou de l'orientation comportementale ; d) d'informations sur les catégories de personnes ou d'organismes auxquelles leurs données, leur profil ou les résultats du traitement peuvent être communiqués.*
2. Non-discrimination et biais	C.2.1	De quelle manière les modalités de déploiement – notamment le lieu, la manière et les acteurs chargés de la mise en œuvre du système d'IA – peuvent-elles influencer l'apparition ou l'amplification de biais ou d'effets discriminatoires identifiés plus tôt dans le cycle de vie ? *Les biais et les schémas discriminatoires identifiés lors de la conception ou du développement peuvent être soit renforcés, soit atténués, selon la manière dont le système d'IA est déployé.*
3. Opérateurs du système	C.3.1	Quels risques potentiels pour les droits humains des personnes chargées de faire fonctionner le système peuvent découler du déploiement du système d'IA ? *Le déploiement du système d'IA peut avoir des effets négatifs sur les droits humains des personnes qui l'exploitent – par exemple en entraînant une surveillance excessive, un stress indu, des conditions de travail dangereuses ou une érosion de l'autonomie professionnelle. De tels impacts peuvent non seulement compromettre leurs droits humains, mais également nuire à l'intégrité et à la responsabilité du fonctionnement du système.*
4. Formation	C.4.1	Les opérateurs du système sont-ils suffisamment formés pour comprendre pleinement les limites du système et intervenir efficacement dans des situations de complexité, d'incertitude, d'anomalie ou de défaillance ? Dans le cas contraire, quels risques pour les droits humains, la démocratie, l'État de droit et la responsabilité du système peuvent en découler ? *Si les opérateurs du système d'IA ne sont pas suffisamment formés, ils peuvent ne pas comprendre les usages prévus et les limites du système ni reconnaître les situations complexes, incertaines, anormales ou défaillantes. Cela pourrait les empêcher d'exercer un jugement humain approprié et d'intervenir efficacement, augmentant ainsi le risque de préjudice pour les personnes potentiellement concernées et compromettant la responsabilité, la transparence et la fiabilité du système.*

		Exemples de facteurs de risque potentiels
5. Intervention humaine dans la boucle	C.5.1	Le système d'IA présente-t-il un niveau élevé d'automatisation ou d'« autonomie opérationnelle » ? Le cas échéant, quels risques peuvent découler d'interactions avec les personnes concernées qui ne sont pas entièrement médiées par un contrôle humain ? *Les systèmes d'IA dotés d'un niveau élevé d'automatisation ou d'autonomie opérationnelle peuvent interagir avec les personnes potentiellement concernées d'une manière non entièrement encadrée par le contrôle humain. Cela crée des risques d'atteinte à la dignité humaine et à l'autonomie individuelle, ainsi que d'érosion de la transparence, de la supervision, de la responsabilité et de l'imputabilité. Si le remplacement total des opérateurs humains ne se produit que dans certains contextes, il existe de nombreuses situations dans lesquelles les opérateurs humains restent impliqués mais tendent à surenchérir sur les résultats produits par l'IA. Cela peut conduire à une illusion de supervision humaine effective, la tâche étant en réalité exécutée principalement par le système, malgré une supervision nominale. En l'absence de garanties appropriées, de telles interactions peuvent avoir des effets négatifs sur les droits humains des personnes potentiellement concernées.*
6. Utilisations hors du champ prévu	C.26.1	Comment les dysfonctionnements, les mauvais usages ou les utilisations malveillantes du système d'IA peuvent-ils affecter les droits humains des personnes potentiellement concernées, ainsi que la démocratie et l'État de droit ? Dans quelle mesure ces risques sont-ils identifiés ? *Il existe un risque que le système d'IA dysfonctionne, soit utilisé de manière inappropriée ou fasse l'objet d'un usage malveillant – par exemple en raison d'une panne ou d'une défaillance technique, d'une utilisation dépassant le champ d'application prévu, ou d'une application intentionnellement abusive. De tels événements peuvent avoir des effets négatifs sur les droits humains des personnes potentiellement concernées, notamment en compromettant la transparence, la supervision, la responsabilité et l'imputabilité du système. Cette catégorie de risques peut également porter atteinte à l'autonomie individuelle et, lorsqu'elle est amplifiée à grande échelle – par exemple par la diffusion de désinformation ou de fausses informations –, avoir des répercussions sur la démocratie et l'État de droit.*
7. Proximité avec la prise de décision ou l'action	C.7.1	Dans quelle mesure le système d'IA prend-il, réexamine-t-il ou influence -t-il de manière substantielle des décisions ou actions susceptibles d'affecter les droits humains, la démocratie ou l'État de droit ? Comment ce degré d'implication influence-t-il le niveau potentiel de risque ? *Le degré de proximité du système d'IA avec l'activité concernée (prise de décision ou action) peut amplifier ou réduire son impact potentiel sur les droits humains. Lorsque le système prend ou réexamine directement des décisions, ou les influence de manière substantielle, le risque d'effets négatifs sur les droits humains est généralement plus élevé que pour les systèmes qui n'apportent qu'un soutien périphérique ou n'exercent aucune influence significative.*

Ressources COBRA E (Domaines illustratifs[20] de préoccupation potentielle du point de vue des droits humains, de la démocratie et de l'État de droit)

Cette section fournit un outil pouvant être utilisé pour réaliser ou éclairer l'évaluation COBRA. Elle présente des domaines de préoccupation potentielle (colonne de gauche) accompagnés, pour chacun d'eux, d'une description indicative, d'exemples de risques potentiels liés à l'IA, des domaines potentiellement concernés, ainsi que de références aux instruments pertinents en matière de droits humains et, le cas échéant, à d'autres instruments juridiques pertinents.

La liste doit être comprise comme un ensemble illustratif d'éléments destinés à aider à l'évaluation des risques liés à l'IA dans différents contextes. Cette ressource est indicative, non exhaustive et évolutive, ce qui signifie qu'elle est appelée à évoluer en fonction des développements intervenant dans les domaines technologiques et dans les régimes juridiques pertinents.

Dix-sept domaines de préoccupation potentielle ont été identifiés. Ces domaines représentent des aspects fondamentaux consacrés dans divers instruments internationaux relatifs aux droits humains, qui soulignent la protection de la dignité humaine et de l'autonomie individuelle.

20. Toute référence aux instruments internationaux relatifs aux droits humains dans ce tableau est incluse à des fins explicatives. Ces références ne s'appliquent qu'aux États qui sont parties à ces instruments. Chaque État est censé appliquer ses propres lois conformément à ses obligations juridiques internationales, ce qui pourrait inclure l'encouragement du secteur privé à respecter et à soutenir les droits humains, notamment tels qu'ils sont énoncés dans les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits humains.

