INTELLIGENCE ARTIFICIELLE, DROITS DE L'HOMME, DÉMOCRATIE ET ÉTAT DE DROIT

GUIDE INTRODUCTIF

David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, & Morgan Briggs

Avec un avant - propos de

Lord Tim Clement-Jones

Préparé pour accompagner *l'étude de faisabilité* publiée par le Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe.

The Alan Turing Institute

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

INTELLIGENCE ARTIFICIELLE, DROITS DE L'HOMME, DÉMOCRATIE ET ÉTAT DE DROIT

GUIDE INTRODUCTIF

Préparé pour accompagner *l'étude de faisabilité* publiée par le Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe.

Auteurs :
David Leslie, Christopher Burr,
Mhairi Aitken, Josh Cowls,
Mike Katell & Morgan Briggs

Le Programme de politiques publiques de l'Institut Alan Turing a été créé en mai 2018 dans le but de développer la recherche et de concevoir des outils et des techniques pour aider les États à innover en matière de technologies à grand volume de données et à améliorer la qualité de vie des citoyens. L'Institut travaille avec des responsables politiques pour examiner comment la science des données et l'intelligence artificielle peuvent éclairer les politiques gouvernementales et améliorer la fourniture des services publics. Nous sommes convaincus que les États ne peuvent récolter les fruits de ces technologies qu'en plaçant l'éthique et la sécurité au premier rang des priorités.

Le présent Guide introductif est un document vivant, qui évoluera et s'améliorera avec les contributions des utilisateurs, des acteurs concernés et des parties intéressées. Nous avons donc besoin de votre participation ! Vous pouvez envoyer vos commentaires à l'adresse policy@turing.ac.uk. Ce travail de recherche a été financé en partie par une subvention de l'ESRC (Conseil de la recherche économique et sociale du Royaume-Uni) (ES/T007354/1) et par des fonds publics, qui ont rendu possible le Programme de politiques publiques de l'Institut Alan Turing (https://www.turing.ac.uk/research/research-programmes/public-policy).

Cet ouvrage a été élaboré pour une large part à partir de l'Étude de faisabilité publiée en décembre 2020 par le Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe. Il est recommandé aux lecteurs de se référer directement à ce document (https://rm.coe.int/0900001680a1160f) pour se faire une idée plus complète des vues exprimées dans le présent guide.

Les vues exprimées dans ce guide sont de la responsabilité de leurs auteurs et ne reflètent pas nécessairement la ligne officielle du Conseil de l'Europe ou de Alan Turing Institute.

Pour citer cet ouvrage :
D. Leslie, C. Burr, M. Aitken, J. Cowls, M. Katell et M. Briggs, Intelligence artificielle, droits de l'homme, démocratie et État de droit. Guide introductif, Conseil de l'Europe, 2021.

Mise en page : The Alan Turing Institute

Images : Shutterstock et Canva Pro

TABLE DES MATIÈRES

AVANT-PROPOS

Il est manifeste, aujourd'hui plus que jamais, en particulier après cette année de pandémie de covid-19 qui a exposé au grand jour notre dépendance toujours plus grande aux technologies numériques, que nous devons préserver la confiance des citoyens dans l'adoption de l'intelligence artificielle (IA).

Pour y parvenir, nous devons, tout en concrétisant les possibilités offertes par l'IA, atténuer les risques liés à la mise en œuvre de cette technologie. D'où la nécessité d'une norme claire en matière de responsabilité et de comportement éthique.

Si 2019 a été marquée par l'adoption, au niveau des pays, de principes éthiques de l'IA arrêtés à l'échelle internationale, notamment ceux de la Recommandation de l'OCDE sur l'IA et les principes non contraignants du G20 sur l'IA, 2020 a été l'année où la communauté internationale de l'IA a entrepris une réflexion sur la manière d'instiller ces principes dans le développement et le déploiement des systèmes d'IA.

Pour que l'IA éthique devienne une réalité, il faut évaluer les risques de cette technologie en contexte, s'agissant en particulier de son incidence sur les droits civils et sociaux, puis, en fonction des risques évalués, élaborer des normes ou édicter des règles pour la conception, le développement et le déploiement éthiques des systèmes d'IA.

Une initiative majeure de ce processus a été l'*Étude de faisabilité* élaborée et adoptée en décembre par le Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe (CAHAI), qui examine les solutions possibles d'une réponse juridique internationale fondée sur les normes du Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie et de l'État de droit.

La question essentielle est de savoir s'il est possible et souhaitable d'apporter des réponses aux risques et opportunités spécifiques des systèmes d'IA en recourant à des instruments juridiques internationaux contraignants ou non contraignants, et ce par l'intermédiaire du Conseil de l'Europe, qui est le gardien de la Convention européenne des droits de l'homme, de la Convention 108+ qui protège le traitement des données à caractère personnel, et de la Charte sociale européenne.

Maintenant que le Conseil de l'Europe et le CAHAI entrent dans la phase de consultation multipartite dans le cadre de l'*Étude de faisabilité*, il est impératif, si l'on veut réaliser le plein potentiel de ce travail et faire les bons choix, s'agissant notamment des instruments juridiques et des mécanismes de contrôle et de conformité, de comprendre parfaitement les incidences sociétales et réglementaires de la démarche adoptée dans cette étude et des propositions fondées sur des principes qui y sont formulées.

Ce superbe Guide introductif, élaboré par l'Institut Alan Turing pour accompagner l'*Étude de faisabilité* et conçu pour en expliquer le contexte et faciliter la consultation, est un modèle de clarté. Nul doute qu'il renforcera la participation du public et permettra la tenue d'un débat large et éclairé. Il porte sur un domaine essentiel de l'action publique, où la participation du plus grand nombre à des discussions vastes et approfondies, en particulier sur les valeurs à adopter, est primordiale. Ce guide permettra de ne pas laisser ces questions à la seule décision de quelques spécialistes.

Lord Tim Clement-Jones
Londres, 2021

01 INTRODUCTION

FINALITÉ DU GUIDE

Il est frappant de constater que les progrès rapides accomplis ces vingt dernières années dans le domaine de l'intelligence artificielle (IA) et des technologies fondées sur les données placent la société contemporaine à un moment charnière où se décide quelle forme prendra le futur de l'humanité. D'un côté, la multiplication des innovations de l'IA bénéfiques pour la société promet de nous aider à lutter contre le changement climatique et la perte de la biodiversité, d'améliorer équitablement les soins médicaux, la qualité de vie, les transports, la production agricole, etc., et de remédier à bon nombre d'injustices sociales et d'inégalités matérielles qui assaillent le monde aujourd'hui. De l'autre, les innovations irresponsables de l'IA qui prolifèrent sont les signes avant-coureurs des problèmes éventuels qui nous attendent si ces technologies poursuivent sur leur lancée inquiétante.

Ces signes avant-coureurs, nous les voyons notamment dans les risques croissants que font peser de plus en plus les infrastructures de surveillance numérique (reconnaissance faciale en direct, etc.) sur le droit au respect de la vie privée, le droit à l'auto-expression, le droit d'association et le droit au consentement et sur d'autres libertés publiques et sociales. Ces signes, nous les voyons aussi dans les effets transformateurs qui apparaissent déjà dans la prolifération à grande échelle de la curation algorithmique ciblée et de la manipulation comportementale guidée par les données : ces techniques dopent les revenus des plates-formes des géants du numérique tout en favorisant les crises mondiales de méfiance sociale, la propagation des fausses informations et la progression de la polarisation culturelle et politique. Ces signes avant-coureurs, nous les voyons encore dans l'utilisation de modèles de risque prédictifs et de capacités de suivi numérique amélioré par algorithme dans des secteurs à fort impact comme le maintien de l'ordre, pratique qui a pour effet de renforcer et d'enraciner davantage les schémas de discrimination structurelle, de marginalisation systémique et d'inégalité.

Reconnaissant la nécessité d'une intervention humaine démocratique pour mettre l'innovation en matière d'IA sur la bonne voie, le Comité des Ministres du Conseil de l'Europe a adopté, en septembre 2019, le mandat du Comité ad hoc sur l'intelligence artificielle (CAHAI). Le CAHAI est chargé d'examiner la faisabilité et les éléments potentiels d'un cadre juridique pour la conception, le développement et le déploiement des systèmes d'IA qui soit compatible avec les normes du Conseil de l'Europe dans les domaines interdépendants des droits de l'homme, de la démocratie et de l'État de droit.

Première et nécessaire étape dans l'exécution de cette tâche, l'Étude de faisabilité adoptée par la plénière du CAHAI en décembre 2020 examine les solutions possibles d'une réponse juridique internationale qui permette de combler les lacunes de la législation et d'adapter l'utilisation d'instruments juridiques contraignants et non contraignants aux risques et opportunités spécifiques des systèmes d'IA. Cette étude examine comment les libertés et droits fondamentaux qui sont déjà codifiés dans le droit international en matière de droits de l'homme peuvent servir d'assise à un tel cadre juridique. Elle propose neuf principes et priorités qui sont adaptés aux nouveaux défis que posent la conception, le développement et le déploiement des systèmes d'IA. Une fois codifiés en droit, ces principes et priorités créeront un ensemble de droits et d'obligations interdépendants qui contribueront à garantir que la conception et l'utilisation des technologies d'IA sont conformes aux valeurs des droits de l'homme, de la démocratie et de l'État de droit. L'Étude de faisabilité conclut que les règles et régimes juridiques en vigueur ne sont ni suffisants à la sauvegarde de ces valeurs fondamentales appliquées à l'IA, ni adaptés, en tant que tels, à la création d'un environnement d'innovation en matière d'IA qui puisse être jugé suffisamment fiable pour orienter dans la bonne direction l'intelligence artificielle et les technologies reposant sur un usage intensif de données. Un nouveau cadre juridique est donc nécessaire.

Le présent Guide introductif a pour objet de présenter à un public général et non technique les grands concepts et principes exposés dans l'Étude de faisabilité du CAHAI. Il vise aussi à fournir des informations générales sur les domaines de l'innovation en matière d'IA, des normes de protection des droits de l'homme et des mécanismes de conformité, qui entrent dans le champ de cette étude. Conformément à l'engagement du Conseil de l'Europe de mener de larges consultations multipartites et de vastes actions de sensibilisation et de mobilisation, ce guide a été conçu pour faciliter la participation éclairée et pleinement pertinente d'un groupe inclusif de parties prenantes, le CAHAI souhaitant obtenir des retours d'information et des orientations générales sur les questions essentielles soulevées par l'Étude de faisabilité.

COMMENT UTILISER CE GUIDE INTRODUCTIF

Ce guide s'adresse aussi bien aux lecteurs qui n'ont pas de formation technique qu'à ceux qui ont déjà un acquis, mais souhaitent néanmoins « rafraîchir » leurs connaissances sur un ou plusieurs sujets traités dans l'Étude de faisabilité. Nous avons donc organisé les chapitres en modules, de sorte que le lecteur pourra choisir les sujets et sections qui l'intéressent le plus (et s'y consacrer en priorité) ou parcourir le guide de bout en bout.

Les trois premiers chapitres fournissent des informations générales sur l'IA et les technologies d'apprentissage automatique (ch. 2), les droits de l'homme, la démocratie et l'État de droit (ch. 3), et les risques et opportunités que présentent les systèmes d'IA dans le contexte des droits de l'homme (ch. 4). Le guide aborde ensuite certains des sujets plus spécifiques traités dans l'Étude de faisabilité. Le chapitre 5 expose les neuf principes et priorités proposés par le CAHAI pour servir de point d'ancrage à un cadre juridique transsectoriel fondé sur des valeurs. Il explique ensuite comment ces principes et priorités s'articulent avec les obligations et droits fondamentaux, afin de permettre leur traduction en lois. Le chapitre 6 brosse un panorama des instruments juridiques existants qui pourraient être intégrés dans un système plus large de mécanismes juridiques contraignants et non contraignants. Enfin, le chapitre 7 présente l'éventail des mécanismes de contrôle de conformité qui sont disponibles pour prendre en charge, mettre en œuvre et valider les contraintes mises en place par un cadre juridique.

À la fin de ce guide, vous trouverez un glossaire terminologique et une liste annotée de publications, dont certains travaux réalisés précédemment par le Conseil de l'Europe et d'autres organismes dans le domaine des normes et de la réglementation de l'IA et dans des domaines connexes des politiques technologiques.

Rien ne remplaçant l'original, nous recommandons vivement au lecteur de consulter directement la remarquable Étude de faisabilité et d'utiliser le présent guide comme un simple document d'accompagnement, où il trouvera, à portée de main, des informations contextuelles et des éclaircissements dans un format condensé.

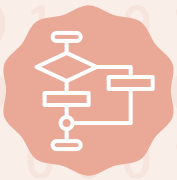
02 COMMENT LES SYSTÈMES D'IA FONCTIONNENT-ILS ?

Avant d'examiner comment la conception, le développement et le déploiement des technologies d'IA peuvent être rendus compatibles avec les droits de l'homme, la démocratie et l'État de droit grâce à un cadre d'instruments juridiques contraignants et non contraignants, nous commençons par une présentation des concepts techniques fondamentaux, des types d'apprentissage automatique et des étapes du cycle de vie de l'IA.

CONCEPTS TECHNIQUES

DONNÉES À CARACTÈRE PERSONNEL

Données pouvant être utilisées pour identifier une personne. Les données à caractère personnel comprennent diverses informations, par exemple le nom et le prénom, l'adresse, des données de localisation et diverses formes d'identification (numéro de passeport, numéro national d'identité, etc.).



ALGORITHME

Processus de calcul ou ensemble de règles qui sont exécutés pour résoudre un problème. Les algorithmes complexes sont généralement exécutés par des ordinateurs, mais un humain peut également suivre un processus algorithmique, par exemple une recette de cuisine ou une formule mathématique pour résoudre une équation.

APPRENTISSAGE AUTOMATIQUE (AA)

Type de calcul numérique utilisé pour découvrir des schémas dans un ensemble de données et faire des prévisions quant aux résultats d'une instance particulière. Le terme « apprentissage » est quelque peu trompeur, car les ordinateurs n'apprennent pas de la même façon que les humains. En fait, l'ordinateur est capable de trouver des similitudes et des différences dans les données en ajustant plusieurs fois ses paramètres (processus souvent appelé « apprentissage »). Lorsque les données d'entrée changent, les données de sortie évoluent en conséquence, et c'est en observant ces modifications que l'ordinateur apprend à repérer de nouveaux schémas. Cet apprentissage s'effectue en appliquant une formule mathématique à de gros volumes de données d'entrée pour produire un résultat correspondant. Ce principe est décrit en détail dans la section suivante.

INTELLIGENCE ARTIFICIELLE (IA)

L'intelligence artificielle ou « IA » a été définie de multiples façons au cours des dernières décennies, mais pour les besoins du présent guide, nous nous contenterons de la définir en décrivant ce qu'elle fait, autrement dit quel rôle elle joue dans le monde humain : les systèmes d'IA sont des modèles algorithmiques qui accomplissent dans le monde des fonctions cognitives et perceptives autrefois réservées aux êtres humains, lesquels ont la faculté de penser, juger et raisonner.

MÉGADONNÉES

Ensembles de données volumineux qui nécessitent souvent de grandes capacités de stockage et contiennent de grands volumes de données quantitatives pouvant être utilisés pour mettre en évidence des schémas ou des tendances. Les données contenues dans ces vastes ensembles peuvent être de différents types (nombres, noms, images, etc.) et être spécifiques à un but donné et tabulaires (structurées) ou générales et variées (non structurées).

SCIENCE DES DONNÉES

Domaine composé d'éléments de diverses disciplines, notamment l'informatique, les mathématiques, les statistiques et les sciences sociales, et généralement centré sur l'extraction d'informations et de schémas à partir d'ensembles de données dans le but de traiter une question ou un problème particulier ou d'y répondre.

POSSIBILITÉ D'INTERPRÉTATION

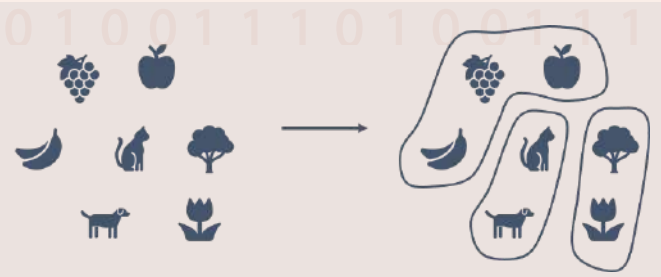
Si un humain est capable de déterminer comment un système d'IA ou d'apprentissage automatique est parvenu à une décision ou d'expliquer pourquoi il s'est comporté d'une certaine façon, alors le système peut être qualifié d'interprétable. La possibilité d'interprétation peut aussi renvoyer à la transparence des processus qui ont permis de mettre au point le système.

TYPES D'APPRENTISSAGE AUTOMATIQUE

APPRENTISSAGE SUPERVISÉ

Les modèles d'apprentissage supervisé sont entraînés sur des ensembles de données étiquetées. Dans ces modèles, de nombreux exemples sont présentés à un algorithme pour qu'il « apprenne » à mettre en correspondance des variables d'entrée (souvent appelées « caractéristiques ») avec des données de sortie souhaitées (aussi appelées « variables cibles » ou « étiquettes »). Les modèles d'AA sont capables de repérer, dans ces exemples, des schémas reliant les entrées aux sorties. Ils reproduisent ensuite ces schémas en appliquant les règles peaufinées pendant l'entraînement à de nouvelles données d'entrée, afin d'en déduire des classifications ou des prévisions. Un exemple classique d'apprentissage supervisé est l'utilisation de diverses variables comme la présence de mots tels que « lotterie » ou « vous avez gagné » pour prédire si un e-mail doit être classé dans la catégorie des courriels non sollicités (spams). L'apprentissage supervisé peut prendre la forme d'une classification (prédire si un e-mail est ou non sollicité par exemple), ou d'une régression, qui consiste à déterminer la relation entre des variables d'entrée et une variable cible. La régression et la classification linéaires sont les formes les plus simples d'apprentissage supervisé, mais d'autres modèles courants relèvent aussi de ce type d'apprentissage, comme les machines à vecteurs supports et les forêts aléatoires.

APPRENTISSAGE NON SUPERVISÉ



Contrairement à l'apprentissage supervisé, qui est un processus de mise en relation de points de données (comparaison de deux images dont l'une contient des objets déjà identifiés par exemple), l'apprentissage non supervisé consiste à laisser le système repérer des schémas dans des données. Dans l'apprentissage non supervisé, le système identifie des schémas et des structures en mesurant les densités ou similitudes des points de données dans l'ensemble des données. Le regroupement (clustering) est une application courante de l'apprentissage non supervisé. Le modèle reçoit des données non étiquetées et il détermine des similitudes et des différences parmi les points de données d'entrée. Des groupes sont ainsi formés en fonction des caractéristiques similaires, qui sont des facteurs importants pour le classement des données d'entrée. Dans l'exemple ci-contre, le modèle reçoit différentes sortes de fruits, des animaux, une fleur et un arbre. En analysant les caractéristiques propres à chaque catégorie, l'algorithme de regroupement est capable de classer les animaux, les fruits et les plantes en trois groupes distincts. La réduction de dimension est une autre forme d'apprentissage non supervisé.

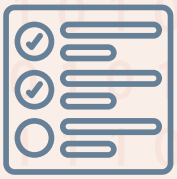
APPRENTISSAGE PAR RENFORCEMENT

Les modèles d'apprentissage par renforcement fondent leur apprentissage sur une observation de leurs interactions avec un environnement réel ou virtuel et non sur des données existantes. Les « agents » d'apprentissage par renforcement recherchent une voie optimale pour effectuer une tâche, en enchaînant une série d'actions qui maximisent la probabilité de réaliser la tâche en question. Si les actions qu'ils prennent conduisent à un échec, ils sont sanctionnés, sinon ils sont récompensés. Les « agents » sont programmés pour choisir leurs actions de façon à maximiser leur récompense. Ils « apprennent » de leurs réussites et de leurs échecs passés, s'améliorent au fil des multiples essais et erreurs, et peuvent être conçus pour élaborer des stratégies à long terme afin de maximiser leur récompense globale plutôt que se concentrer exclusivement sur l'action suivante. Par exemple, l'apprentissage par renforcement est couramment utilisé pour mettre au point les véhicules autonomes (voitures sans chauffeur). Il permet aux ingénieurs d'améliorer les performances du véhicule dans un environnement simulé, en testant différents éléments tels que l'accélération ou la réponse du véhicule à la régulation du trafic. Grâce à ces interactions avec l'environnement simulé, les « agents » d'apprentissage par renforcement sont sanctionnés ou récompensés en fonction de l'achèvement de la tâche, ce qui a une incidence sur les performances futures du véhicule.

ÉTAPES DU CYCLE DE VIE DE L'IA

CONCEPTION

Planification du projet



La toute première tâche de l'équipe de projet est de fixer les objectifs du projet. Cette étape appelée « planification » comprend, entre autres, des activités de mobilisation des parties prenantes, une analyse d'impact général, un recensement des étapes clés du projet et l'évaluation des ressources et des capacités au sein de l'équipe ou de l'organisation. À titre d'exemple, nous considérons une équipe de projet d'IA dans un contexte agricole. À cette étape, l'équipe doit décider s'il convient ou non d'utiliser une application d'IA pour prévoir quels terrains seront cultivables dans les cinq prochaines années et quel sera a priori le rendement des cultures. L'étape de planification permet à l'équipe de réfléchir aux questions éthiques, socioéconomiques, juridiques et techniques avant d'investir des ressources dans le développement du système d'IA.

Formulation du problème

L'équipe de projet doit définir le problème que le modèle devra résoudre et déterminer quelles données d'entrée sont nécessaires et à quelles fins. Elle doit aussi examiner les incidences éthiques et juridiques de l'utilisation des données et fournir une description détaillée des conséquences voulues et non voulues de cette utilisation. Dans notre exemple, l'équipe décide que le projet concerne essentiellement le rendement des cultures. Cette formulation plus précise permet d'identifier une question spécifique qui peut être appréhendée à l'aide de données et de s'assurer que le résultat sera en phase avec des considérations éthiques et juridiques telles que la diversité biologique ou l'utilisation des sols.

Extraction ou acquisition de données

Cette étape met en jeu des processus qui permettent de recueillir des données afin résoudre le problème. L'extraction de données consiste par exemple à collecter automatiquement des données en ligne (web scraping) ou à collecter des données via des enquêtes ou d'autres méthodes analogues, tandis que l'acquisition de données passe, entre autres, par la signature d'un accord juridique permettant d'obtenir des jeux de données existants. Dans notre exemple, l'équipe décide que pour résoudre le problème, il faut déterminer les facteurs qui sont importants pour la prévision du rendement des cultures au cours d'une saison agricole donnée. Elle décide donc de se procurer des données auprès d'un organisme public et de coopératives agricoles, ce qui, dans les deux cas, nécessite la signature d'accords juridiques de partage de données.

Analyse des données

À ce stade, l'équipe de projet peut commencer à examiner les données. Il s'agit avant tout de réaliser une analyse exploratoire des données (AED) à un niveau poussé. L'AED consiste à déterminer la composition des données au moyen d'une représentation visuelle et de statistiques sommaires. On se demandera notamment s'il existe des données manquantes (données incomplètes), des valeurs aberrantes (données inattendues), des classes non équilibrées (données déséquilibrées) ou des corrélations. Dans notre exemple, l'équipe pourra créer des représentations visuelles pour déterminer la distribution des types de cultures sur l'ensemble des exploitations, les conditions météorologiques et les niveaux de pH des sols, et voir s'il y a des données manquantes.

Pré-traitement

Dans la phase de développement du cycle de vie de l'IA, l'étape de prétraitement est souvent la partie qui demande le plus de temps. Le prétraitement comprend diverses tâches, notamment le nettoyage des données (reformatage ou élimination des informations incomplètes) et la préparation préalable des données (transformation des données dans un format adapté à la modélisation), tâches qui, parmi d'autres, viennent alimenter le processus d'entraînement du modèle. Dans notre exemple, pendant l'étape de prétraitement, les membres de l'équipe s'aperçoivent que les niveaux de pH des sols sont représentés par des données numériques et des données de type « chaîne de caractères », ce qui posera des problèmes lors de l'exécution des modèles. Ils décident donc que toutes les données de pH doivent être du même type et transforment les chaînes de caractères en nombres.

Choix du modèle et entraînement

Les modèles doivent être choisis de façon à répondre au problème défini pendant la phase de conception. Si leur complexité, plus ou moins grande, est à prendre en compte, d'autres paramètres comme le type, la quantité et la disponibilité des données interviennent dans le choix du modèle. Avec un modèle trop simple, il y a un risque de sous-ajustement par rapport aux données (c'est-à-dire de non-prise en compte des données). Par ailleurs, les données prétraitées sont réparties en deux groupes, l'un servant à l'apprentissage, l'autre aux essais, afin d'éviter le sur-ajustement. Ce phénomène se produit lorsque le modèle reflète de trop près les données d'apprentissage et n'est pas en mesure de s'adapter pour faire des prévisions exactes à partir de « nouvelles » données (données d'entrée ne figurant pas dans la base d'apprentissage). Les données d'apprentissage servent à affiner les paramètres du modèle choisi. Dans notre exemple, l'équipe de projet choisit un modèle de régression linéaire afin d'utiliser des données antérieures pour prévoir les rendements des futures récoltes. L'objectif est de choisir un modèle interprétable afin de pouvoir expliquer totalement les résultats. Le choix se porte donc logiquement sur une technique simple, en l'occurrence la régression linéaire.

Essai et validation du modèle

Après l'apprentissage, le modèle est ajusté et testé au moyen de « nouvelles » données.Des jeux de validation sont utilisés pour ajuster des attributs de haut niveau du modèle (par exemple, des hyperparamètres qui régissent la manière dont le modèle apprend). Ces jeux de validation sont souvent créés en divisant d'emblée le jeu de données en trois parties, par exemple 60 % de données pour l'apprentissage, 20 % pour les essais et 20 % pour la validation. Au cours de la validation, des éléments de l'architecture du modèle peuvent être modifiés pour jouer sur la performance de ce dernier. Dans notre exemple, l'équipe s'aperçoit, en exécutant le modèle, que le nombre de variables intégrées provoque un sur-ajustement. Elle décide donc d'ajouter un terme de régularisation (méthode permettant de diminuer l'erreur du modèle) afin de supprimer des variables non essentielles. Puis le modèle est testé sur des données qui ne lui sont pas familières pour mimer une application du monde réel et confirmer sa performance et sa justesse.

Rapport sur le modèle

Après avoir entraîné, validé et testé le modèle, l'équipe doit produire un rapport d'évaluation (comprenant diverses mesures de performance et des analyses d'impact) et fournir des informations détaillées sur le flux de travaux du modèle, afin de permettre des discussions ouvertes et transparentes sur les produits en sortie. Dans notre exemple, pour clore la phase de développement, l'équipe établit un document contenant diverses mesures de performance de son modèle et décrivant les processus qui ont permis d'aboutir à son itération actuelle, notamment le prétraitement et la décision éventuelle d'appliquer une régularisation lors des étapes d'essai et de validation.

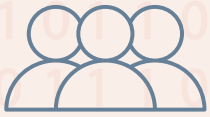
DÉPLOIEMENT

Mise en œuvre du modèle

La phase suivante du cycle de vie de l'IA consiste à déployer le modèle entraîné dans le monde réel. La mise en œuvre effective permet d'intégrer le modèle dans un système plus vaste. Une fois mis en service, le modèle traite de nouvelles données et remplit ainsi sa mission définie pendant la phase de conception. Dans notre exemple, l'équipe de projet d'IA décide que le modèle de rendement des cultures est prêt à l'emploi. Elle le met à disposition de plusieurs coopératives agricoles et leur demande de l'exécuter sur leurs propres données pour voir s'il apporte des éclairages utiles.

Formation des utilisateurs

Les personnes chargées de la mise en œuvre doivent être formées à la logique du système pour être en mesure d'expliquer ses décisions aux responsables en langage clair et d'évaluer, en toute indépendance et sans préjugé, la qualité, la fiabilité et l'impartialité des résultats. Dans notre exemple, l'équipe forme certains exploitants agricoles à l'utilisation de son modèle, puis ces utilisateurs lui font savoir s'ils jugent le système utile, fiable et précis, entre autres paramètres.



Suivi

Une fois mis en service par l'équipe de projet, le modèle doit faire l'objet d'un suivi afin de s'assurer qu'il remplit toujours sa mission, qu'il est utilisé de manière responsable et dans les limites de son champ d'application, et qu'il réagit correctement aux nouvelles situations du monde réel. Dans notre exemple, l'équipe remarque qu'une nouvelle variable destinée à mesurer la qualité de l'eau a été publiée par un organisme de normalisation, ce qui pourrait créer une incompatibilité des données par rapport aux normes, cette variable n'ayant pas été prévue à l'origine dans la base de données d'apprentissage. L'équipe décide donc d'intégrer ce changement dans le modèle pour rester compatible avec les normes et pratiques agricoles.