(1) Intégrité physique et psychique, et morale et dignité humaine	**Description illustrative** Les dispositions relatives aux droits humains prévoient un certain nombre de protections en ce qui concerne l'intégrité physique et mentale des individus, renforçant l'importance de l'autonomie personnelle, de la propriété de soi-même et de l'autodétermination des êtres humains en ce qui concerne leur corps. Plusieurs de ces protections peuvent s'appliquer, par exemple dans les situations impliquant le recours à la force par les autorités exerçant des pouvoirs de maintien de l'ordre et lorsque des individus sont placés en détention. Dans certains pays, ces protections peuvent inclure des enquêtes sur le comportement des autorités publiques dans des situations de santé ou de danger de mort. La protection des droits humains peut s'étendre à d'autres situations où l'intégrité physique et mentale des individus peut être en jeu. **Exemples de risques potentiels liés à l'IA** Les risques liés à l'IA en matière d'intégrité personnelle, physique et mentale, ainsi que de dignité humaine, pourraient découler de questions telles que le manque de transparence dans la prise de décision, les défauts dans la conception du système (y compris lorsque cette conception n'est pas adaptée à l'âge), le développement ou l'utilisation des systèmes d'IA et le risque que les systèmes pilotés par l'IA déshumanisent et objectifient les individus. Les systèmes d'IA qui manipulent les individus peuvent également présenter des risques. **Obligations en matière de droits humains potentiellement pertinentes** Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques et ses protocoles facultatifs :** articles 6, 7, 9 et 10 ; **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** articles 7 (travail décent) et 12 (santé) ; **Convention européenne des droits de l'homme (STE n° 5) et ses protocoles additionnels :** articles 2 (Droit à la vie), 3 (Interdiction de la torture) et 13 (Droit à un recours effectif) ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 7 (protection des enfants et des adolescents), 8 (protection de la maternité), 11 (protection de la santé), 13 (assistance sociale et médicale), 14 (systèmes de protection sociale), 16 (protection des familles), 17 (protection des enfants), 18 (protection des travailleurs migrants et de leur famille), 23 (protection sociale des personnes âgées), 26 (dignité dans le travail), 30 (protection contre la pauvreté et l'exclusion sociale) et 31 (logement) ; **Charte des droits fondamentaux de l'Union européenne (ci-après « Charte de l'Union européenne ») :** articles 1 (Dignité humaine), 3 (Droit à l'intégrité de la personne) et 4 (Interdiction de la torture et des peines ou traitements inhumains ou dégradants) ; **Convention américaine relative aux droits de l'homme :** articles 3 (Droit à la reconnaissance de la personnalité juridique), 4 (Droit à la vie) et 5 (Droit à l'intégrité de la personne), et article 12 (Droit à l'alimentation) du Protocole de San Salvador.
(2) Liberté physique et sûreté, mouvement et résidence	**Description illustrative** Des préoccupations en matière de droits humains et d'État de droit peuvent apparaître en relation avec le traitement physique des personnes en situation d'arrestation, de détention, de punition ou qui sont victimes de trafic d'êtres humains. Des protections existent contre l'arrestation ou la détention arbitraires, la torture, les peines ou traitements cruels, inhumains ou dégradants, les restrictions injustifiées à la liberté de circulation, le trafic d'êtres humains, et pour garantir l'accès à un recours et à la justice.

(2) Liberté physique et sûreté, mouvement et résidence	**Exemples de risques potentiels liés à l'IA**
	La collecte de données à caractère personnel et/ou l'utilisation d'outils d'IA peuvent permettre ou entraîner la déduction des comportements ou des activités telles que l'arrestation ou la détention, ou les restrictions au déplacement et au séjour, même si ces déductions sont incorrectes ; de telles décisions incorrectes prises ou éclairées par des outils d'IA qui sont mises en œuvre sans examen complémentaire suffisant peuvent conduire à un traitement injustifié ou empêcher une personne accusée de contester la légalité de sa détention. Les systèmes d'IA peuvent être utilisés pour faciliter ou permettre la traite des êtres humains, par exemple en générant du contenu utilisé pour solliciter des enfants à des fins sexuelles ou attirer des personnes dans des situations d'exploitation. Comme c'est le cas pour la traite des êtres humains en général, les femmes et les enfants sont exposés à un risque plus élevé. L'IA peut également être utilisée pour effectuer une surveillance de masse et pour identifier des personnes sans objectif justifiable ou raisonnablement prévisible.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** articles 7, 8, 9, 10, 11, 12, 13 et 14 ; **Convention n° 105 de l'Organisation internationale du travail (OIT)** (abolition du travail forcé) ; **Convention européenne des droits de l'homme :** articles 3 (Interdiction de la torture), 4 (Interdiction de l'esclavage et du travail forcé), 5 (Droit à la liberté et à la sûreté) et 13 (Droit à un recours effectif), et article 2 du Protocole n° 4 (Liberté de circulation) ; **Charte sociale européenne :** article 1.2 (Droit au travail) ; **Charte de l'Union européenne :** articles 3 (Droit à l'intégrité de la personne), 4 (Interdiction de la torture et des peines ou traitements inhumains ou dégradants), 5 (Interdiction de l'esclavage et du travail forcé), 6 (Droit à la liberté et à la sûreté) et 45 (Liberté de circulation et de séjour) ; **Convention américaine relative aux droits de l'homme :** articles 5 (Droit à l'intégrité de la personne), 6 (Interdiction de l'esclavage et de la servitude), 7 (Droit à la liberté de la personne) et 22 (Droit de déplacement et de résidence).
(3) Justice et administration de la justice	**Description illustrative**
	Ce domaine couvre l'accès à la justice et un ensemble complexe de règles minimales régissant l'administration de la justice dans la sphère plus large de la justice et de l'administration publique. Les protections des droits humains dans ce domaine couvrent divers droits et garanties concernant l'accès aux voies de recours, la qualité de l'examen des affaires par les tribunaux (par exemple prévoir le prononcé public d'un jugement), la qualité de la participation aux procédures judiciaires (par exemple prévoir une audience équitable et publique) et les exigences relatives à la composition des tribunaux (par exemple obliger la mise en place d'un tribunal compétent, indépendant et impartial). Cependant, dans de nombreux pays, un large éventail de questions est régi par les organes de l'administration publique, les citoyens ayant accès aux tribunaux à un stade ultérieur de la procédure en ce qui concerne certaines questions. L'administration publique est dotée de pouvoirs de décision qui lui permettent d'influencer de manière significative les individus et les intérêts importants de leur vie.
	Exemples de risques potentiels liés à l'IA
	L'utilisation de l'IA dans le système judiciaire présente des risques tels que le biais nuisible, le manque de transparence, une surveillance humaine insuffisante, l'utilisation de données erronées (ce qui est particulièrement problématique dans le contexte de l'administration de la justice), l'érosion de l'indépendance judicaire et la perte de confiance du public dans les institutions judiciaires. Elle suscite également des inquiétudes quant aux atteintes aux procédures régulières, qui pourraient entraîner des violations des droits humains.

(3) Justice et administration de la justice	**Obligations en matière de droits humains potentiellement pertinentes**
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** articles 2,3, 9, 14 et 26 ; **Convention européenne des droits de l'homme :** articles 5 (Droit à la liberté et à la sûreté), 6 (Droit à un procès équitable), 8 (Droit au respect de la vie privée et familiale) et 13 (Droit à un recours effectif), et article 3 du Protocole n° 7 (Droit d'indemnisation en cas d'erreur judiciaire) ; **Charte de l'Union européenne :** articles 41 (Droit à une bonne administration), 42 (Droit d'accès aux documents), 47 (Droit à un recours effectif et à accéder à un tribunal impartial) et 48 (Présomption d'innocence et droits de la défense) ; **Convention américaine relative aux droits de l'homme :** articles 8 (Garanties judiciaires), 10 (Droit au dédommagement), 24 (Égalité devant la loi) et 25 (Protection judiciaire).
(4) Vie privée et protection des données	**Description illustrative**
	Les droits relatifs à la protection de la vie privée garantissent que l'influence extérieure sur les affaires privées est limitée et que les personnes peuvent empêcher que leurs données personnelles, leurs comportements et leurs décisions soient divulgués ou contrôlés sans leur consentement, ce qui préserve l'autonomie et la dignité de la personne. Spécifiquement en ce qui concerne le secteur public, les lois sur la protection de la vie privée et des données protègent les informations personnelles des individus en garantissant que les autorités publiques ou les entités agissant en leur nom ne traitent que les informations personnelles qu'elles sont habilitées à traiter. Les protections relatives à la vie privée permettent aussi aux individus de garder le contrôle sur la manière dont ils développent, définissent et présentent leur identité personnelle. Elles protègent contre l'utilisation non autorisée ou la représentation erronée des attributs personnels d'un individu, tels que son nom, son image ou d'autres caractéristiques distinctives. Les aspects externes et sociaux de la vie privée et de l'autonomie individuelle fondés sur l'identité, qui concernent l'interaction avec les autres, sont au moins aussi importants. Souvent, les gens ne peuvent être eux-mêmes et se développer que s'ils peuvent établir et maintenir des relations avec d'autres. Des exemples spécifiques de ce droit fondamental sont le droit au respect de la vie familiale, le droit de se marier et, plus généralement, le droit d'établir des relations avec le monde extérieur, aussi bien dans la sphère privée que dans un cadre professionnel.
	Exemples de risques potentiels liés à l'IA
	La collecte de grandes quantités de données à caractère personnel (même lorsque ces données sont collectées de manière transparente) ainsi que l'utilisation de techniques d'IA pour en déduire des informations personnelles peuvent créer de sérieux risques pour la vie privée et l'identité. De la surveillance invasive à la réidentification de données anonymisées, en passant par l'usurpation d'identité et la manipulation comportementale, les technologies de l'IA peuvent porter atteinte à l'autonomie personnelle et exposer des informations sensibles.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** article 17 ; **Convention relative aux droits de l'enfant et ses protocoles facultatifs :** article 16 ; **Convention européenne des droits de l'homme :** articles 8 (Droit au respect de la vie privée et familiale), 10 (Liberté d'expression) et 12 (Droit au mariage) ; **Charte de l'Union européenne :** articles 7 (Respect de la vie privée et familiale), 8 (Protection des données à caractère personnel) et 9 (Droit de se marier et droit de fonder une famille) ; **Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel**, tel qu'elle a été modifiée (STE n° 108, STCE n° 223) **et ses protocoles :** articles 3, 5, 6 et 9. **Convention américaine relative aux droits de l'homme :** articles 11 (Protection de l'honneur et de la dignité de la personne), 17 (Protection de la famille), 18 (Droit à un nom) et 20 (Droit à une nationalité).

(5) Égalité et interdiction de la discrimination[21]	Description illustrative
	Les droits à une égale protection de la loi et à la non-discrimination dans la jouissance des droits et des libertés complètent les autres dispositions de fond des divers instruments internationaux relatifs aux droits humains. La discrimination peut être directe ou indirecte, résulter d'une association, être fondée sur un ou plusieurs motifs, ou être fondée sur une action ou une absence d'action. À titre de référence générale, les caractéristiques protégées mentionnées dans le corpus du droit international des droits humains – aux niveaux mondial, régional et national, selon ce qui est applicable dans chaque État – peuvent inclure, sans s'y limiter : ▸ le sexe ; ▸ la « race »[22] et la couleur ; ▸ la langue ; ▸ la religion ; ▸ les opinions politiques ou autres ; ▸ l'origine nationale ou sociale ; ▸ l'appartenance à une minorité nationale ; ▸ la fortune ; ▸ la naissance ; ▸ l'âge ; ▸ l'identité de genre et son expression ; ▸ l'orientation sexuelle ; ▸ les caractéristiques sexuelles ; ▸ l'état de santé et le handicap ; ▸ la situation matrimoniale et parentale ; ▸ la situation au regard du droit des étrangers ; ▸ la situation au regard de l'emploi. Le respect de l'égalité peut aller au-delà de l'interdiction d'un traitement moins favorable, sans justification, fondé sur une ou plusieurs caractéristiques protégées.
	Exemples de risques potentiels liés à l'IA
	Les tendances discriminatoires peuvent être créées ou exacerbées par : ▸ un système dont l'objectif est intrinsèquement discriminatoire ; ▸ des ensembles de données qui ne sont pas suffisamment représentatifs ou qui encodent des inégalités historiques ; ▸ des systèmes algorithmiques qui ne tiennent pas suffisamment compte des schémas de biais[23] dans les ensembles de données ou qui génèrent des résultats biaisés ; ▸ des tests et des évaluations inadéquats liés aux biais ; ▸ un manque de représentation ou de prise en compte des perspectives de divers groupes au cours du développement de l'IA ou d'autres activités d'IA ; ▸ une formation insuffisante et d'autres mesures de gouvernance pour les développeurs ou les utilisateurs de systèmes d'IA en ce qui concerne la discrimination potentielle. Parallèlement, un système formé sur la base de données théoriquement totalement impartiales peut néanmoins avoir des effets injustes en aval pour des raisons liées à la manière dont il est déployé.

21. Les dispositions relatives à l'égalité et à la non-discrimination sont complémentaires à d'autres dispositions substantielles de divers instruments internationaux relatifs aux droits humains, et ces questions sont donc pertinentes pour tous les domaines de préoccupation potentiels figurant dans le présent tableau.
22. Étant donné que tous les êtres humains appartiennent à la même espèce, les théories fondées sur l'existence de différentes « races » sont rejetées. Toutefois, le terme « race » est utilisé afin de garantir que les personnes qui sont généralement et erronément perçues comme « appartenant à une autre race » ne sont pas exclues de la protection prévue.
23. Dans le contexte de l'intelligence artificielle, le terme « biais » désigne un résultat faussé d'une manière injuste. Le biais est susceptible de conduire à des résultats discriminatoires dès lors : a) qu'il est lié à des caractéristiques protégées ; et b) qu'une action est entreprise sur la base des résultats biaisés d'un système d'IA (que ce biais soit introduit dans le système par les données d'entraînement ou d'ajustement, par le modèle lui-même, ou encore par la manière dont les résultats du modèle sont interprétés par un être humain).

(5) Égalité et interdiction de la discrimination	**Obligations en matière de droits humains potentiellement pertinentes**
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** articles 2, 3, 23, 24 et 26 ; **Pacte international relatif aux droits économiques, sociaux et culturels :** articles 2.2 et 3 ; **Conventions de l'OIT n° 111** (sur la discrimination en matière d'emploi et de profession) et **n° 190** (sur la violence et le harcèlement dans le monde du travail) ; **Convention européenne des droits de l'homme :** article 14 (Interdiction de discrimination), et article 1 du Protocole n° 12 ; **Charte sociale européenne :** articles 1.2, 4.3, 19.4-5 et 20 ; **Charte de l'Union européenne :** diverses dispositions du chapitre III relatif à l'égalité (articles 20 à 26), articles 14 (Droit à l'éducation), 15 (Liberté professionnelle et droit de travailler) et 16 (Liberté d'entreprise) ; **Convention américaine relative aux droits de l'homme :** article 24 (Égalité devant la loi) ; et d'autres instruments juridiques dédiés, tels que : la **Convention internationale sur l'élimination de toutes les formes de discrimination raciale**, la **Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes**, la **Convention relative aux droits de l'enfant** et la **Convention relative aux droits des personnes handicapées**.
(6) Pensée, conscience, religion et croyance	**Description illustrative**
	Les dispositions relatives aux droits humains protègent la capacité d'une personne à avoir, pratiquer et exprimer ses convictions religieuses ou l'absence de celles-ci sans interférence ou discrimination. Elles comprennent des protections pour la capacité d'exercer un culte, de changer de religion ou de croyances, et d'observer des pratiques religieuses (telles que l'habillement, les restrictions alimentaires ou les fêtes religieuses) en public ou en privé. Elles protègent également contre la contrainte d'adopter ou de renoncer à une religion et garantissent l'égalité de traitement quelle que soit l'appartenance religieuse, qu'elle soit réelle ou supposée. La liberté de conscience protège la capacité d'avoir des convictions profondes sur le bien et le mal, qui ne sont pas nécessairement fondées sur des systèmes religieux, et d'agir en fonction. Il peut s'agir par exemple de l'athéisme, du végétarisme ou de l'objection de conscience au service militaire. Bien que le contenu de la liberté de pensée fasse l'objet d'un débat, elle pourrait inclure de ne pas avoir à révéler ses pensées, de ne pas avoir à être puni pour ses pensées et la protection contre l'altération de ses pensées de manière inadmissible.
	Exemples de risques potentiels liés à l'IA
	Les risques liés à l'IA pour les droits religieux peuvent découler d'algorithmes qui créent ou perpétuent des biais préjudiciables affectant les personnes sur la base de leurs croyances religieuses réelles ou perçues, par des violations de la vie privée, ou en facilitant le profilage religieux, la suppression de l'expression religieuse et la manipulation des croyances. Ces risques menacent la capacité des individus à pratiquer et à exprimer leur religion sans craindre la discrimination, la surveillance ou la coercition. De plus, les systèmes d'IA tels que les systèmes de reconnaissance émotionnelle et les systèmes qui perpétuent la désinformation pourraient interférer avec la liberté de pensée ou de conscience et avoir un effet négatif. L'utilisation de systèmes d'IA peut compromettre l'autonomie individuelle en matière de croyance en influençant subtilement les perceptions et les processus de pensée et en limitant potentiellement la liberté de former et d'exprimer des convictions religieuses ou spirituelles sans ingérence extérieure indue.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** articles 18, 19 et 20 ; **Convention européenne des droits de l'homme :** articles 9 (Liberté de pensée, de conscience et de religion), 10 (Liberté d'expression) et 14 (Interdiction de discrimination) ; **Charte de l'Union européenne :** articles 10 (Liberté de pensée, de conscience et religion) et 11 (Liberté d'expression et d'information), chapitre III sur l'égalité ; **Convention américaine relative aux droits de l'homme :** articles 12 (Liberté de conscience et de religion) et 13 (Liberté de pensée et d'expression) ; **Convention relative aux droits de l'enfant et ses protocoles facultatifs :** article 14.