Mise à jour et retrait

Avec le temps, le modèle peut perdre en efficacité, obligeant l'équipe de contrôle à réexaminer des étapes antérieures de la phase de développement, y compris le choix du modèle et son apprentissage. Si des changements plus importants s'imposent, le système devra éventuellement être retiré, et le processus de conception devra reprendre à l'étape de planification du projet. Dans notre exemple, l'équipe a dû réentraîner le modèle plusieurs fois pour prendre en compte de nouvelles variables et corriger les jeux de données qui n'étaient plus normalisés. Elle continue de suivre le modèle tout en envisageant des alternatives, notamment le développement d'un nouveau système.

03 BRÈVE INTRODUCTION AUX DROITS DE L'HOMME, À LA DÉMOCRATIE ET À L'ÉTAT DE DROIT

« Tous les droits de l'homme sont universels, indissociables, interdépendants et intimement liés. »

-United Nations Vienna Declaration, 1993

Les droits de l'homme, la démocratie et l'État de droit sont intimement liés. La capacité des États légitimes à sauvegarder efficacement les droits de l'homme repose sur l'interdépendance d'institutions démocratiques robustes et responsables, sur des mécanismes décisionnels inclusifs et transparents et sur un système judiciaire indépendant et impartial qui garantit l'État de droit. De façon générale, les droits de l'homme sont les libertés et droits fondamentaux dont jouit toute personne dans le monde, du berceau au tombeau, et qui préservent et protègent la dignité inviolable de chaque individu, indépendamment de sa race, de son origine ethnique, de son sexe, de son âge, de son orientation sexuelle, de sa classe, de sa religion, de son handicap, de sa langue, de sa nationalité ou de toute autre caractéristique qui lui est attribuée.Ces libertés et droits fondamentaux créent des obligations qui contraignent les États à respecter, protéger et réaliser les droits de l'homme. Lorsque ces obligations ne sont pas respectées, les personnes concernées ont le droit de former des recours juridiques afin d'obtenir réparation de toute violation des droits de l'homme.

LES DROITS DE L'HOMME EN QUELQUES DATES

Historiquement, l'ensemble des droits et principes fondamentaux connus sous le nom de droits de l'homme est apparu au milieu du xxe siècle, dans le sillage des atrocités et du traumatisme de la seconde guerre mondiale.

1948 Les Nations Unies adoptent la **Déclaration universelle des droits de l'homme** (DUDH), qui définit une première norme internationale en matière de libertés et droits fondamentaux. Bien que ce document ne soit pas juridiquement contraignant, il servira de base aux nombreux traités, conventions et chartes en matière de droits de l'homme adoptés dans le monde entier jusqu'à ce jour.

1953 La **Convention européenne des droits de l'homme** (CEDH) entre en vigueur. Rédigé par le Conseil de l'Europe en 1950, ce traité international consacre les droits civils et politiques que les 47 États membres du Conseil de l'Europe sont juridiquement contraints de respecter. Outre l'établissement de droits fondamentaux visant à sauvegarder la dignité inviolable de tout individu, la CEDH fait obligation aux États de protéger les citoyens ordinaires contre les violations des droits de l'homme.

1961 Le Conseil de l'Europe publie sa **Charte sociale européenne** (ESC) et l'ouvre à la signature. Ce traité étend les droits fondamentaux pour y inclure des droits sociaux et économiques concernant la santé, les conditions de travail, le logement, le travail des migrants, l'égalité entre les femmes et les hommes et la sécurité sociale. Des protocoles additionnels sont ajoutés en 1988 pour renforcer l'égalité des chances sur le lieu de travail, la participation des travailleurs et la protection des personnes démunies et des personnes âgées. Une Charte révisée est adoptée en 1996.

1966 L'ONU adopte son **Pacte international relatif aux droits civils et politiques** (PIDCP) et son **Pacte international relatif aux droits économiques, sociaux et culturels** (PIDESC). Le PIDCP comprend le droit de ne pas subir de torture, le droit à un procès équitable, la non-discrimination et le droit au respect de la vie privée. Le PIDESC étend les droits fondamentaux pour y inclure le droit à des conditions de travail justes, le droit à la santé, le droit à un niveau de vie suffisant, le droit à l'éducation et le droit à la sécurité sociale. Ensemble, la DUDH, le PIDCP et le PIDESC de l'ONU sont désormais connus sous le nom de **Charte internationale des droits de l'homme.**

2009 **La Charte des droits fondamentaux de l'Union européenne** (CDF UE) se voit donner force de loi par le Traité de Lisbonne. Cet instrument codifie dans le droit communautaire un ensemble fondamental de droits civils, politiques, sociaux, économiques et culturels pour les citoyens de l'Union européenne. Les domaines des droits de l'homme visés par la CDF UE comprennent les droits relatifs à la dignité humaine, aux libertés fondamentales, à l'égalité, à la solidarité, aux droits économiques et aux droits à la participation à la vie de la communauté.

EUX FAMILLES DE DROITS DE L'HOMME

Les principes qui constituent les droits de l'homme peuvent être classés en deux groupes :

Droits civils et politiques

Droits essentiels:

-Droit à la vie et à la dignité humaine
-Droit à l'intégrité physique et mentale
-Droit à la liberté et à la sûreté des personnes
-Droit de ne pas être soumis à la torture ni à des traitements cruels
-Droit à un procès équitable et une procédure régulière
-Droit à un recours effectif
-Liberté de pensée, de conscience et de religion
-Liberté d'expression et d'opinion
-Droit au respect de la vie privée et familiale
-Droit à la protection des données à caractère personnel
-Droit à la non-discrimination
-Droit à l'égalité devant la loi
-Liberté de réunion et d'association
-Droit de participer à la gestion des affaires publiques

Droits sociaux, économiques et culturels

Droits essentiels:

-Droit à des conditions de travail équitables et à la sécurité et à l'hygiène dans le travail
-Droit à une rémunération équitable
-Droit à la formation professionnelle
-Droit à l'égalité des chances au travail
-Droit syndical et de négociation collective
-Droit à la sécurité sociale
-Droit à l'éducation
-Droit à un niveau de vie suffisant
-Droit à l'assistance sociale et médicale
-Droit à la protection de la santé
-Droit à la protection des travailleurs migrants
-Droit des personnes âgées à une protection sociale
-Droit à la protection contre le harcèlement sexuel
-Droit à la protection contre la pauvreté et l'exclusion sociale

INTERDÉPENDANCE DES DROITS DE L'HOMME, DE LA DÉMOCRATIE ET DE L'ÉTAT DE DROIT

L'interdépendance des droits de l'homme, de la démocratie et de l'État de droit trouve son origine dans le fait que ces principes sont, par nature, imbriqués et symbiotiques. La légitimité des institutions démocratiques est enracinée dans la notion selon laquelle tous les citoyens ont le droit, à égalité, de participer à la vie commune de la collectivité et à la conduite des décisions collectives qui les concernent. Cependant, pour qu'ils puissent exercer ce droit de participer à la gestion des affaires publiques, ils doivent d'abord posséder de nombreux autres droits civils, politiques, sociaux, culturels et économiques interdépendants :

- Ils doivent jouir de la liberté de pensée, d'association, de réunion et d'expression.
- Ils doivent bénéficier d'un respect égal devant la loi et d'une protection contre toute forme de discrimination qui entraverait leur participation pleine et équitable à la vie de la collectivité.
- Ils doivent avoir accès aux moyens matériels de leur participation grâce à une éducation appropriée, à un niveau de vie suffisant et à des conditions de travail satisfaisantes, à la santé, à la sécurité et à la sécurité sociale.
- Ils doivent avoir accès à des recours judiciaires effectifs en cas d'atteinte à l'un de leurs droits fondamentaux.

Pour répondre à ce dernier point, l'État de droit apporte le fondement institutionnel qui permet de garantir la participation démocratique et la protection des libertés et droits fondamentaux. Un pouvoir judiciaire indépendant et impartial, qui garantit aux citoyens des procédures judiciaires régulières et l'égalité et l'équité de traitement au regard de la loi, agit comme le garant d'une possibilité de recours chaque fois que les libertés ou droits fondamentaux sont bafoués.

04 OPPORTUNITÉS ET RISQUES DE L'IA ET DE L'APPRENTISSAGE AUTOMATIQUE, ET LEURS RÉPERCUSSIONS SUR LES DROITS DE L'HOMME, LA DÉMOCRATIE ET L'ÉTAT DE DROIT

Les technologies d'intelligence artificielle (IA) offrent de nombreuses possibilités d'amélioration de la qualité de vie et du fonctionnement de l'État. Par leur puissance, leur portée et leur rapidité, les systèmes d'IA peuvent apporter un gain d'efficience et d'efficacité dans de nombreux domaines, notamment la santé, les transports, l'éducation et l'administration. Ils peuvent prendre le relais des travailleurs humains pour accomplir des tâches fastidieuses, dangereuses, désagréables et complexes. Cependant, les technologies d'IA peuvent aussi porter atteinte aux droits de l'homme, à la démocratie et à l'État de droit. Cette symbiose de possibilités et de risques doit être comprise à la lumière du caractère « **sociotechnique** » de l'IA, laquelle couvre un large éventail de technologies très élaborées, mises en œuvre dans des contextes humains et conçues pour atteindre des objectifs définis par des humains. On peut donc dire que les technologies d'IA sont le reflet des valeurs et des choix de leurs concepteurs et utilisateurs.

L'IA peut être utilisée pour faire des prévisions concernant le comportement humain, déceler des signes de maladie et évaluer les risques auxquels sont exposés les intérêts et le bien-être d'autrui. Mais toutes ces tâches peuvent porter atteinte aux droits, aux possibilités et au bien-être de ceux qu'elles visent. C'est pourquoi la responsabilité est un aspect essentiel de la mise au point et de l'utilisation de ces systèmes. Si l'IA peut remplacer les êtres humains pour accomplir des tâches fastidieuses ou complexes, les choix qui sont faits pendant la construction et l'utilisation des systèmes d'IA peuvent aboutir à la reproduction de préjugés néfastes et autres erreurs du jugement humain qui nuisent aux personnes concernées et à la société dans son ensemble, par des voies plus difficiles à identifier que lorsque ces tâches sont réalisées par des humains.

Par conséquent, outre l'évaluation des caractéristiques techniques de tel système ou telle technologie, la responsabilité en matière d'IA nous impose d'examiner avec soin les dangers et les avantages potentiels de l'IA pour les individus et pour les groupes. L'un de ces dangers réside dans les biais injustifiés, qui peuvent se manifester ouvertement, par exemple lorsque des modèles d'IA font des prévisions discriminatoires ou traitent un groupe démographique ou un individu particulier différemment des autres sans justification. Mais l'opacité de certains systèmes d'IA rend difficile l'évaluation de leur potentiel de nuisance. De plus, le fonctionnement des technologies d'IA est difficile à interpréter ou à expliquer non seulement parce que ces systèmes sont construits par des experts, mais aussi en raison de leur complexité technique et des droits de propriété intellectuelle.

Les répercussions des systèmes d'IA sur les droits de l'homme peuvent être examinées sur la base des dispositions de la Convention européenne des droits de l'homme (CEDH) et de la Charte sociale européenne (CSE), notamment leurs garanties spécifiques concernant **la liberté et la justice, le respect de la vie privée, la liberté d'expression, l'égalité et la non-discrimination, ainsi que les droits sociaux et économiques.** L'IA a aussi des répercussions sur la démocratie et l'État de droit qui ne relèvent pas clairement des dispositions de la CEDH et de la CSE, mais qui n'en sont pas moins des considérations importantes. Un examen approfondi des risques et des possibilités des systèmes d'IA nous aidera à déterminer les libertés et droits existants qui offrent les protections nécessaires, ceux qui nécessitent une clarification et ceux qui doivent être adaptés aux nouveaux défis et possibilités de l'IA et de l'apprentissage automatique.

Liberté et justice: L'IA peut avoir des répercussions négatives sur la liberté des individus et sur la justice qui les protège, en particulier lorsque cette technologie est mise en œuvre dans des contextes à fort impact comme la justice pénale. La complexité et l'opacité des systèmes d'IA peuvent faire obstacle au droit à un procès équitable, notamment le droit à l'égalité des armes, qui veut qu'une partie faisant l'objet d'une décision algorithmique puisse dûment examiner le raisonnement de ces systèmes et le contester. Si l'utilisation de l'IA dans ce contexte permet parfois de limiter l'arbitraire et les mesures discriminatoires, les décisions judiciaires étayées ou guidées par l'IA peuvent aussi compromettre l'élaboration du droit et l'autonomie décisionnelle de la justice. Les acteurs judiciaires devraient donc avoir une compréhension suffisante de l'IA qu'ils utilisent, pour que l'exercice des responsabilités au regard des décisions prises au moyen de cette technologie soit garanti.

Tout système qui étaye les décisions en matière de sanction pénale au moyen d'une note représentant le risque qu'un condamné récidive doit être interprétable, vérifiable et contestable par l'accusé, afin que le processus judiciaire soit équitable et ouvert.

Un système qui analyse les expressions du visage, le ton de la voix, le choix des mots et d'autres signes biométriques et les compare à des modèles pour prédire si un candidat à un emploi sera une « bonne » recrue peut porter atteinte à l'intimité corporelle et émotionnelle de ce dernier.

Liberté d'expression, d'association et de réunion: Le bon fonctionnement d'une démocratie exige un discours social et politique ouvert et la réduction au minimum de toute influence ou manipulation indue par toute personne ou institution particulière. Or l'IA met ces valeurs en danger lorsqu'elle enregistre et analyse l'utilisation des sites web et des réseaux sociaux ou qu'elle extrait des données au moyen d'une surveillance biométrique dans le but de recueillir ou de traiter des informations sur l'activité des personnes en ligne ou hors ligne. Utilisée ainsi, l'IA contribue à donner l'impression que l'on est observé et écouté, ce qui peut avoir un effet dissuasif et empêcher l'expression libre et l'action politique. Les algorithmes d'IA utilisés par les plates-formes des réseaux sociaux déterminent les messages et publicités à afficher, créant ainsi une expérience d'utilisateur qui exploite les centres d'intérêt et les préjugés individuels tout en renforçant potentiellement des conceptions du monde qui sont clivantes, antidémocratiques ou violentes. L'IA est également employée pour produire des vidéos très réalistes mais factices, de faux comptes d'utilisateur et d'autres contenus fabriqués de toute pièce qui peuvent altérer la faculté des individus à se faire une opinion éclairée reposant sur des faits.