(7) Opinions, expression et information	**Description illustrative**
	La protection des opinions et de la liberté d'expression s'applique à la conservation, à la recherche, à la transmission et à la réception d'informations et d'idées par le biais de tous les médias. Les formes d'expression peuvent être de nature politique, artistique, y compris la production de pièces de théâtre et de spectacles, personnelle ou commerciale, et peuvent inclure des reportages, des photographies et des formes de comportement, telles que des boycotts ou des campagnes, des vêtements, des symboles, etc. Ces protections peuvent aussi s'appliquer à certaines relations régies par la règle du droit privé (par exemple les relations de travail) et aux déclarations faites dans la correspondance privée ou lors de réunions à huis clos.
	Exemples de risques potentiels liés à l'IA
	Les risques liés à l'IA pour les opinions, la liberté d'expression et l'accès à l'information peuvent potentiellement inclure la surcensure, la partialité des systèmes algorithmiques utilisés pour la sélection des contenus, ainsi que la création et la diffusion d'informations inexactes. En outre, la surveillance facilitée par l'IA peut étouffer la liberté d'expression, tandis que l'utilisation d'algorithmes peut réduire l'accès à tous les points de vue. L'utilisation de systèmes d'IA peut porter atteinte à l'autonomie d'expression individuelle en influençant de manière subtile les opinions, en façonnant le discours et en limitant la capacité de formuler, d'articuler et de partager librement des idées sans manipulation ou contrainte occulte. En fonction de la manière dont une personne s'exprime publiquement, un système d'IA peut faire des déductions incorrectes, ce qui pourrait affecter les informations ou leur accès dont la personne pourrait autrement bénéficier (par exemple, une personne portant certains vêtements pourrait être catégorisée comme appartenant à un groupe auquel elle n'appartient pas).
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** articles 19 et 20 ; **Convention relative aux droits de l'enfant et ses protocoles facultatifs :** article 13 ; **Convention européenne des droits de l'homme :** articles 10 (Liberté d'expression) et 14 (Interdiction de discrimination) ; **Charte de l'Union européenne :** articles 10 (Liberté de pensée, de conscience et religion), 11 (Liberté d'expression et d'information) et 42 (Droit d'accès aux documents) ; **Convention américaine relative aux droits de l'homme :** articles 13 (Liberté de pensée et d'expression) et 14 (Droit de rectification ou de réponse).
(8) Réunion et association pacifiques	**Description illustrative**
	Les droits humains sont protégés en ce qui concerne les rassemblements non violents et les marches, y compris les réunions dans les lieux privés et publics. Il existe des protections pour les groupements volontaires visant un objectif commun et la possibilité de former ou d'être affilié à un groupe ou à une organisation poursuivant des objectifs particuliers. Les partis politiques, les associations de minorités et les associations religieuses, ainsi que les syndicats, sont des exemples marquants de ces associations.
	Exemples de risques potentiels liés à l'IA
	Les risques potentiels de l'IA dans ce domaine peuvent inclure la surveillance arbitraire ou discriminatoire des membres d'associations et des participants à des réunions, la prévision policière, la censure des réseaux sociaux, la désinformation et le harcèlement ciblé, tous ces éléments pouvant dissuader les individus d'exercer leurs droits de se réunir pacifiquement et de s'associer librement, et de participer aux affaires publiques.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques et ses protocoles facultatifs :** article 21 ; **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** article 15 ; **Convention européenne des droits de l'homme (STE n° 5) et ses protocoles additionnels :** article 11 (Liberté de réunion et d'association) ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 5, 6.1-2, 21, 28 et 29 ; **Charte de l'Union européenne :** articles 12 (Liberté de réunion et d'association) et 28 (Droits de négociation et d'actions collectives) ; **Convention américaine relative aux droits de l'homme :** articles 15 (Droit de réunion) et 16 (Liberté d'association).

(9) Propriété	**Description illustrative**
	Les protections des droits humains dans ce domaine ne se limitent généralement pas à la propriété de biens matériels, mais peuvent s'étendre à d'autres droits et intérêts, y compris les droits de propriété intellectuelle, qui constituent des actifs et peuvent également être considérés comme des « droits de propriété ».
	Exemples de risques potentiels liés à l'IA
	Les droits de propriété peuvent être défavorablement affectés par des systèmes d'IA dans certains domaines, notamment dans les cas suivants : l'utilisation d'algorithmes erronés dans les domaines de la finance et de l'immobilier pour prendre des décisions automatisées concernant la valeur des biens immobiliers, les prêts ou le crédit ; le contenu généré par l'IA, comme les œuvres d'art, la musique ou les textes, et le développement de modèles d'IA peuvent soulever des questions complexes liées aux droits de propriété intellectuelle ; sans contrôle humain, les progrès rapides de la technologie de l'IA pourraient dépasser les cadres juridiques existants, laissant des lacunes dans la protection des droits de propriété ; l'utilisation de la propriété intellectuelle pour la formation de modèles lorsque les lois ne sont pas claires ou appliquées de manière incohérente quant à la question de savoir si une telle utilisation est autorisée. Ces risques peuvent compromettre le contrôle des individus sur leurs biens et leurs créations intellectuelles, ce qui peut conduire à une exploitation et à une discrimination potentielles.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Convention européenne des droits de l'homme (STE n° 5) et ses protocoles additionnels :** article 1 du Protocole n° 1 (Protection de la propriété) ; **Charte de l'Union européenne :** article 17 (Droit de propriété) ; **Convention américaine relative aux droits de l'homme :** article 21 (Droit à la propriété privée).
(10) Éducation	**Description illustrative**
	La protection des droits humains dans ce domaine concerne l'accès des individus à une éducation de qualité sans discrimination. Les protections peuvent englober non seulement l'accès à l'enseignement primaire, mais aussi à l'enseignement secondaire, supérieur et professionnel, ainsi qu'aux possibilités d'apprentissage tout au long de la vie
	Exemples de risques potentiels liés à l'IA
	Les risques liés à l'IA comprennent les biais préjudiciables produits par les algorithmes dans des domaines tels que les admissions, la notation et l'apprentissage personnalisé. D'autres risques liés à l'IA (y compris ceux résultant d'un manque de culture numérique) peuvent nuire à l'efficacité et à l'équité des systèmes d'éducation, de formation professionnelle et de formation, en ayant un impact sur l'expérience d'apprentissage globale des étudiants ou sur leurs chances sur le marché du travail.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** article 13 ; **Convention relative aux droits de l'enfant et ses protocoles facultatifs :** article 28 ; **Convention européenne des droits de l'homme (STE n° 5) et ses protocoles additionnels :** article 2 du Protocole n° 1 (Droit à l'instruction) ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 1.4, 7.6, 9, 10, 15, 17, 19.11-12, 20.*b* et 30 ; **Charte de l'Union européenne :** article 14 (Droit à l'éducation) ; **Convention américaine relative aux droits de l'homme :** article 13 (Droit à l'éducation) du Protocole de San Salvador.

(11) Arts, sciences, culture et langues	**Description illustrative**
	Les protections des droits humains dans ce domaine permettent l'accès universel à une variété de ressources culturelles, la jouissance et les bénéfices du progrès scientifique et de ses applications, ainsi que la protection des intérêts moraux et matériels résultant de toute production scientifique, littéraire ou artistique des auteurs. Elles protègent également la capacité des individus, y compris les membres des minorités ethniques, religieuses ou linguistiques, à jouir de leur culture, à pratiquer leur religion et à utiliser leur langue.
	Exemples de risques potentiels liés à l'IA
	Les considérations liées à l'IA dans les domaines de la culture, des arts et des sciences peuvent impliquer des questions complexes liées à la propriété intellectuelle, à l'exactitude des données et au contrôle de la qualité, à des sources de données peu fiables et à des risques pour la pensée critique et les pratiques de recherche éthiques. Ces risques peuvent nuire à la richesse et à la diversité de l'expression humaine, à la créativité et au développement des connaissances. De même, ces protections comprennent la liberté de la recherche scientifique et de l'activité créatrice.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques et ses protocoles facultatifs :** article 27 ; **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** article 15 ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 15.3, 22.*c*, 23 et 30 ; **Charte de l'Union européenne :** articles 13 (Liberté des arts et des sciences) et 22 (Diversité culturelle, religieuse et linguistique) ; **Convention américaine relative aux droits de l'homme :** article 14 (Droit aux bienfaits de la culture) du Protocole de San Salvador.
(12) Travail et emploi	**Description illustrative**
	Les droits du travail protègent les travailleurs sur le lieu de travail en matière de traitement équitable, d'environnement de travail sûr et sain, de conditions de travail saines, sûres et décentes, de salaires raisonnables, de respect de la vie privée, de niveau de vie décent, d'absence de discrimination, de traitement juste et équitable, et de possibilité de former des syndicats et d'y adhérer. Les droits du travail peuvent inclure, par exemple, des normes minimales de salaire et un niveau de vie décent, des limites au temps de travail, la protection contre le travail forcé, les questions liées au travail des enfants, à la liberté d'association, à la reconnaissance effective du droit de négociation collective et à l'élimination de la discrimination.
	Exemples de risques potentiels liés à l'IA
	L'IA peut avoir divers effets transformateurs sur le travail et l'emploi. Les risques liés à l'IA comprennent l'utilisation de l'IA pour surveiller et gérer les employés, ce qui suscite des inquiétudes en ce qui concerne la protection de la vie privée et crée des environnements stressants en raison du suivi constant des performances, du comportement et de la productivité, ainsi que des déplacements d'emplois potentiellement liés à l'IA. Les systèmes d'IA utilisés pour informer ou prendre des décisions concernant l'embauche, le licenciement ou les promotions risquent d'être biaisés, opaques ou de constituer des pratiques de travail injustes, y compris ceux qui concernent les personnes travaillant dans le domaine de l'IA. L'IA peut être utilisée pour automatiser et optimiser les chaînes d'approvisionnement, ce qui pourrait involontairement augmenter la demande de main-d'œuvre bon marché, y compris le travail des enfants dans certains secteurs. L'utilisation de l'IA dans le suivi et la gestion de la main-d'œuvre peut réduire la surveillance humaine, ce qui pourrait permettre à des pratiques d'exploitation de passer inaperçues, en particulier en ce qui concerne le travail des enfants.

(12) Travail et emploi	**Obligations en matière de droits humains potentiellement pertinentes** Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** articles 6, 7, 8 et 10 ; **Convention relative aux droits de l'enfant et ses protocoles facultatifs :** article 32 ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 1-10, 21, 22-26, 28 et 29 ; **Charte de l'Union européenne :** articles 8 (Protection des données à caractère personnel), 12 (Liberté de réunion et d'association), 15 (Liberté professionnelle et droit de travailler), 27 (Droit à l'information et à la consultation des travailleurs au sein de l'entreprise), 28 (Droit de négociation et d'actions collectives), 29 (Droit d'accès aux services de placement), 30 (Protection en cas de licenciement injustifié), 31 (Conditions de travail justes et équitables), 32 (Interdiction du travail des enfants et protection des jeunes au travail) et 33 (Vie familiale et vie professionnelle) ; **Convention américaine relative aux droits de l'homme :** articles 7 (Conditions de travail justes, équitables et satisfaisantes) et 8 (Droits syndicaux) du Protocole de San Salvador. En outre, les conventions de l'OIT suivantes peuvent être utiles dans ce contexte : la Convention n° 87 sur la liberté syndicale et la protection du droit syndical (1948) ; la Convention n° 98 sur le droit d'organisation et de négociation collective (1949) ; la Convention n° 111 concernant la discrimination (emploi et profession) (1958) ; la Convention n° 138 sur l"âge minimum (1973) ; la Convention n° 155 sur la santé et la sécurité des travailleurs (1981) ; et la Convention n° 190 sur la violence et le harcèlement (2019).
(13) Santé, soins de santé et aide sociale/protection sociale	**Description illustrative** Le secteur de la santé englobe l'accès aux services médicaux, les initiatives de santé publique et les infrastructures de soins de santé afin de garantir le bien-être physique et mental des individus. Les protections dans ce domaine sont généralement liées à la capacité des individus à accéder aux soins de santé afin de pouvoir jouir du meilleur état de santé physique et mentale possible. Le secteur de la sécurité sociale permet aux personnes en situation de chômage, d'invalidité, de vieillesse ou de difficulté économique de bénéficier d'un soutien financier et d'une protection par le biais de prestations et de programmes sociaux. Les protections dans ce domaine concernent généralement la capacité des individus à recevoir une telle assistance. La protection sociale, l'assistance sociale et la protection contre la pauvreté et l'exclusion sociale visent à exclure ou à réduire la pauvreté et l'inégalité, et à promouvoir la dignité humaine. **Exemples de risques potentiels liés à l'IA** L'utilisation de systèmes d'IA peut donner lieu à des données biaisées préjudiciables menant à un accès inégal aux soins de santé ou aux prestations sociales, en particulier pour les groupes marginalisés. L'utilisation de l'IA dans le traitement de données personnelles sensibles en matière de santé et de prestations sociales expose à des risques significatifs pour la vie privée et la sécurité (y compris les questions relatives au consentement et à l'autonomie individuelle), notamment en cas de violation ou d'utilisation abusive des données. Une dépendance excessive à l'égard de l'IA pour les diagnostics médicaux ou les décisions en matière de prestations sociales pourrait entraîner des erreurs, réduire le contrôle humain et nuire aux résultats pour les patients, ou rejeter à tort des demandes de prestations. En ce qui concerne la protection sociale et la sécurité sociale, l'utilisation de l'IA peut présenter des risques lorsqu'elle est utilisée pour déterminer l'éligibilité des personnes aux prestations sociales ou pour détecter les cas où les prestations sociales ont été versées à tort et où l'autorité a des demandes de remboursement : les conséquences des erreurs (éventuellement causées par des préjugés) peuvent être graves et touchent souvent les personnes en situation de vulnérabilité en particulier. **Obligations en matière de droits humains potentiellement pertinentes** Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** articles 9, 10, 11 et 12 ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 2, 3, 7, 8, 11, 12, 13, 14, 15, 17.1, 19.2, 22.*b*, 23, 27.1.*b* et 30 ; **Code européen de sécurité sociale (CESS) :** articles 7-12 (Soins médicaux), 13-18 (Indemnités de maladie), 19-24 (Prestations de chômage), 25-30 (Prestations de vieillesse), 31-38 (Prestations en cas d'accidents du travail et de maladies professionnelles), 46-52 (Prestations de maternité), 53-58 (Prestations d'invalidité) et 59-64 (Prestations de survivants) ; **Convention relative aux droits de l'enfant et ses protocoles facultatifs :** article 24 ; **Charte de l'Union européenne :** articles 8 (Protection des données à caractère personnel), 34 (Sécurité sociale et aide sociale) et 35 (Protection de la santé) ; **Convention américaine relative aux droits de l'homme :** articles 9 (Droit à la sécurité sociale) et 10 (Droit à la santé) du Protocole de San Salvador.