Respect de la vie privée: L'IA peut accéder à d'immenses volumes de données à caractère personnel et les traiter à une vitesse incroyable. Elle peut faire des prédictions sur le comportement des individus, leur état d'esprit et leur identité en captant des informations qui ne sont pas nécessairement considérées comme étant personnelles ou privées, notamment les expressions du visage, le rythme cardiaque, l'emplacement physique et d'autres données apparemment banales ou librement accessibles. Ces prédictions peuvent constituer une atteinte à la vie privée d'un individu telle qu'il la conçoit, mais aussi avoir des « effets panoptiques » en l'amenant à modifier son comportement parce qu'il soupçonne qu'on l'observe ou qu'on analyse ses faits et gestes.

Les systèmes de reconnaissance faciale en direct peuvent empêcher les citoyens d'exercer leur liberté de réunion et d'association, ce qui les prive de la protection de l'anonymat et a un effet dissuasif sur la solidarité sociale et la participation démocratique. La surveillance biométrique utilisant l'IA peut aussi leur confisquer le droit à un consentement explicite et éclairé concernant la collecte de données à caractère personnel.

Égalité et non-discrimination: Les systèmes d'IA ont la capacité de reproduire et de renforcer les schémas discriminatoires qui sont déjà à l'œuvre dans la société où ils sont développés et utilisés. Ce phénomène se produit lorsque les choix faits pendant la conception et le déploiement sont déterminés par les préjugés et les « angles morts » des concepteurs. Il se produit également lorsque des inégalités et des discriminations historiques structurelles sont enracinées dans les jeux de données servant à entraîner les modèles d'IA et d'apprentissage automatique. En effet, les décisions discriminatoires des humains qui ont créé ces jeux de données biaisées peuvent être à l'origine de décisions et de comportements algorithmiques discriminatoires.

Les systèmes de prévision policière qui s'appuient sur des données historiques présentent le risque de reproduire les résultats de précédentes pratiques discriminatoires et d'engendrer un phénomène de « rétroaction », par lequel toute nouvelle décision de police produit de nouvelles données et aboutit in fine à la suspicion et à l'arrestation disproportionnées d'individus appartenant à des groupes marginalisés.

Les sociétés qui proposent des services de livraison et de VTC coordonnés par des applications mobiles peuvent automatiser la gestion et la supervision d'une vaste main-d'œuvre, et déshumaniser les relations de travail et donc les pratiques de management. Cette automatisation peut limiter la liberté d'action et les possibilités de recours des employés qui sont confrontés à des décisions erronées ou injustes prises par des managers algorithmiques au sujet de leur salaire ou de leur emploi.

Droits sociaux et économiques: Les systèmes d'IA sont de plus en plus souvent utilisés par les employeurs et les pouvoirs publics selon des modalités qui mettent en danger les droits sociaux et économiques des citoyens. Ainsi, les employeurs se servent de l'IA pour suivre le comportement des salariés, faire échec aux tentatives de syndicalisation et prendre des décisions concernant le recrutement, les salaires et les promotions. Dans certains milieux professionnels, les humains sont essentiellement gérés par des systèmes de décision algorithmique, ce qui peut restreindre leurs possibilités économiques. De même, les pouvoirs publics qui utilisent l'IA pour déterminer les bénéficiaires des avantages sociaux et des soins de santé ont un impact sur la prospérité économique des individus.En effet, si ce type de gestion est insuffisamment contrôlé, certains peuvent se voir refuser des prestations auxquelles ils ont droit et leur qualité de vie est alors menacée. L'automatisation des décisions d'éligibilité et de l'attribution des prestations sociales peut certes permettre un gain d'efficacité, mais les non-bénéficiaires se retrouvent parfois sans recours ou face à des formulaires ou autres procédures complexes, sans possibilité d'être assistés avec la bienveillance nécessaire.

Ces enjeux en matière de droits de l'homme s'accompagnent d'une concentration des pouvoirs aux mains des grands acteurs de l'IA, sociétés publiques ou privées qui conçoivent les systèmes ou les mettent en œuvre. Les exploitants des grandes plates-formes en ligne utilisent cette technologie pour définir les contenus à afficher et les auteurs à mettre en avant, non pas au service d'intérêts démocratiques, mais de leurs propres intérêts. Les États, pour leur part, utilisent l'IA pour classer et ordonner les informations et pour suivre et pister les citoyens. Tant les entreprises que les États peuvent, au moyen de l'IA, façonner les opinions et réprimer la contestation.

Au vu de ces considérations et de ces inquiétudes, les États devraient suivre une politique prudente en matière d'adoption et de réglementation de l'IA, pour permettre une réalisation équilibrée du potentiel de cette technologie tout en limitant le plus possible les risques qu'elle présente pour les êtres humains et leurs intérêts. Lorsque cela ne suffit pas à atténuer les risques, les États devraient envisager des mesures d'interdiction d'utilisation de l'IA. Et lorsque des incertitudes demeurent quant au niveau ou à l'impact des risques, ils devraient renforcer le contrôle et le suivi réglementaires des systèmes d'IA et être prêts à en interdire l'usage.

05 PRINCIPES ET PRIORITÉS POUR LA DÉFINITION D'UN CADRE JURIDIQUE

En septembre 2019, le Comité des Ministres du Conseil de l'Europe a adopté le mandat du Comité ad hoc sur l'intelligence artificielle (CAHAI). Le CAHAI est chargé d'examiner la faisabilité et les éléments potentiels d'un cadre juridique pour le développement, la conception et le déploiement des systèmes d'IA, sur la base des normes du Conseil de l'Europe dans les domaines interdépendants des droits de l'homme, de la démocratie et de l'État de droit. Première et nécessaire étape dans l'exécution de cette tâche, l'*Étude de faisabilité* adoptée par la plénière du CAHAI en décembre 2020 propose neuf principes et priorités qui doivent constituer le socle de ce cadre d'instruments juridiques contraignants et non contraignants:

DIGNITÉ HUMAINE

Tous les individus sont intrinsèquement et incontestablement dignes de respect du simple fait de leur statut d'être humain. Les êtres humains devraient être traités comme des sujets moraux et non comme des objets que l'on note ou manipule avec des algorithmes.

LIBERTÉ ET AUTONOMIE HUMAINES

Les êtres humains devraient avoir le droit de déterminer, en toute connaissance de cause et de façon autonome, si, quand et comment les systèmes d'IA doivent être utilisés. Ces systèmes devraient être utilisés non pas pour conditionner ou contrôler les humains, mais pour enrichir leurs capacités.

PRÉVENTION DES PRÉJUDICES

L'intégrité physique et mentale des êtres humains et la durabilité de la biosphère doivent être protégées, et des garanties supplémentaires doivent être mises en place pour protéger les personnes en situation de vulnérabilité. Les systèmes d'IA ne doivent pas être autorisés à porter préjudice au bien-être des êtres humains et à la santé de la planète.

NON-DISCRIMINATION, ÉGALITÉ ENTRE LES FEMMES ET LES HOMMES, ÉQUITÉ ET DIVERSITÉ

Tous les êtres humains ont le droit de ne pas subir de discrimination et ont droit à l'égalité et à un traitement égal de par la loi. Les systèmes d'IA doivent être conçus de façon à être justes, équitables et inclusifs quant à leurs effets bénéfiques et à la répartition de leurs risques.

TRANSPARENCE ET EXPLICABILITÉ DES SYSTÈMES D'IA

Lorsqu'un produit ou un service utilise un système d'IA, cela doit être clairement annoncé aux personnes visées. Doivent également être fournies des informations utiles concernant la logique qui sous-tend les résultats produits par le système.

PROTECTION DES DONNÉES ET DROIT À LA VIE PRIVÉE

La conception et l'utilisation des systèmes d'IA qui reposent sur le traitement de données à caractère personnel doivent garantir le droit au respect de la vie privée et familiale, y compris le droit de la personne à contrôler ses propres données. Le consentement libre, éclairé et non ambigu doit intervenir à cet égard.

RESPONSABILITÉ ET OBLIGATION DE RENDRE DES COMPTES

Toutes les personnes participant à la conception et au déploiement de systèmes d'IA doivent rendre des comptes lorsque des normes juridiques en vigueur sont bafouées ou que des utilisateurs finaux ou d'autres personnes subissent un préjudice injuste. Les victimes doivent avoir accès à des recours effectifs pour demander réparation.

DÉMOCRATIE

Des mécanismes de contrôle transparents et inclusifs doivent veiller à ce que les processus de décision démocratiques, le pluralisme, l'accès à l'information, l'autonomie et les droits économiques et sociaux soient garantis dans le contexte de la conception et de l'utilisation des systèmes d'IA.

ÉTAT DE DROIT

Les systèmes d'IA ne doivent pas mettre en danger l'indépendance de la justice, la régularité des procédures, ni l'impartialité. Pour cela, il faut garantir la transparence, l'intégrité et l'impartialité des données et des méthodes de traitement des données.

RELATIONS ENTRE LES PRINCIPES, LES DROITS ET LES OBLIGATIONS

Ces neuf principes et priorités **s'appliquent horizontalement**. Ils s'appliquent à la conception, au développement et au déploiement des systèmes d'IA **dans tous les secteurs et tous les cas d'utilisation**, mais ils pourraient être associés à une approche sectorielle énonçant des exigences contextuelles (plus détaillées) sous la forme d'instruments non contraignants, comme des normes sectorielles, des lignes directrices ou des listes d'évaluation.

Le cadre juridique est censé partir de ce point de vue très large. Son objectif sera de garantir les neuf principes et priorités, et ce en définissant, d'une part, les **droits concrets** qui permettent la réalisation de ces principes transsectoriels au niveau individuel et, d'autre part, les **obligations et exigences clés** que devront respecter les développeurs et les utilisateurs pour construire et utiliser des systèmes d'IA conformes aux droits de l'homme, à la démocratie et à l'État de droit. Les droits qui seront définis pourront être 1) des droits existants, 2) de nouveaux droits adaptés aux défis soulevés par l'IA et aux opportunités qu'offre cette technologie ou 3) des droits existants qui auront été précisés.

Table de correspondances entre les principes et priorités et les droits et obligations :

	DROITS SUBSTANTIELS	OBLIGATIONS CLÉS
DIGNITÉ HUMAINE	- Le droit à la dignité humaine, le droit à la vie **(art. 2 de la Convention européenne des droits de l'homme – CEDH)** et le droit à l'intégrité physique et mentale. - Le droit de toute personne d'être informée du fait qu'elle interagit avec un système d'IA et non avec un être humain. - Le droit de refuser d'interagir avec un système d'IA chaque fois que cela pourrait nuire à la dignité humaine.	- Lorsque des tâches réalisées par des machines plutôt que par des êtres humains risquent de nuire à la dignité humaine, les États membres devraient veiller à ce qu'elles soient confiées exclusivement à des êtres humains. - Les États membres devraient exiger que les acteurs du déploiement de l'IA informent les êtres humains qu'ils interagissent avec un système d'IA et non avec un être humain dans tous les contextes où une confusion est possible.
LIBERTÉ ET AUTONOMIE HUMAINES	- Le droit à la liberté et à la sécurité **(art. 5 de la CEDH).** - Le droit de la personne humaine à l'autonomie et à l'autodétermination. Le droit de ne pas être soumis à une décision reposant uniquement sur un traitement automatisé lorsque cela produit des effets juridiques sur des personnes physiques ou, de façon similaire, leur porte atteinte de manière significative. - Le droit de contester et de dénoncer effectivement les décisions éclairées et/ou prises par un système d'IA et d'exiger que de telles décisions soient réexaminées par une personne. - Le droit de décider librement de ne pas faire l'objet de manipulations, de profilages individualisés et de prédictions basés sur l'IA, y compris dans les cas de traitement de données à caractère non personnel. - Le droit d'avoir la possibilité, lorsque des motifs impérieux et légitimes qui s'y opposent ne l'excluent pas, de choisir d'être en contact avec un être humain plutôt qu'avec un robot.	- Les manipulations, profilages individualisés et prédictions fondés sur l'IA nécessitant le traitement de données à caractère personnel doivent tous respecter les obligations énoncées dans la Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. - Les États membres devraient mettre en œuvre de manière effective la version modernisée de cette Convention (« Convention 108+ ») pour mieux traiter les questions liées à l'IA. - Les États membres devraient imposer aux acteurs du développement et du déploiement de l'IA de mettre en place des mécanismes de contrôle humains pour préserver l'autonomie humaine, d'une manière qui soit adaptée aux risques spécifiques découlant du contexte dans lequel le système d'IA concerné est développé et utilisé. - Les États membres devraient imposer aux acteurs du développement et du déploiement de l'IA de faire dûment connaître les possibilités de recours en temps utile.

	DROITS SUBSTANTIELS	OBLIGATIONS CLÉS
PRÉVENTION DES PRÉJUDICES	- Le droit à la vie **(art. 2 de la CEDH)** et le droit à l'intégrité physique et mentale. - Le droit à la protection de l'environnement. - Le droit à des communautés et une biosphère durables.	- Les États membres devraient veiller à ce que les acteurs du développement et du déploiement de systèmes d'IA prennent des mesures suffisantes pour réduire le plus possible tout préjudice physique ou mental pour les individus, la société et l'environnement. - Les États membres devraient s'assurer qu'il existe des exigences suffisantes en matière de sécurité, de sûreté et de robustesse (dès la conception) et que les acteurs du développement et du déploiement de systèmes d'IA les respectent. - Les États membres devraient veiller à ce que les systèmes d'IA soient développés et utilisés de manière durable, dans le plein respect des normes de protection environnementale qui s'appliquent.
NON-DISCRIMINATION, ÉGALITÉ ENTRE LES FEMMES ET LES HOMMES, ÉQUITÉ ET DIVERSITÉ	- Le droit à la non-discrimination **(sur la base des motifs protégés énoncés à l'article 14 de la CEDH et dans le Protocole no 12 à la CEDH)**, y compris la discrimination intersectionnelle. - Le droit à la non-discrimination et le droit à l'égalité de traitement. - Les systèmes d'IA peuvent également donner lieu à une classification injuste fondée sur de nouveaux types de différenciation qui ne sont pas traditionnellement protégés. - Ce droit doit être garanti tout au long du cycle de vie des systèmes d'IA (conception, développement, mise en œuvre et utilisation) et dans le cadre des choix humains relatifs à la conception, à l'adoption et à l'utilisation de l'IA, que l'utilisation relève du secteur public ou du secteur privé.	- Les États membres sont tenus de veiller à ce que les systèmes d'IA qu'ils déploient n'entraînent pas de discrimination illégale, de stéréotypes nuisibles (y compris, mais sans s'y limiter, les stéréotypes de genre) ni d'inégalités sociales au sens large, et doivent donc appliquer le plus haut niveau de contrôle lorsqu'ils utilisent ou encouragent l'utilisation de systèmes d'IA dans des domaines sensibles de l'action publique, notamment, mais sans s'y limiter, la police, la justice, l'asile et la migration, la santé, la sécurité sociale et l'emploi. - Les États membres devraient intégrer des exigences de non-discrimination et de promotion de l'égalité dans les processus de passation de marchés publics pour des systèmes d'IA, et veiller à ce que les systèmes d'IA fassent l'objet, avant leur mise en service, d'un audit indépendant pour déterminer s'ils ont des effets discriminatoires. - Les États membres devraient imposer des exigences visant à remédier efficacement aux effets discriminatoires que pourraient avoir les systèmes d'IA déployés par les secteurs public et privé et à protéger les personnes contre les conséquences négatives de ces systèmes. Ces exigences devraient être proportionnées aux risques encourus. - Les États membres devraient encourager la diversité et la parité femmes/hommes dans le milieu professionnel de l'IA ainsi que les remontées d'information périodiques d'un large éventail de parties prenantes. La sensibilisation au risque de discrimination, y compris les nouveaux types de différenciation, et de préjugés dans le contexte de l'IA devrait être encouragée.