(14) Enfants	**Description illustrative**
	Ce groupe bénéficie à la fois de protections générales en matière de droits humains et de protections spécifiques aux enfants. Elles visent à garantir que les enfants peuvent grandir, apprendre, jouer, se développer et s'épanouir dans la dignité et que les vulnérabilités et les besoins spécifiques des enfants sont dûment pris en compte. Dans toutes les actions concernant les enfants, leur intérêt supérieur doit constituer une considération primordiale. En gardant à l'esprit les considérations ci-dessus, de nombreux domaines de préoccupation potentiels présentés dans ce tableau seront pertinents pour les enfants.
	Exemples de risques potentiels liés à l'IA
	Les systèmes d'IA pourraient présenter des risques pour les enfants s'ils sont utilisés pour prendre des décisions qui ne tiennent pas compte de leurs besoins ou de leurs droits, ou s'ils créent des images ou vidéos fausses mais réalistes pouvant troubler les enfants, les induire en erreur ou leur nuire. Les contenus générés, sélectionnés ou recommandés par des systèmes d'IA (par exemple sur les réseaux sociaux ou les plateformes vidéo) peuvent exposer les enfants à des contenus illégaux, nuisibles ou inadaptés à leur âge. La publicité ciblée par l'IA et les contenus fondés sur l'IA peuvent exploiter la vulnérabilité des enfants, influençant leur comportement et leurs préférences, y inclus à des fins commerciales. Les plateformes sociales pilotées par l'IA peuvent contribuer à l'anxiété, à la dépression et à une faible estime de soi chez les enfants, à travers le cyberharcèlement, l'addiction et les comparaisons nuisibles. Les enfants sont particulièrement sensibles aux dommages liés à l'IA générative. Les enfants sont moins capables de discerner un contenu synthétique d'un contenu authentique, d'identifier des informations inexactes et de savoir s'ils sont manipulés par des schémas obscurs, et de comprendre qu'ils interagissent avec une machine plutôt qu'avec un être humain. L'IA est de plus en plus utilisée pour produire du matériel pédopornographique et exploiter les enfants. L'IA générative peut permettre la création de matériel pédopornographique synthétique et nuire aux enfants, à leurs familles et à la société. Les prédateurs peuvent utiliser également l'IA générative pour préparer, extorquer et exploiter leurs victimes.
	Obligations en matière de droits humains potentiellement pertinentes
	Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** articles 10.3 et 12 ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 7 et 17 ; **CESS :** articles 39-45 (Prestations aux familles) ; **Charte de l'Union européenne :** articles 7 (Respect de la vie privée et familiale), 8 (Protection des données à caractère personnel), 24 (Droits de l'enfant) et 32 (Interdiction du travail des enfants et protection des jeunes au travail) ; **Convention américaine relative aux droits de l'homme :** article 19 (Droit de l'enfant). En plus des dispositions relatives aux droits humains contenues dans les traités internationaux sur les droits humains, les instruments spécialisés en matière de droits humains suivants peuvent être utiles : la Convention de 1989 des Nations Unies relative aux droits de l'enfant et ses protocoles facultatifs ; l'Observation générale n° 25 sur les droits de l'enfant en relation avec l'environnement numérique ; la Convention de 2007 du Conseil de l'Europe sur la protection des enfants contre l'exploitation et les abus sexuels (STCE n° 201, Convention de Lanzarote).
(15) Environnement	**Description illustrative**
	Dans certaines juridictions, les dispositions relatives aux droits humains peuvent reconnaître un certain niveau de responsabilité des autorités publiques dans la protection des individus contre les atteintes à l'environnement.
	Exemples de risques potentiels liés à l'IA
	Les systèmes d'IA, en particulier ceux qui utilisent d'importantes ressources informatiques pour l'apprentissage et l'inférence, consomment de grandes quantités d'énergie. Cela contribue aux émissions de carbone, surtout lorsqu'ils sont alimentés par des sources d'énergie non renouvelables. De même, la production du matériel nécessaire à la formation et au fonctionnement des systèmes d'IA nécessite de nombreux matériaux à base de terres rares, dont l'extraction a des effets polluants plus importants. Les applications de l'IA dans les domaines de l'agriculture, de l'économie, de la gestion des ressources ou de la surveillance de la faune et de la flore, tout en offrant de nombreux avantages potentiels pour l'environnement, peuvent également avoir des incidences environnementales imprévues si elles ne sont pas gérées avec soin, telles que la perturbation des écosystèmes, la surexploitation des ressources naturelles, la pollution ou l'atteinte à la biodiversité.

(15) Environnement	**Obligations en matière de droits humains potentiellement pertinentes** Si l'on considère la question sous l'angle des droits humains, les problèmes liés à la protection de l'environnement pourraient être abordés en vertu des instruments suivants : **Pacte international relatif aux droits économiques, sociaux et culturels et son protocole facultatif :** article 12 ; **Convention européenne des droits de l'homme (STE n° 5) et ses protocoles additionnels :** articles 2 (Droit à la vie), 3 (Interdiction de la torture), 6 (Droit à un procès équitable), 8 (Droit au respect de la vie privée et familiale), 10 (Liberté d'expression) et 11 (Liberté de réunion et d'association) ; **Charte de l'Union européenne :** article 37 (Protection de l'environnement) ; **Convention américaine relative aux droits de l'homme :** article 11 (Droit à un environnement salubre) du Protocole de San Salvador ; **Convention de la Commission économique pour l'Europe des Nations Unies sur l'accès à l'information, la participation du public au processus décisionnel et l'accès à la justice en matière d'environnement**.
(16) Démocratie	**Description illustrative** Les dispositions relatives aux droits humains protègent la capacité des individus à participer à la gestion des affaires publiques, à voter lors d'élections libres et équitables, à se présenter à des fonctions publiques et à avoir accès aux services gouvernementaux. Une caractéristique importante des systèmes démocratiques de gouvernement est le pluralisme politique, qui est en grande partie garanti par la protection des droits humains, dont le respect est essentiel pour une démocratie florissante (comme la liberté d'expression, la liberté d'association et la liberté de réunion pacifique), ainsi que par l'existence de médias pluralistes et indépendants, d'une gamme de partis politiques représentant différents intérêts et points de vue, d'un accès équitable et d'une participation significative au débat public et à la prise de décisions publiques, et d'un accès à des informations exactes et fiables. Les droits relatifs à une participation significative aux processus de décision publique et au gouvernement garantissent que les individus ont la capacité de s'engager dans le processus politique, notamment en votant, en se présentant à des élections et en exprimant leurs opinions sur les politiques publiques. Cela favorise la gouvernance démocratique, la responsabilité et la représentation, et permet aux citoyens d'influencer les décisions qui affectent leur vie et leurs communautés. **Exemples de risques potentiels liés à l'IA** Les systèmes d'IA jouant un rôle dans l'influence ou l'information des processus démocratiques pourraient avoir un impact négatif sur l'accès équitable des individus au débat public et à la participation aux élections libres et équitables, ainsi que sur leur capacité à former librement leur opinion, à travers les éléments suivants, qui ne sont pas exhaustifs : la tromperie, la désinformation ou la mésinformation à l'échelle locale, nationale ou mondiale causées par le déploiement du système ; la manipulation, à l'échelle locale, nationale ou mondiale, facilitée par le déploiement du système ; l'intimidation ou le contrôle comportemental, à l'échelle locale, nationale ou mondiale, facilités par le déploiement du système. L'intelligence artificielle pourrait être utilisée d'une manière qui entrave la participation à la vie publique, par exemple par l'utilisation d'algorithmes pour filtrer les candidats politiques ou les participants à des processus publics, ou encore pour restreindre la liberté d'expression sur des questions liées aux politiques publiques. **Obligations en matière de droits humains potentiellement pertinentes** Les instruments juridiques suivants peuvent être pertinents : **Pacte international relatif aux droits civils et politiques :** articles 19, 20, 21 et 25 ; **Convention européenne des droits de l'homme (STE n° 5) et ses protocoles additionnels :** articles 10 (Liberté d'expression) et 11 (Liberté de réunion et d'association), et article 3 du Protocole n° 1 (Droit à des élections libres) ; **Charte sociale européenne (STE n° 35) et ses protocoles, et Charte sociale européenne révisée (STE n° 163) :** articles 5 et 6.1-2 ; **Charte de l'Union européenne :** articles 11 (Liberté d'expression et d'information), 12 (Liberté de réunion et d'association), 39 (Droit de vote et d'éligibilité aux élections au Parlement européen), 40 (Droit de vote et d'éligibilité aux élections municipales) et 42 (Droit d'accès aux documents) ; **Convention américaine relative aux droits de l'homme :** articles 13 (Liberté de pensée et d'expression), 14 (Droit de rectification ou de réponse), 15 (Droit de réunion), 16 (Liberté d'association) et 23 (Droits politiques).

(17) État de droit	
	Description illustrative
	L'État de droit est un principe de gouvernance selon lequel toutes les personnes, institutions et entités, publiques et privées, y compris l'État lui-même, sont responsables devant des lois qui sont publiquement promulguées, appliquées de manière égale, et jugées de manière indépendante, et qui sont conformes au respect des droits humains internationaux. L'indépendance judiciaire est fondamentale pour la démocratie et pour garantir la confiance du public dans l'administration de la justice. Les institutions de l'État de droit sont des organisations (la justice, le législatif et l'exécutif, la profession juridique et des organismes semi-indépendants tels que les autorités anticorruptions, les autorités de protection des données, les médiateurs, etc.) et des systèmes (système juridique) qui garantissent que les lois sont appliquées de manière équitable, cohérente et transparente, protégeant les droits des individus et maintenant l'ordre dans la société. Ensemble, ces institutions préservent l'État de droit en veillant à ce que ce soient les lois qui régissent la société et non des décisions arbitraires, et à ce que chacun soit traité de manière égale devant la loi.
	Exemples de risques potentiels liés à l'IA
	L'utilisation de l'IA présente des risques (notamment en raison du manque de transparence et de supervision humaine suffisante, de l'utilisation de systèmes entraînant des conséquences imprévues, des biais nuisibles, des menaces pour la cybersécurité, entre autres) pour l'État de droit, l'équité procédurale, la transparence et la responsabilité dans des contextes tels que : l'intégrité des institutions et des processus démocratiques, y compris le principe de la séparation des pouvoirs, le respect de l'indépendance judiciaire ; l'accès à la justice ; les mécanismes de responsabilité (supervision de l'exécutif) ainsi que les organismes et politiques de lutte contre la corruption.
	Obligations en matière de droits humains potentiellement pertinentes
	L'exercice effectif et la protection des droits inscrits dans les instruments juridiques suivants sont essentiels pour garantir le respect de l'État de droit. **Pacte international relatif aux droits civils et politiques :** articles 2.3, 4 et 14 ; **Convention européenne des droits de l'homme :** articles 6 (Droit à un procès équitable) et 13 (Droit à un recours effectif) ; **Charte de l'Union européenne :** articles 41 (Droit à une bonne administration), 42 (Droit d'accès aux documents), 47 (Droit à un recours effectif et à accéder à un tribunal impartial) et 48 (Présomption d'innocence et droits de la défense) ; **Convention américaine relative aux droits de l'homme :** articles 8 (droit à un procès équitable), 10 (droit à une indemnisation) et 25 (droit à la protection judiciaire).

Ressources COBRA F (Liste illustrative[24] des secteurs/domaines et domaines possibles de préoccupation potentiels)

Cette ressource énumère (sans ordre de priorité) les secteurs et domaines de préoccupation potentiels du point de vue des droits humains, de la démocratie et de l'État de droit. Elle est destinée à être utilisée conjointement avec les **ressources COBRA E**, afin de réaliser la cartographie des impacts dans le cadre de la COBRA. De même, cette liste est non exhaustive et devrait être révisée régulièrement.

	Domaines	Domaines de préoccupation
1. Administration publique	**(a) Soins de santé**, notamment l'accès aux services de santé, aux diagnostics, aux pronostics et aux soins préventifs, l'administration de traitements de maintien en vie, la prise en charge des pathologies engageant le pronostic vital, les services de soins d'urgence, les services de consultation et de traitement en santé mentale, les décisions relatives à la fin de vie	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (4) Vie privée et protections de données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (13) Santé, soins de santé et aide sociale/protection sociale ; (14) Enfants ; (15) Environnement
	(b) Vie familiale et protection sociale, notamment la jouissance mutuelle des parents avec enfants, la garde, le droit de visite, les droits contractuels, la prise en charge par l'assistance publique, les familles d'accueil, l'adoption et les services de reproduction, l'accès aux services publics et la fourniture de prestations publiques	(1) Intégrité physique, psychique et dignité humaine ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (9) Propriété ; (13) Santé, soins de santé et aide sociale/protection sociale ; (14) Enfants
	(c) Migration et contrôle des frontières, notamment l'expulsion, l'extradition, l'éloignement, les changements de statut, le refus du droit d'entrée, la notification des droits, les services de traduction/d'interprétation, la production de transcriptions, le recueil et l'appréciation des éléments de preuve, les conditions et les modalités d'entrée sur le territoire de l'État et d'éloignement de ce dernier	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (12) Travail et emploi ; (13) Santé, soins de santé et aide sociale/protection sociale ; (14) Enfants ; (15) Environnement ; (17) État de droit
	(d) Développement et entretien des infrastructures, notamment la sécurité sanitaire et la jouissance de l'espace public, les transports publics et la mobilité, la gestion des risques environnementaux, les plans fonciers et l'aménagement du territoire et l'urbanisme, le logement, l'infrastructure digitale, la gestion de l'énergie et la consommation d'énergie	(1) Intégrité physique, psychique et dignité humaine ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (9) Propriété ; (13) Santé, soins de santé et aide sociale/protection sociale ; (14) Enfants ; (15) Environnement
	(e) Services d'urgence, notamment la gestion des opérations de sauvetage, les infrastructures de communication d'urgence et la gestion des suites des catastrophes	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (14) Enfants ; (15) Environnement
	(f) Éducation publique, notamment l'accès aux établissements d'enseignement, aux examens de l'éducation et à la reconnaissance officielle des études	(5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (14) Enfants
	(g) Emploi, notamment le recrutement, l'accès à l'emploi, la gestion des performances, la réglementation sur les travailleurs, et l'accessibilité et les aménagements raisonnables pour les personnes handicapées	(1) Intégrité physique, psychique et dignité humaine (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (8) Réunion et association pacifiques ; (12) Travail et emploi ; (13) Santé, soins de santé et aide sociale/protection sociale

24. Les références aux instruments internationaux relatifs aux droits humains dans ce tableau sont fournies à titre illustratif. Ces références ne s'appliquent qu'aux États qui sont parties à ces instruments. Chaque État est censé appliquer ses propres lois applicables conformément à ses obligations juridiques internationales, ce qui pourrait inclure l'encouragement du secteur privé à respecter et à soutenir les droits humains, y compris tels qu'énoncés dans les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme.