	DROITS SUBSTANTIELS	OBLIGATIONS CLÉS
TRANSPARENCE ET EXPLICABILITÉ	- Le droit d'être rapidement informé qu'un système d'IA éclaire ou prend une décision produisant des effets juridiques ou ayant pareillement des effets importants sur la vie d'une personne **(Convention 108+)**. - Le droit de recevoir une explication utile sur le fonctionnement d'un tel système d'IA, la logique d'optimisation qu'il suit, le type de données qu'il utilise et ses répercussions sur les intérêts de la personne concernée, chaque fois qu'il produit des effets juridiques ou qu'il a pareillement des effets sur la vie de cette personne. L'explication doit être adaptée au contexte et fournie d'une manière qui soit utile et compréhensible pour la personne, pour lui permettre de protéger efficacement ses droits. - Le droit des utilisateurs d'un système d'IA à être assisté par un être humain lorsque le système est utilisé pour interagir avec des personnes, en particulier dans le contexte des services publics.	- Les utilisateurs devraient être clairement informés de leur droit à être assistés par un être humain chaque fois qu'est utilisé un système d'IA susceptible d'avoir une incidence sur leurs droits ou pareillement de les toucher de manière significative, en particulier dans le contexte des services publics, et de recevoir une information claire sur la manière de demander cette assistance. Les États membres devraient exiger que les acteurs du développement et du déploiement des systèmes d'IA assurent une communication suffisante. - Chaque fois que l'utilisation de systèmes d'IA risque de porter atteinte aux droits de l'homme, à la démocratie ou à l'État de droit, les États membres devraient imposer des exigences aux acteurs du développement et du déploiement de l'IA en matière de traçabilité et d'information. - Les États membres devraient rendre publiques et accessibles toutes les informations pertinentes sur les systèmes d'IA qui sont utilisés pour fournir des services publics (notamment leur fonctionnement, leurs méthodes d'optimisation, la logique sous-jacente et le type des données utilisées), tout en préservant les intérêts légitimes que sont notamment la sécurité publique et les droits de propriété intellectuelle, mais en veillant toutefois au plein respect des droits de l'homme.
PROTECTION DES DONNÉES ET DROIT AU RESPECT DE LA VIE PRIVÉE	- Le droit au respect de la vie privée et familiale et à la protection des données à caractère personnel **(art. 8 de la CEDH)**. - Le droit à l'intégrité physique, psychologique et morale au vu du profilage basé sur l'IA et de la reconnaissance des émotions/de la personnalité. - Tous les droits consacrés par la Convention 108 et sa version modernisée, en particulier en ce qui concerne le profilage basé sur l'IA et la géolocalisation.	- Les États membres doivent veiller à ce que le droit à la vie privée et à la protection des données soit garanti tout au long du cycle de vie des systèmes d'IA qu'ils déploient ou qui sont déployés par des acteurs privés. - Les États membres devraient prendre des mesures pour protéger efficacement les personnes contre la surveillance de masse fondée sur l'IA, par exemple grâce à la technologie de reconnaissance biométrique à distance ou à d'autres technologies de suivi utilisant l'IA. - Lors de l'acquisition ou de la mise en œuvre de systèmes d'IA, les États membres devraient évaluer et limiter tout effet négatif sur le droit à la vie privée et à la protection des données et, plus largement, sur le droit au respect de la vie privée et familiale. Il faut s'interroger en particulier sur la proportionnalité du caractère invasif du système par rapport à l'objectif légitime qu'il est censé atteindre et sur sa nécessité pour y parvenir. - Les États membres devraient mettre en place des garanties appropriées pour ce qui concerne les flux de données transfrontaliers afin de s'assurer que les règles de protection des données ne sont pas contournées.

	DROITS SUBSTANTIELS	OBLIGATIONS CLÉS
 **OBLIGATION DE RENDRE DES COMPTES ET RESPONSABILITÉ**	- Le droit à un recours effectif en cas de violation des droits et libertés **(art. 13 de la CEDH)**. - Ce droit devrait également inclure le droit à des recours effectifs et accessibles chaque fois que le développement ou l'utilisation de systèmes d'IA par des organismes publics ou privés cause un préjudice injuste ou viole les droits légalement protégés d'un individu.	- Les États membres doivent veiller à ce qu'il existe des recours effectifs dans les juridictions nationales respectives, y compris en matière de responsabilité civile et pénale, ainsi que des mécanismes de réparation accessibles pour les personnes dont les droits sont mis à mal par le développement ou l'utilisation d'applications d'IA. - Les États membres devraient mettre en place des mécanismes publics de contrôle des systèmes d'IA susceptibles d'enfreindre les normes juridiques dans le domaine des droits de l'homme, de la démocratie ou de l'État de droit. - Les États membres devraient veiller à ce que les responsables du développement et du déploiement de systèmes d'IA 1) identifient, étayent par des documents et signalent les effets négatifs que pourraient avoir ces systèmes sur les droits de l'homme, la démocratie et l'État de droit, et 2) mettent en place des mesures d'atténuation adéquates pour qu'en cas de préjudice, les responsabilités soient établies et les auteurs rendent des comptes. - Les États membres devraient prendre des dispositions garantissant que les pouvoirs publics demeurent en mesure de vérifier les systèmes d'IA utilisés par des acteurs privés, afin d'évaluer leur conformité avec la législation en vigueur et de rendre les acteurs privés comptables de leurs actes.
 **DÉMOCRATIE**	- Le droit à la liberté d'expression et à la liberté d'association et de réunion **(art. 10 et 11 de la CEDH)**. - Le droit de vote et d'éligibilité, le droit à des élections libres et équitables, et en particulier le suffrage universel, égal et libre, y compris l'égalité des chances et la liberté des électeurs de se forger une opinion. À cet égard, les personnes ne doivent être soumises à aucune tromperie ou manipulation. - Le droit à une information (diversifiée), au libre discours et à l'accès à la pluralité des idées et des opinions. - Le droit à une bonne gouvernance.	- Les États membres devraient prendre des mesures appropriées pour lutter contre l'utilisation ou le dévoiement de systèmes d'IA à des fins d'ingérence illégale dans les processus électoraux, de ciblage politique personnalisé sans qu'il existe de mécanismes adéquats de transparence, de définition des responsabilités et d'obligation de rendre des comptes, ou plus généralement d'orientation des comportements politiques des électeurs ou de manipulation de l'opinion publique. - Les États membres devraient adopter des stratégies et mettre en place des mesures pour lutter contre la désinformation et repérer les discours de haine en ligne afin de garantir une pluralité informationnelle équitable. - Les États membres devraient soumettre leurs procédures de passation de marchés publics à des exigences juridiquement contraignantes qui garantissent l'utilisation responsable de l'IA dans le secteur public en veillant au respect des principes susmentionnés, notamment la transparence, l'équité, la responsabilité et l'obligation de rendre des comptes. - Les États membres devraient mettre en place des mesures propres à renforcer la maîtrise de l'outil numérique et les compétences numériques dans toutes les catégories de la population. Leurs programmes d'enseignement devraient s'adapter pour promouvoir une culture de l'innovation responsable qui respecte les droits de l'homme, la démocratie et l'État de droit.

ÉTAT DE DROIT

DROITS SUBSTANTIELS

- Le droit à un procès équitable et une procédure régulière (art. 6 de la CEDH), y compris la possibilité d'obtenir des informations utiles sur les décisions fondées sur l'IA et de contester ces décisions dans le cadre de l'application de la loi ou de la justice, notamment le droit à un réexamen de ces décisions par un être humain.
- Le droit à l'indépendance et à l'impartialité de la justice, et le droit à l'assistance d'un avocat.
- Le droit à un recours effectif **(art. 13 de la CEDH)** également en cas de préjudice illégal ou de violations des droits de l'homme d'un justiciable dans le contexte des systèmes d'IA.

OBLIGATIONS CLÉS

- Les États membres doivent veiller à ce que les systèmes d'IA utilisés dans le domaine de la justice et de l'application de la loi soient conformes aux exigences fondamentales du droit à un procès équitable. À cette fin, ils devraient garantir la qualité et la sécurité des décisions de justice et des données judiciaires, ainsi que la transparence, l'impartialité et l'équité des méthodes de traitement des données. Des garanties d'accessibilité et d'explicabilité des méthodes de traitement des données, notamment la possibilité d'audits externes, devraient être mises en place dans ce but.
- Les États membres doivent veiller à ce que des voies de recours effectives soient disponibles et que des mécanismes de réparation accessibles soient mis en place pour les personnes qui ont subi une atteinte à leurs droits résultant du développement ou de l'utilisation de systèmes d'IA dans des contextes liés à l'État de droit.
- Les États membres devraient fournir aux personnes des informations utiles sur l'utilisation de systèmes d'IA dans le secteur public chaque fois que cela peut avoir des conséquences significatives sur leur vie. Ces informations devraient être fournies en particulier lorsque des systèmes d'IA sont utilisés dans le domaine de la justice et de la répression, tant en ce qui concerne le rôle des systèmes d'IA dans le cadre de la procédure judiciaire que la possibilité de contester les décisions éclairées ou prises de cette façon.
- Les États membres devraient veiller à ce que l'utilisation des systèmes d'IA n'interfère pas avec le pouvoir décisionnel des juges ou l'indépendance judiciaire et à ce que toute décision judiciaire soit soumise à un contrôle humain utile.

CONSIDÉRATIONS ADDITIONNELLES TOUCHANT AUX PRINCIPES, AUX DROITS ET AUX OBLIGATIONS

Lorsqu'il est envisagé d'introduire de nouveaux droits et de nouvelles obligations dans un futur cadre juridique sur les systèmes d'IA fondé sur des principes, d'autres facteurs devraient être pris en considération. Tout d'abord, ces droits et obligations devraient être nécessaires, utiles et proportionnés à l'objectif qui est de protéger les citoyens contre les effets négatifs des systèmes d'IA sur les êtres humains, la démocratie et l'État de droit, et les bénéfices que procurent ces systèmes devraient être justement et équitablement répartis. Ces réflexions sur les risques et bénéfices devraient englober tous les aspects de la question et tenir compte de l'équilibre des intérêts légitimes qui sont en jeu. Toute approche fondée sur les risques et prenant en considération les bénéfices devrait en outre faire la distinction entre les différents niveaux de risque et cet aspect devrait être pris en compte au moment de l'élaboration et de l'adoption des mesures réglementaires.

Principaux éléments d'une approche fondée sur les risques et tenant compte des bénéfices:

- Examiner le **contexte d'utilisation** et les **effets potentiels** de la technologie d'IA
- Examiner le **domaine d'application** et les **parties prenantes concernées**
- **Évaluer et réexaminer les risques périodiquement** et **systématiquement**, et **adapter toutes mesures d'atténuation** à ces risques
- **Optimiser les bénéfices sociétaux** de l'innovation en matière d'IA en **ciblant les mesures réglementaires** selon une approche fondée sur ces risques

En ce qui concerne les obligations et les exigences, les autorités nationales devraient jouer un rôle central en évaluant systématiquement leur législation pour vérifier qu'elle est conforme avec les principes et priorités de la mise en conformité de la conception et de l'utilisation de l'IA avec les droits de l'homme, la démocratie et l'État de droit, et pour repérer les éventuels vides juridiques. De plus, des mécanismes nationaux d'audit et de contrôle des systèmes d'IA devraient prévenir les cas préjudiciables de non-conformité. Enfin, étant donné que les infrastructures numériques critiques du secteur public qui ont une incidence sur l'intérêt général sont de plus en plus souvent fournies par des acteurs privés, ces derniers ont le devoir de mettre la conception, le développement et le déploiement de leurs technologies en conformité avec ces principes et priorités.