	Domaines	Domaines de préoccupation
2. Maintien de l'ordre et sécurité	**(a) Police et services assimilés**, notamment l'utilisation de la force létale, le recours à la force physique lors des arrestations, les contrôles d'identité et l'identification des personnes à des fins d'application de la loi, les programmes concernant la protection des personnes en danger (par exemple les victimes de violence domestique ou les témoins protégés), les arrestations et les détentions, la gestion de programmes concernant le contrôle des fonctionnaires, la gestion des opérations de sauvetage et de libération d'otages, la gestion des foules lors d'événements publics, la prévision policière, l'analyse des émotions et des sentiments, la surveillance et les restrictions par la police et d'autres organismes chargés de l'application de la loi	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (7) Opinions, expression et information ; (8) Réunion et association pacifiques ; (9) Propriété ; (14) Enfants ; (15) Environnement ; (17) État de droit
	(b) Poursuites, notamment le recueil et l'appréciation des éléments de preuve	(1) Intégrité physique, psychique et dignité humaine ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (7) Opinions, expression et information ; (14) Enfants ; (15) Environnement ; (17) État de droit
3. Administration de la justice	**a) Tribunaux et justice**, notamment les arrestations, les détentions, les décisions concernant la libération sous caution, la libération conditionnelle et le port de bracelet électronique, la notification des droits et des décisions, les services de traduction/d'interprétation, la production de transcriptions, le recueil et l'appréciation des éléments de preuve (y compris l'appréciation de la fiabilité des témoins), l'octroi de l'assistance d'un défenseur, le prononcé de toute accusation pénale, la détermination des droits et obligations civils, les décisions concernant la récusation des juges ou des jurés, les décisions concernant l'accès au niveau chargé de la révision des procédures, la détermination de la sanction pénale, les procédures automatisées	(2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (7) Opinions, expression et information ; (9) Propriété ; (14) Enfants ; (17) État de droit
	(b) Aspects institutionnels de l'organisation du système judiciaire, notamment le respect de l'indépendance judiciaire, y compris, mais sans s'y limiter, la gestion des processus de contrôle des antécédents, de nomination et de révocation des juges/officiers de justice, l'attribution des affaires à traiter à certains juges/officiers de justice, la gestion des affaires dans le cadre des procédures judiciaires	(3) Justice et administration de la justice ; (5) Égalité et interdiction de la discrimination ; (9) Propriété ; (12) Travail et emploi ; (17) État de droit

	Domaines	Domaines de préoccupation
4. Processus démocratiques	**(a) Système électoral**, notamment les conditions et modalités d'exercice du droit de vote, l'âge d'éligibilité, les règles d'exclusion, les conditions et modalités du vote, les méthodes et procédures de vote, les conditions et modalités du dépouillement, le droit de se présenter à des élections, l'organisation d'élections et de référendums, le redécoupage électoral, la gestion des litiges électoraux et les recours effectifs à cet égard, la diffusion de l'information électorale	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (7) Opinions, expression et information ; (8) Réunion et association pacifiques ; (16) Démocratie
	(b) Institutions et processus politiques, notamment la suprématie de la constitution, le rôle du pouvoir judiciaire dans l'équilibre des pouvoirs, l'indépendance judiciaire, la délégation de la fonction législative	(3) Justice et administration de la justice ; (5) Égalité et interdiction de la discrimination ; (7) Opinions, expression et information ; (16) Démocratie ; (17) État de droit
	(c) Opinions et discours public, notamment, mais sans s'y limiter, l'expression d'un discours protégé sous diverses formes, la protection des sources journalistiques, les activités de collecte d'informations, l'accès aux données, le recueil et le traitement automatique des données et les activités de propriété, de recherche et d'enquête, le régime de divulgation concernant les informations obtenues à titre confidentiel, la protection des lanceurs d'alerte, la démocratie participative et les consultations publiques, y compris les questions relatives à l'organisation des réunions de comités	(3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (14) Enfants ; (16) Démocratie ; (17) État de droit
	(d) Réunion et association pacifiques, notamment le moment, le lieu et le mode de conduite des réunions, les conditions et modalités de formation d'un groupe ou d'une organisation poursuivant des objectifs particuliers, ou d'affiliation à ces derniers, la surveillance des réunions et l'identification des participants, la participation des citoyens à la vie publique	(5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (8) Réunion et association pacifiques ; (9) Propriété ; (11) Arts, sciences, culture et langues ; (14) Enfants ; (16) Démocratie
	(e) Accès à l'information, notamment l'accès aux informations personnelles, aux informations financières et aux informations sur les transactions commerciales des particuliers, l'obligation de fournir des informations fiables et précises, les responsabilités en matière de vérification et de transmission des informations, l'accès aux informations détenues par l'État	(3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (9) Propriété ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (14) Enfants ; (16) Démocratie ; (17) État de droit
	(f) Médias, notamment la transparence concernant la propriété des médias, le pluralisme des médias, la liberté d'expression pendant les élections (hors ligne et en ligne), les devoirs et responsabilités des portails d'information sur internet, la production automatisée de bulletins d'information, les plateformes médiatiques, la désinformation et la mésinformation, la modération des contenus en ligne	(5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (8) Réunion et association pacifiques ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (14) Enfants ; (16) Démocratie

	Domaines	Domaines de préoccupation
5. Prison et probation	**(a) Gestion des prisons et des centres de détention**, notamment le profilage des détenus, le dépistage psychologique des détenus potentiellement vulnérables, la gestion des détenus dangereux, la gestion de la population carcérale, les fouilles des visiteurs et des détenus, la surveillance des communications	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (13) Santé, soins de santé et aide sociale/protection sociale ; (15) Environnement
	(b) Libération conditionnelle et services de probation, notamment la libération conditionnelle, et la surveillance des personnes et de tout dispositif électronique portable	(1) Intégrité physique, psychique et morale et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination
6. Services essentiels proposés par le secteur privé	(a) Services de communication	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (9) Propriété ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (14) Enfants
	(b) Éducation et formation professionnelle	(5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (7) Opinions, expression et information ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (12) Travail et emploi ; (14) Enfants
	(c) Applications biomédicales, sciences de la vie, épidémiologie et soins de santé	(1) Intégrité physique, psychique et morale et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (9) Propriété ; (13) Santé, soins de santé et aide sociale/protection sociale ; (14) Enfants ; (15) Environnement
	(d) Gestion de l'environnement et des déchets	(1) Intégrité physique, psychique et morale et dignité humaine ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (13) Santé, soins de santé et aide sociale/protection sociale ; (15) Environnement
	(e) Gestion de l'énergie	(5) Égalité et interdiction de la discrimination
	(f) Infrastructures urbaines et urbanisme	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (9) Propriété ; (14) Enfants ; (15) Environnement
	(g) Industrie manufacturière et automatisation industrielle	(1) Intégrité physique, psychique et dignité humaine ; (12) Travail et emploi ; (15) Environnement
	(h) Construction et bâtiment	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (12) Travail et emploi ; (15) Environnement
	(i) Sécurité et sûreté publique	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (8) Réunion et association pacifiques ; (15) Environnement
	(g) Industrie manufacturière et automatisation industrielle	(1) Intégrité physique, psychique et dignité humaine ; (12) Travail et emploi ; (15) Environnement

6. Services essentiels proposés par le secteur privé	(h) Construction et bâtiment	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (12) Travail et emploi ; (15) Environnement
	(i) Sécurité et sûreté publique	(1) Intégrité physique, psychique et dignité humaine ; (2) Liberté physique et sûreté, mouvement et résidence ; (3) Justice et administration de la justice ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (8) Réunion et association pacifiques ; (15) Environnement
	(j) Domotique (technologies domestiques intelligentes)	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (14) Enfants
	(k) Logement et offre d'hébergement social	(1) Intégrité physique, psychique et morale et dignité humaine ; (5) Égalité et interdiction de la discrimination ; (13) Santé, soins de santé et aide sociale/protection sociale
	(l) Emploi, ressources humaines et gestion du travail (notamment le recrutement, l'accès à l'emploi, la gestion des performances, la réglementation sur les travailleurs, l'accessibilité, et les adaptations raisonnables pour les personnes affectées d'un handicap	(2) Liberté physique et sûreté, mouvement et résidence ; (4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (10) Éducation ; (12) Travail et emploi
	(m) Services financiers	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (9) Propriété
	(n) Technologies de l'information et réseaux	(4) Vie privée et protection des données ; (5) Égalité et interdiction de la discrimination ; (7) Opinions, expression et information ; (9) Propriété ; (10) Éducation ; (11) Arts, sciences, culture et langues ; (14) Enfants
	(o) Industrie automobile et infrastructures de transport	(2) Liberté physique et sûreté, mouvement et résidence ; (5) Égalité et interdiction de la discrimination ; (15) Environnement
	(p) Agriculture et ressources alimentaires	(1) Intégrité physique et dignité humaine, (5) Égalité et interdiction de la discrimination ; (6) Pensée, conscience, religion et croyance ; (12) Travail et emploi ; (13) Santé, soins de santé et aide sociale/protection sociale ; (15) Environnement