06 PANORAMA DES INSTRUMENTS JURIDIQUES

CADRES JURIDIQUES INTERNATIONAUX

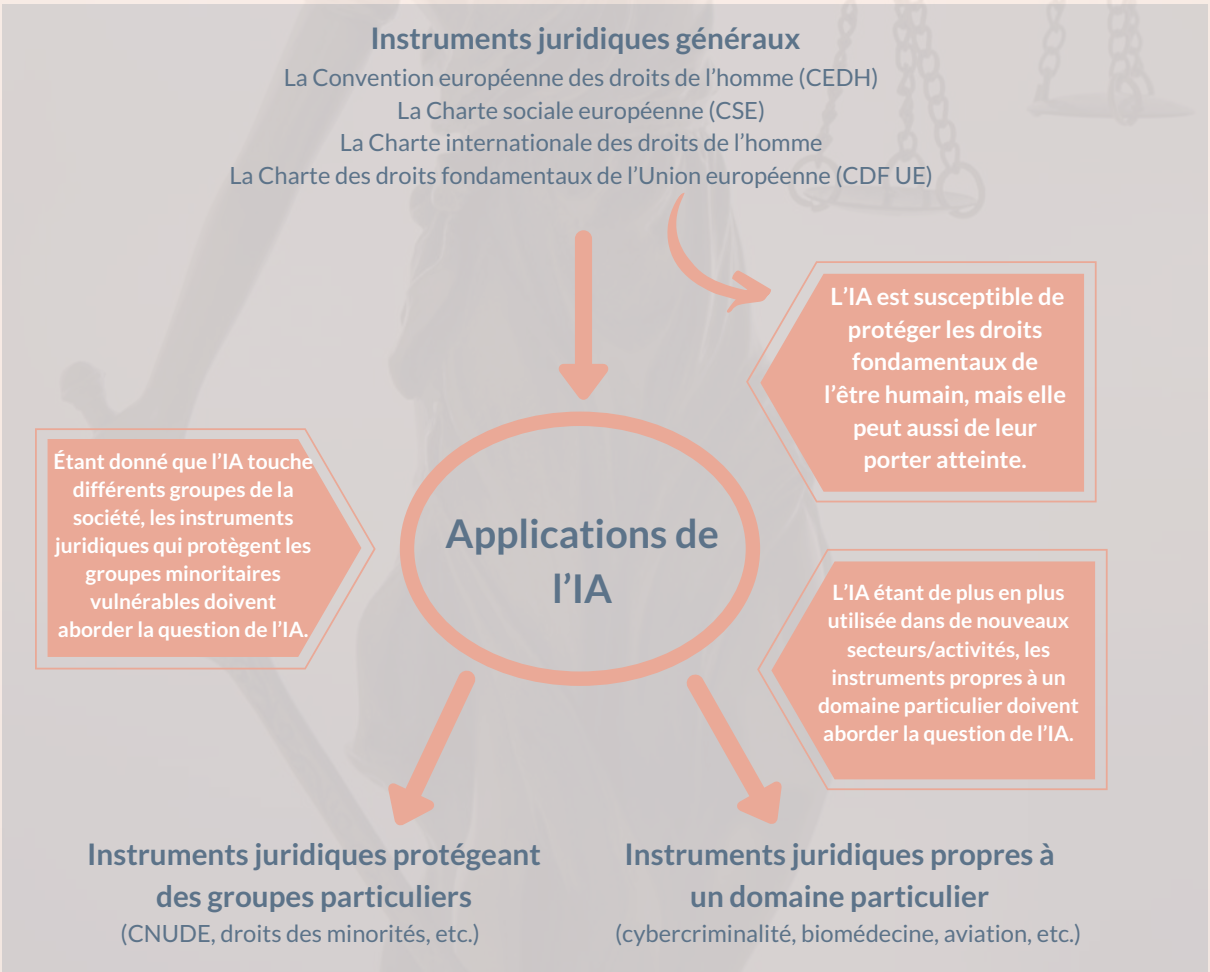
À ce jour, il n'existe pas de législation internationale portant expressément sur l'IA – ou la prise de décision automatisée –, mais plusieurs cadres juridiques s'appliquent en la matière. Citons en particulier (voir récapitulatif supra):

- La Convention européenne des droits de l'homme (CEDH)
- La Charte sociale européenne (CSE)
- La Charte internationale des droits de l'homme
- La Charte des droits fondamentaux de l'Union européenne (CDF UE).

Ces instruments juridiques énoncent les droits fondamentaux des êtres humains, dont beaucoup s'appliquent aux utilisations de l'IA, notamment le droit à la non-discrimination et le droit à la vie privée.

De même, un certain nombre d'instruments juridiques énoncent les droits des personnes en lien avec des activités et/ou secteurs particuliers, notamment la cybercriminalité, la biomédecine et l'aviation. L'IA étant de plus en plus utilisée dans des secteurs variés et selon des modalités qui touchent des pans toujours plus nombreux de nos vies, elle revêt une importance croissante dans chacune de ces branches du droit.

Les instruments juridiques qui protègent les groupes vulnérables ou minoritaires sont aussi concernés par l'IA. De ce fait, même s'il n'existe aucun instrument spécifique en matière d'IA, un nombre de plus en plus important de mécanismes juridiques en vigueur sont pertinents au regard des modalités de développement et de déploiement de cette technologie.



APPROCHES ACTUELLES DE DROIT SOUPLE

À l'heure actuelle, les principales approches en matière de gouvernance ou de réglementation de l'IA relèvent du « droit souple ». L'encadré ci-dessous présente les différences entre droit dur et droit souple.

Ces dernières années ont été marquées par une prolifération de lignes directrices et de principes pour une pratique éthique de l'IA. Ces principes et lignes directrices, élaborés par des organisations privées, publiques et universitaires, servent généralement à montrer que les processus de développement et de déploiement de l'IA sont dignes de confiance. Très souvent, l'élaboration de lignes directrices internes et de bonnes pratiques est vue comme un moyen de soutenir l'idée qu'en matière d'IA, il n'est pas nécessaire de légiférer ni de réglementer davantage au niveau central. Ainsi, bon nombre d'organisations qui ont proposé des principes ou des lignes directrices pour une IA éthique militent ardemment en faveur de l'autoréglementation.

Les codes de conduite adoptés volontairement au sein des organisations utilisant l'IA peuvent jouer un rôle important dans le développement d'une culture organisationnelle et retentir notablement sur les pratiques. De plus, ils ont pour avantages la flexibilité, l'adaptabilité, l'immédiateté de la mise en œuvre, une force d'attraction plus large et la possibilité d'être révisés et modifiés rapidement. Cela dit, certains les critiquent au motif qu'ils seraient symboliques et purement formels.

On observe une certaine cohérence dans les principes mis en avant par les ensembles de lignes directrices existants. Ainsi la transparence est-elle couramment mentionnée. Les recommandations pratiques en revanche manquent de cohérence, et de ce fait, les approches adoptées sont très différentes et les exigences éthiques et modalités de réglementation de l'IA diversement comprises. De plus, si les codes de bonne pratique et les lignes directrices pour une IA éthique sont pléthores, leur application pèche en général par un manque de transparence et de responsabilisation. La mise en œuvre de ces règles par des comités internes ou des commissions de contrôle a été critiquée pour son manque de transparence ou d'efficacité.

Il y a donc d'excellentes raisons d'associer approches volontaires de droit souple et gouvernance contraignante.

Droit dur

- Instruments juridiquement contraignants
- Sanctions prédéterminées
- Opposable par contentieux ou procédures judiciaires

Droit souple

- Recommandations, lignes directrices, certifications ou déclarations qui consolident des principes courants et des normes de bonnes pratiques
- Laisse souvent une marge à l'interprétation
- Pas de sanctions juridiques

INSTRUMENTS JURIDIQUES NATIONAUX

Partout dans le monde, l'élaboration de stratégies pour régir ou réglementer l'IA suscite un intérêt croissant. Les approches de droit souple prédominent. Une consultation menée auprès de membres du CAHAI a donné les résultats suivants:

- 30 États membres et 4 États observateurs disposent de stratégies et de politiques concernant les systèmes d'IA ;
- 1 État membre a lancé un programme de certification volontaire en matière d'IA ;
- 2 États membres ont officiellement adopté des cadres éthiques internationaux ou européens non contraignants sur l'IA ;
- 12 États membres et 4 États observateurs ont adopté un ou plusieurs instruments.

Ces projets ont été menés par diverses institutions, notamment des conseils nationaux, des comités, des institutions publiques spécialisées en IA et des organismes gouvernementaux.

En ce qui concerne le droit dur, la consultation des membres du CAHAI a fait ressortir que:

- 4 États membres ont adopté des cadres juridiques spécifiques sur l'IA concernant l'essai et l'utilisation des véhicules autonomes (voitures sans chauffeur) ;
- 2 États membres sont en train d'élaborer des cadres juridiques sur l'utilisation de l'IA dans les domaines du recrutement et de la prise de décision automatisée dans l'administration.

LE RÔLE DES ACTEURS PRIVÉS

Les acteurs privés (entreprises par exemple) contribuent largement à façonner le domaine de l'éthique de l'IA, notamment en créant et en adoptant des codes de conduite non contraignants. En outre, certains se disent favorables à la mise en place d'un cadre réglementaire pour assurer une plus grande sécurité juridique dans le domaine de l'IA.

Incontestablement, les acteurs privés ont un rôle important à jouer. Les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme énoncent le devoir qui incombe aux acteurs privés de respecter les droits de l'homme sur l'ensemble de leurs activités, produits et services.

Si une nouvelle approche réglementaire est adoptée, la participation et la coopération des acteurs privés seront primordiales pour l'élaboration d'un droit souple sectoriel, dont le rôle important sera de compléter et d'étayer la mise en œuvre du droit dur via des règles propres aux différents contextes (lignes directrices sectorielles, programmes de certification, etc.).

Un cadre réglementaire efficace en matière d'IA doit refléter la diversité des intérêts et des perspectives et nécessite donc la coopération étroite de toutes les parties prenantes : États, institutions publiques, société civile, entreprises, etc.

LIMITATIONS ACTUELLES

Bon nombre des instruments juridiques actuellement utilisés pour réglementer les différents aspects de l'IA ont été élaborés avant que les systèmes d'IA ne se banalisent. Ils ne sont donc peut-être pas adaptés aux divers effets et risques de l'IA.

Les approches de droit souple sont non contraignantes et reposent sur la base d'un libre respect des règles. Elles peuvent donc mener à des pratiques et des résultats divers et variés. De plus, la diversité des approches adoptées par les organisations qui appliquent des règles de droit souple peut aboutir à des engagements superficiels et de pure forme en matière d'éthique de l'IA. Cela étant, de nombreux travaux en cours dans le domaine des normes et de la certification pourraient contribuer à l'élaboration de futures lois.

Par ailleurs, certains principes importants en matière de gouvernance de l'IA ne sont pas protégés par la loi, notamment la nécessité de garantir un niveau suffisant de contrôle et de supervision par les humains, et la transparence et l'explicabilité effectives des systèmes d'IA. Il manque des instruments juridiques pour traiter ces importants facteurs technologiquement spécifiques de l'IA.

Si les mécanismes juridiques actuels protègent dans une certaine mesure les droits individuels, les dimensions sociétales des dangers de l'IA ne sont pas suffisamment prises en compte (risques pour les processus électoraux ou les institutions démocratiques par exemple). La protection de la démocratie et de l'État de droit nécessite un contrôle public et l'intervention de la sphère publique dans la conception, le développement et l'utilisation responsables des systèmes d'IA.

Enfin, les lacunes réglementaires actuelles créent de l'insécurité et de l'ambiguïté dans le champ de l'IA, ce qui est problématique pour ceux qui conçoivent, mettent en œuvre ou utilisent cette technologie et pour la société dans son ensemble. Cette insécurité risque d'empêcher les effets positifs de l'innovation en matière d'IA et de faire obstacle à d'importantes avancées qui auraient pu bénéficier aux citoyens et aux collectivités dans lesquelles ils vivent.

BESOINS ET OPPORTUNITÉS FUTURS

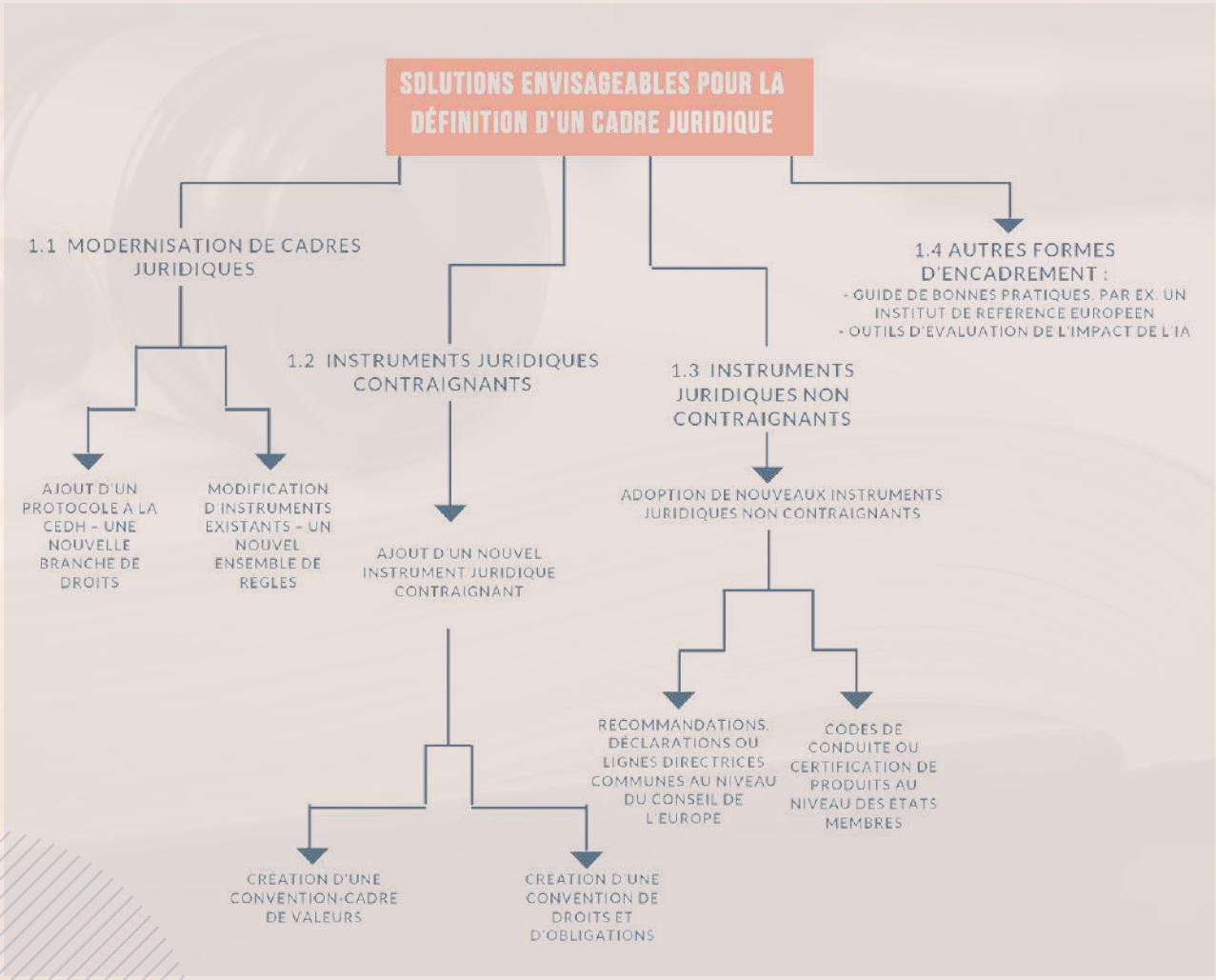
Il conviendrait que les futures approches en matière de réglementation remédient aux limitations mentionnées ci-dessus. Elles devraient transcender les secteurs et contenir des dispositions contraignantes permettant de sauvegarder les droits de l'homme, la démocratie et l'État de droit et d'assurer un niveau de protection plus complet.Ces approches pourraient venir en complément des règles sectorielles existantes.

L'élaboration d'un instrument juridiquement contraignant reposant sur les normes du Conseil de l'Europe – si telle est la solution retenue par le Comité des Ministres – contribuerait à faire de l'initiative du Conseil de l'Europe une entreprise unique en son genre, qui se distinguerait des autres initiatives internationales, qui, soit visent à élaborer un autre type d'instrument, soit diffèrent en termes de portée et de contexte.

SOLUTIONS ENVISAGEABLES POUR LA DÉFINITION D'UN CADRE JURIDIQUE

Pour établir des règles en matière d'IA afin de protéger les droits de l'homme, la démocratie et l'État de droit, le Conseil de l'Europe peut appliquer diverses stratégies, chacune ayant ses avantages et ses inconvénients en termes de résultats attendus.

Deux grandes alternatives sont à examiner. D'une part, il faut faire la distinction entre les instruments juridiques contraignants et non contraignants. De ce choix découlera l'obligation ou non pour les États d'appliquer les règles adoptées par le Conseil de l'Europe. D'autre part, il faut décider dans quelle mesure des instruments existants seront renforcés et modernisés et dans quelle mesure de nouveaux seront créés. Le graphique ci-dessous illustre ces différentes approches et indique où trouver un complément d'information dans la présente section.



1.1: Modernisation des instruments juridiques contraignants en vigueur

L'une des solutions à l'étude consiste à modifier les règles existantes pour les rendre applicables à l'IA. Il s'agira par exemple d'ajouter un *protocole* (ensemble de droits) à la Convention européenne des droits de l'homme. Un protocole additionnel traduirait l'engagement ferme des États membres à protéger les droits de l'homme, la démocratie et l'État de droit dans le contexte de l'IA, mais ne permettrait pas, en soi, de prescrire des exigences ou des normes spécifiques à ce domaine. De plus, les protocoles additionnels ne lient que les États qui les ont ratifiés, d'où le risque d'un contrôle fragmentaire. La Cour européenne des droits de l'homme est par ailleurs déjà surchargée.

Autrement, le Conseil de l'Europe pourrait décider de modifier des *instruments* existants (ensembles de règles) pour prendre en compte les questions soulevées par l'IA. Deux instruments existants pourraient être modifiés de cette manière : la Convention de Budapest sur la cybercriminalité et la Convention 108+, qui protège les personnes contre le traitement des données à caractère personnel. L'avantage de cette solution est qu'il existe déjà des mécanismes de suivi et d'application de ces règles, qui sont opérationnels. L'inconvénient est qu'il serait difficile d'adapter les instruments existants pour qu'ils répondent entièrement aux besoins. Les défis posés par la cybercriminalité et la protection des données sont liés, mais pas identiques à ceux soulevés par l'IA, notamment l'explicabilité des systèmes automatisés et l'obligation de rendre des comptes à leurs utilisateurs.

Enfin, ces deux solutions pourraient être *combinées* afin de remédier à leurs inconvénients respectifs. L'ajout d'un protocole permettrait de définir des valeurs et des principes généraux, tandis que la modification d'instruments existants permettrait de préciser les obligations auxquelles les États doivent se plier pour protéger ces principes dans la pratique, tout en veillant à ce qu'il existe des moyens suffisants pour contrôler le respect de ces obligations. Reste à savoir si une approche combinée serait suffisamment rapide et maniable par rapport au rythme de développement et d'adoption de l'IA.

1.2: Adoption d'un nouvel instrument juridique contraignant

Une autre solution consisterait à élaborer et à adopter un ensemble de règles contraignantes totalement nouvelles, spécialement pour l'IA. Cet ensemble pourrait prendre la forme d'une *convention* ou d'une *convention-cadre*. À l'instar de la distinction faite plus haut entre *protocoles* et *instruments*, une *convention-cadre* définit des principes généraux et des domaines d'action, tandis qu'une *convention* réglemente une question particulière de manière concrète en énonçant des droits et des obligations. Cela dit, elles sont toutes deux des traités et ont donc le même statut au regard du droit international. Nous nous proposons d'examiner ces deux possibilités l'une après l'autre.

Une *convention-cadre* permettrait de consacrer les grands principes et les valeurs fondamentales à respecter pendant la conception et le déploiement des systèmes d'IA, tout en laissant aux États une large marge d'appréciation dans la mise en pratique de ces principes et valeurs. Une fois la convention-cadre établie, les signataires pourraient décider d'élaborer des protocoles et des dispositions spécifiques plus détaillés. Cette approche se prêterait bien à l'essor rapide de l'IA et aux nouvelles questions éthiques que pose cette technologie. Une convention-cadre pourrait comprendre un ensemble de principes et de règles convenus applicable au développement de l'IA ainsi que des recommandations spécifiques sur la façon de contrôler leur application et de coopérer entre les pays. Des accords analogues sont déjà en place parmi les États membres du Conseil de l'Europe dans le cadre de la protection des minorités nationales et de la protection des personnes dans le contexte du traitement médical et de l'expérimentation médicale, ce qu'il convient de noter, car ces deux questions ne sont pas sans rapport avec les dangers potentiels des systèmes d'IA. Mais le plus souvent, les conventions-cadres s'en tiennent à définir des obligations générales pour

les États plutôt que des droits concrets pour les personnes, ce qui laisse aux États une certaine marge de manœuvre quant aux modalités d'application des principes qui y sont consacrés.

Les *conventions* permettent une réglementation plus poussée. Dans le cas de l'IA, une convention pourrait définir les droits et obligations permettant de sauvegarder les droits de l'homme, la démocratie et l'État de droit, et conférerait ainsi aux personnes une plus grande protection juridique. La solution de la convention permettrait d'encourager les États à prendre d'urgence des dispositions pour adopter des lois nationales, fixant ainsi des règles du jeu équitables pour le développement de produits d'IA fiables et responsables, même au-delà des frontières nationales.

Cela dit, une convention risque d'être trop rigide et de faire obstacle à de nouveaux usages de l'IA susceptibles de bénéficier à la société. Un ensemble concret de règles contraignantes au niveau international apporterait néanmoins une sécurité juridique à tous les acteurs de la chaîne, offrirait une puissante protection aux victimes de l'IA et jetterait les bases d'un développement de l'IA véritablement responsable.

Que le choix se porte sur une *convention-cadre* ou sur une *convention*, les destinataires de cet instrument (c'est-à-dire ceux à qui ces règles s'adressent principalement) seraient les États qui, en adoptant formellement l'instrument, accepteraient d'être liés par ses dispositions en vertu du droit international. Cela étant, le calendrier d'adoption d'une convention reste flou, et même les États ayant voté en faveur d'un tel instrument au Conseil de l'Europe ne seraient pas tenus de l'adopter formellement. De plus, il faudrait s'assurer que d'autres acteurs comme les pays non européens adoptent des règles équivalentes, sans quoi il y aurait un risque de fragmentation des règles et normes internationales applicables à l'IA.

1.3: *Instruments juridiques non contraignants*

Les instruments non contraignants ou dits de « droit souple » ne sont pas étayés par la force du droit international, mais peuvent néanmoins jouer un rôle en guidant les États et d'autres acteurs dans une direction favorable. Même si le droit souple ne peut, en soi, garantir que l'IA est tournée vers les droits de l'homme, la démocratie et l'État de droit, il peut y contribuer et présente en outre l'avantage d'être souple, adaptable et rapide à mettre en œuvre. Les instruments juridiques non contraignants peuvent être divisés en deux catégories, d'un côté ceux qui sont adoptés au niveau du Conseil de l'Europe, de l'autre, ceux qui doivent être approuvés par les États membres. Ces deux catégories ne s'excluent pas mutuellement. Examinons-les l'une après l'autre.

Un vaste instrument de droit souple au niveau du *Conseil de l'Europe* pourrait prendre la forme d'une recommandation ou d'une déclaration, soit en tant que document autonome, soit pour compléter l'un des instruments contraignants examinés plus haut. Une autre solution serait d'élaborer des guides ou des manuels contribuant à faire mieux connaître les incidences de l'IA sur les droits de l'homme, la démocratie et l'État de droit. Ces documents seraient élaborés avec l'ensemble des parties prenantes, notamment des représentants de l'État, le secteur privé, la société civile et les milieux universitaires. Ils ne seraient pas figés, mais évolueraient dans le temps pour tenir compte des innovations.

Au niveau des *États membres*, les instruments de droit souple pourraient prendre la forme de lignes directrices, de codes de conduite ou de labels, marques ou logos de certification pour les produits d'IA. Ces exemples de droit souple pourraient être intégrés aux pratiques de gouvernance, de passation des marchés et d'audit mises en place par les entreprises privées et autres organismes. Toutefois, si cette forme d'« autoréglementation » peut compléter d'autres principes et règles, elle ne saurait représenter ou remplacer les obligations faites aux États membres de s'employer activement à sauvegarder les droits de l'homme, la démocratie et l'État de droit.

1.4: Autres formes d'encadrement de l'IA

Outre les instruments juridiques contraignants et non contraignants, d'autres formes d'encadrement de l'IA pourraient être proposées aux États membres et autres acteurs. Il s'agira par exemple de définir de bonnes pratiques pour guider les actions positives. La création d'un « Institut européen d'évaluation comparative » pourrait être un bon moyen de définir et de faire largement accepter ce que devraient être les bonnes pratiques et ce qu'il convient de faire pour les encourager. De plus, la création d'un modèle ou d'un outil permettant d'évaluer l'impact de l'IA au niveau du Conseil de l'Europe contribuerait à harmoniser l'application des normes et valeurs relatives à l'IA sur l'ensemble du continent.

En résumé, toute solution visant à garantir effectivement que l'IA respecte la démocratie, les droits de l'homme et l'État de droit passera sans doute par une combinaison des approches horizontales (contraignantes et non contraignantes) décrites ici et de principes, normes et exigences plus sectoriels.

07 MÉCANISMES PRATIQUES À L'APPUI DU CADRE JURIDIQUE

Quels sont les mécanismes pratiques qui contribuent à l'efficacité du cadre juridique, assurent la conformité avec ce cadre et encouragent les bonnes pratiques ? Nous examinerons quelques réponses à cette question en nous intéressant au rôle de ces mécanismes et des différents acteurs concernés, puis en présentant quelques exemples de mécanismes destinés a) à assurer la conformité avec le cadre juridique et b) à soutenir les activités de suivi.

LE RÔLE DES MÉCANISMES DE CONFORMITÉ

Divers mécanismes pratiques ont été conçus pour contribuer et veiller au respect de la conformité. Citons notamment la due diligence en matière de droits de l'homme, les évaluations d'impact, la certification et la standardisation, l'audit et le suivi, mais aussi les bacs à sable réglementaires. Ces mécanismes favorisent la conformité avec le cadre juridique et apportent des avantages supplémentaires comme le renforcement de la transparence et de la confiance. Ils contribuent aussi à promouvoir les bonnes pratiques dans l'industrie et entre les secteurs, notamment par l'évaluation réflexive et anticipée des systèmes utilisant l'IA, depuis les premières étapes de la conception du projet jusqu'au suivi continu du système après son déploiement.

Le cadre juridique devrait fixer des exigences de haut niveau pour la conception de ces mécanismes.Il proposera par exemple de faire évoluer l'utilisation des mécanismes de conformité, parallèlement au développement et au déploiement d'un système donné, afin de tenir compte des éventuelles modifications dans son fonctionnement.

S'il est souhaitable que le cadre juridique fixe pour les mécanismes de conformité des exigences de conception fondées sur des principes, il demeure de la responsabilité des États membres de mettre en œuvre ces mécanismes en respectant les rôles des institutions et les habitudes de réglementation du pays.

De la conformité à l'assurance

Les mécanismes pratiques peuvent aussi être utilisés pour apporter une *assurance* aux opérateurs ou utilisateurs concernés et pour promouvoir de bonnes pratiques. Le cadre conceptuel proposé ici considère que les mécanismes pratiques ne se limitent pas à un *pur contrôle* de *conformité*, mais contribuent aussi à promouvoir un écosystème d'assurance qui présente de multiples avantages, entre autres:

- faciliter la **réflexion** et la **délibération** internes, en proposant des solutions concrètes pour évaluer la conception, le développement et le déploiement des systèmes ou produits utilisant l'IA, au moyen d'une approche dynamique qui évolue en même temps que le système (surveillance des changements de comportement du système après le déploiement par exemple) ;
- favoriser une **communication transparente** entre les développeurs, les organismes de contrôle, les opérateurs et les utilisateurs, et plus largement les parties prenantes ;
- faciliter les processus de **documentation** (ou de reporting) pour **veiller à la mise en œuvre des responsabilités** (audits, etc.) ;
- renforcer la **confiance** en encourageant et en adoptant de bonnes pratiques (normes ou programme de certification par exemple).

LE RÔLE DES DIFFÉRENTS ACTEURS

Sur un plan général, les trois catégories suivantes permettent d'identifier les acteurs qui contribuent, chacun de façon complémentaire, à assurer la conformité réglementaire au niveau national.

ACTEUR	RÔLE DE L'ACTEUR
Contrôleurs de systèmes	Il serait souhaitable que des organismes de contrôle indépendants comme des comités d'experts, des régulateurs sectoriels ou des auditeurs privés représentent des groupes de parties prenantes clairement identifiés et concernés par les applications pratiques de l'IA et soient comptables devant eux. Cela étant, leur champ d'action ne saurait couvrir tous les produits et systèmes reposant sur l'IA.
Développeurs de systèmes	Les développeurs du secteur public et du secteur privé peuvent contribuer au respect de la conformité en adoptant des politiques qui permettent de mieux savoir où ces technologies sont déployées (publication des contrats passés par le secteur public, constitution de registres publics, mise en place de systèmes de notification, etc.). Les outils standard d'audit interne et d'autocertification, quoique limités, peuvent aider.
Opérateurs et utilisateurs de systèmes	Les opérateurs et utilisateurs avisés de l'IA créent de la demande et peuvent utiliser ce pouvoir d'achat pour inciter les fournisseurs et vendeurs d'applications d'IA à se conformer au futur cadre juridique. Cela vaut en particulier pour le secteur public, qui a un pouvoir d'achat important.

Il est à noter que bon nombre des systèmes d'IA ainsi que les flux de données sur lesquels ils s'appuient sont déployés dans de multiples juridictions. Il faut donc s'assurer qu'il existe des mécanismes appropriés de partage d'information et d'établissement de rapports pour faciliter la réalisation des différentes tâches des acteurs concernés.

EXEMPLES DE TYPES DE MÉCANISMES DE CONFORMITÉ

Il existe toutes sortes de mécanismes de conformité. Certains donneront de meilleurs résultats dans tels contextes (selon les habitudes en matière de réglementation notamment) et en fonction des composants du système d'IA qui sont subordonnés aux impératifs de conformité (caractéristiques des données d'entraînement par exemple). Pour déterminer les mécanismes les mieux adaptés à chaque contexte, il convient d'associer les parties prenantes concernées selon un processus inclusif et participatif.

Les mécanismes pratiques qui donnent de bons résultats présentent des caractéristiques communes, lesquelles pourraient figurer dans un cadre juridique sous forme de principes à respecter. Exemples de principes:

- **Une évaluation dynamique (et non statique)** devrait être menée au début et tout au long du cycle de vie du projet pour expliquer les décisions qui sont prises à tout moment ;
- Les mécanismes devraient **évoluer avec la technologie** afin de soutenir les efforts de durabilité ;
- Les processus et produits des mécanismes devraient être **accessibles à différentes catégories de personnes** et **compréhensibles** par les spécialistes comme par les non-spécialistes afin de faciliter les recours et les réparations ;
- Le contrôle exercé par l'organisme ou la partie approprié (auditeur par exemple) devrait être **indépendant** ;
- Les normes techniques, certifications et pratiques **scientifiquement fondées** devraient être mises en avant et utilisées.

Les mécanismes présentés ci-dessous constituent une boîte à outils qui respecte un grand nombre de ces principes tout en offrant des possibilités d'amélioration et d'innovation réglementaire.

Due diligenceen matière de droits de l'homme

Pour que la conception, le développement et le déploiement des systèmes d'IA ne portent pas atteinte aux droits de l'homme, il est essentiel que les organismes fassent preuve de diligence raisonnable en la matière. Les analyses d'impact sont l'un des moyens pratiques de détecter, prévenir, atténuer et expliquer les éventuelles atteintes aux droits de l'homme résultant de l'utilisation de systèmes fondés sur l'IA.L'efficacité de ces analyses dépend des indicateurs socioéconomiques utilisés et des données collectées.Ainsi, telle analyse d'impact examinera les effets d'un système donné sur le bien-être des personnes, la santé publique, les libertés, l'accessibilité de l'information, les inégalités socioéconomiques, la durabilité environnementale, etc.

Certification et Labellisation de la qualité

Les normes et les dispositifs de certification, qui sont couramment utilisés comme des indicateurs de sécurité et de qualité, pourraient être étendus aux systèmes utilisant l'IA (par exemple pour certifier qu'un système donné a été soumis à une évaluation et à des essais poussés sur la base de normes industrielles). Ces dispositifs pourraient s'appliquer aux produits et aux systèmes eux-mêmes, mais aussi aux organismes chargés de les concevoir.

Audit

Des audits périodiques réalisés par des organismes spécialisés et indépendants, chargés de contrôler un secteur donné (la santé par exemple) ou un domaine particulier (véhicules autonomes, etc.), peuvent faciliter le passage à une utilisation plus transparente et plus responsable des systèmes utilisant l'IA.

Bacs à sable réglementaires

Les bacs à sables réglementaires offrent aux entreprises autorisées la possibilité de tester dans un environnement sûr et contrôlé (nommé « bacs à sable ») des produits ou systèmes utilisant l'IA et qui ne sont pas protégés par la réglementation en vigueur. Ces bacs à sables réglementaires peuvent contribuer à réduire le délai de commercialisation et les frais supportés par l'organisme et donc à soutenir l'innovation de façon contrôlée.

Surveillance continue et automatisée

Une fois déployés, les systèmes d'IA doivent faire l'objet d'un suivi permanent dont le but est de s'assurer qu'ils continuent de fonctionner comme prévu. Pour détecter et traiter au plus tôt tout écart de fonctionnalité, ce processus peut-être automatisé. Toutefois, le suivi automatisé n'est pas sans risque, car il peut entraîner une perte de contrôle par les humains et une baisse du niveau de compétence des professionnels chargés du contrôle de conformité.

La promotion et la prescription de mécanismes pratiques tels que ceux décrits ci-dessus devraient s'accompagner d'initiatives de soutien plus étendues afin de maximiser le potentiel de ces dispositifs. Ainsi, l'investissement dans l'éducation numérique au sein de la société ainsi que le développement des compétences et le renforcement des capacités parmi les développeurs, les décideurs et les régulateurs sont des préalables très importants qui contribuent à l'efficacité de tout cadre juridique. Les centres d'expertise sont particulièrement bien placés pour soutenir ces initiatives plus générales en facilitant les échanges permanents, la collaboration et la mise en commun de bonnes pratiques entre les différents acteurs au niveau national et international.

MÉCANISMES DE SUIVI

Outre les mécanismes décrits ci-dessus, il existe toutes sortes de dispositifs et mesures de suivi applicables à ce contexte. Citons notamment les groupes ou comités d'experts indépendants chargés de suivre la mise en œuvre et l'utilisation effective des instruments juridiques (convention par exemple) ou les effets de l'utilisation des systèmes d'IA sur la société. Comme indiqué précédemment, le caractère plurijuridictionnel des systèmes d'IA suppose souvent une coopération à l'échelle internationale. Parmi les mécanismes de suivi offrant cette possibilité on peut citer les réseaux créés entre les États parties pour stimuler l'entraide et la coopération en matière pénale ou civile.

08 CONCLUSION

Dans le présent Guide introductif, nous avons présenté les principaux éléments de l'*Étude de faisabilité* du CAHAI et fourni des informations générales sur les aspects techniques de l'IA et les imbrications entre cette technologie et les droits de l'homme, la démocratie et l'État de droit. Pris ensemble, ces éléments pourront, nous l'espérons, servir de tremplin à une réflexion pleinement pertinente sur les perspectives d'un cadre juridique – fondé sur des principes – régissant la recherche et l'innovation en matière d'IA, conformément à la mission du Conseil de l'Europe, qui est le gardien des libertés et droits fondamentaux, de la justice et des valeurs démocratiques. Pour mettre ces technologies de plus en plus puissantes et à fort pouvoir de transformation sur la bonne voie, tant pour les citoyens que pour la société en général, il faudra élaborer des politiques visionnaires et avisées et mener une réflexion préventive et rigoureuse. L'*Étude de faisabilité* et ce Guide introductif qui l'accompagne constituent les premières étapes dans cette direction.

Alors que les travaux du CAHAI entrent dans leur phase de consultation multipartite et de sensibilisation, il convient de souligner que la qualité et la réussite de cet important projet dépendent maintenant des connaissances et des éclairages apportés par un panel de participants qui doit être aussi large et inclusif que possible. Nous faisons appel à vous, lecteurs, à cette étape essentielle du projet, et cela se justifie pleinement.Cette orientation démocratique de la technologie et des politiques technologiques est au cœur même de l'approche centrée sur l'humain et guidée par les valeurs qui place les droits de l'homme, la démocratie et l'État de droit au premier rang des principes à suivre pour dessiner le futur de la gouvernance de l'IA et, plus largement, de l'innovation numérique. De fait, ce n'est qu'au travers de nombreux retours d'expérience et jugements critiques que les voix des personnes et des communautés concernées peuvent être dûment entendues et prises en compte. Et ce n'est que par la voie d'une consultation rigoureuse des parties prenantes que l'expérience vécue peut véritablement guider cette entreprise fondée sur la coopération, permettant ainsi de bâtir un écosystème technologique durable, garant de l'épanouissement de la société de demain.

09 ANNEXES

ANNEXE 1 : GLOSSAIRE

- **Algorithme (algorithm):** Un algorithme est une procédure ou une série d'étapes qui fournit des instructions sur la manière de traiter un ensemble d'entrées pour produire une sortie. Par exemple, une recette de cuisine peut être considérée comme un algorithme qui fournit des instructions pour produire une sortie (un gâteau) à partir d'un ensemble d'entrées (les ingrédients). Dans le cas de l'apprentissage automatique, l'algorithme consiste généralement en une série d'instructions qui donne l'ordre à un logiciel de prendre un ensemble de données (l'entrée) et d'en tirer un modèle ou de découvrir un schéma sous-jacent (la sortie).

- **Audit algorithmique (algorithmic audit)**: Il existe diverses approches de l'audit algorithmique, qui vont de l'évaluation ciblée d'un système selon un paramètre donné (par exemple, le niveau de biais) à une approche plus large visant à déterminer si le système est conforme à un ensemble de normes ou à un domaine réglementaire. Bien que généralement réalisés par des professionnels à des fins d'évaluation indépendante, les audits algorithmiques sont également utilisés par des journalistes, des universitaires et des militants pour augmenter le niveau de transparence et de responsabilité.

- **Biais d'automatisation (automation bias)**: Le biais d'automatisation désigne un phénomène psychologique qui peut se produire lorsque les opérateurs d'un système d'IA ne tiennent pas compte des résultats produits par le système, s'y conforment de manière excessive ou ne sont pas à même d'évaluer de façon appropriée la fiabilité de ses décisions et de ses résultats en raison de préjugés technologiques.L'utilisateur risque donc de a) se reposer sur le système de manière excessive et lui accorder une trop grande confiance, et, partant, ne pas repérer les prédictions ou classements inexacts, ou b) se méfier du système et le sous-utiliser, alors que celui-ci peut être plus performant que lui dans la réalisation de certaines tâches.

- **Décision automatisée (automated decision)**: On entend par décision automatisée le choix d'une action ou d'une recommandation au moyen de processus de calcul. Sont qualifiées d'automatisées les décisions qui complètent ou remplacent le travail décisionnel généralement effectué par des humains seuls. Plus couramment, les décisions automatisées sont des prédictions sur des personnes ou des situations du monde qui sont tirées d'une analyse par apprentissage automatique portant sur des données concernant des événements passés et leurs similitudes avec un ensemble donné de situations.

- **Données d'entraînement ou d'apprentissage/Données d'essai (training/testing data)**: Pour construire un modèle et s'assurer de son exactitude, on scinde généralement le jeu de données en deux sous-ensembles : les données d'entraînement et les données d'essai. Les données d'entraînement, injectées dans un algorithme, servent à élaborer le modèle initial. Puis, le modèle ainsi entraîné est testé avec les données restantes. Les données sont scindées de cette manière pour s'assurer que le modèle peut être généralisé à de nouvelles situations. En effet, les données collectées ne représentent qu'un petit échantillon de la population générale. Si toutes les données étaient utilisées pour entraîner le modèle, il y aurait un risque de sur-ajustement ; autrement dit, le modèle fonctionnerait bien avec le jeu de données d'origine, mais donnerait de mauvais résultats avec de nouvelles données. En outre, l'essai du modèle avec de « nouvelles » données

permet aux spécialistes des données de repérer un éventuel *sous-ajustement*. Ce phénomène apparaît lorsque la fonction de mise en correspondance du modèle s'adapte mal à la distribution des données et n'est donc pas en mesure de rendre compte précisément des schémas complexes qu'elle tente de classer ou de prédire.

- **Droit souple (soft law)**: Les droits souples désignent les dispositifs réglementaires qui imposent ou restreignent, sans avoir la force de sanctions ou de peines infligées par l'État. Les « bonnes pratiques » et les lignes directrices en matière d'éthique élaborées par les entreprises et les syndicats sont des exemples de droit souple. Dans certaines professions comme la pratique du droit ou de la santé, le droit souple désigne l'ensemble des pratiques éthiques nécessaires à la certification. Dans le domaine de la santé, la violation de l'éthique médicale peut aboutir au retrait du droit d'exercer la médecine. Ces règles de droit souple ont des effets punitifs variables sur ceux qui y sont subordonnés. Ainsi, l'Association of Computing Machinery (ACM) s'est dotée d'un « Code d'éthique et de conduite professionnelle » que ses membres sont censés respecter,mais il n'y a pas de sanction prévue ni de système d'arbitrage pour les membres de l'association qui violeraient le code. Le droit souple peut aussi désigner les dispositifs d'incitation figurant dans les politiques gouvernementales. Par exemple, les crédits d'impôt proposés aux producteurs de technologies « vertes » encouragent, sans les imposer, certains choix de production.

- **Égalité des armes (equality of arms)**: L'égalité des armes désigne l'obligation qu'un procès intenté contre une personne soit équitable. Dans la doctrine des droits de l'homme, l'égalité des armes est consacrée dans le droit à une défense adéquate, ce qui inclut le droit de faire appel à un avocat et de citer et contre-interroger un témoin. Lorsque des technologies sont utilisées dans la conduite de poursuites pénales, l'égalité des armes peut signifier qu'il faut pouvoir interpréter et contester les fonctions et les performances de ces technologies.

- **Équité (fairness):** L'équité peut être définie de multiples façons. Elle peut notamment indiquer dans quelle mesure un système d'IA favorise ou empêche l'atténuation des biais et l'élimination des facteurs discriminatoires qui influent sur ses résultats et ses mises en œuvre. Étant donné que le cycle de vie de l'IA, y compris la décision d'utiliser cette technologie, est impacté, à toutes les étapes, par des choix humains, l'équité de l'IA est déterminée en évaluant le biais humain et son incidence sur ce que fait l'IA et sur ceux qui bénéficient et ceux qui ne bénéficient pas de son utilisation. Dans le contexte de l'IA, garantir l'équité suppose de prêter attention aux données employées, à la conception globale du système, aux résultats de son utilisation et aux décisions concernant sa mise en œuvre.
    - *L'équité des données (data fairness)* signifie que les jeux de données utilisés par l'IA sont suffisamment représentatifs de la population qui est susceptible d'être concernée et qu'ils sont de qualité et pertinents, que les choix qui ont présidé à la collecte des données à l'origine ont été analysés pour y rechercher d'éventuels biais, et que les données peuvent être auditées.

- *L'équité de conception (design fairness)* signifie que les activités des concepteurs du système sont sérieuses, réfléchies et attentives au risque de préjugés de la part de l'équipe de développement.L'équité de conception suppose d'évaluer la formulation globale du problème et le résultat choisi, la sélection et la gestion des données, et le choix des paramètres, et d'évaluer également dans quelle mesure le système fournit des résultats analogues pour des personnes appartenant à des groupes différents et ayant des identités différentes. En résumé, les concepteurs doivent s'assurer que les systèmes qu'ils construisent ne contribuent pas à favoriser des situations sociales indésirables, notamment une discrimination préjudiciable, un épuisement des ressources ou des structures de pouvoir coercitives.

- *L'équité des résultats (outcome fairness)* indique dans quelle mesure les décisions ou autres résultats produits par l'IA sont équitables et justes et aboutissent à une juste répartition des droits, obligations et biens publics. L'équité des résultats est également une évaluation des valeurs défendues ou remises en cause par l'utilisation de l'IA.

- Il est aussi possible de décomposer l'équité d'une instance de l'IA selon les points de vue des parties qui ont une incidence sur son utilisation ou sont concernées par elle. Chaque instance de l'IA est associée à un ensemble différent et éventuellement variable de parties prenantes, dont les grandes catégories sont les personnes concernées, les organismes chargés de la mise en œuvre et les sociétés.

- Pour déterminer *l'équité perçue par la personne concernée* (subject fairness), il est possible de demander à la personne faisant l'objet d'une décision ou d'une action prise ou appuyée par un système d'IA si elle estime que le processus et le résultat sont justifiables et légitimes. Pour évaluer le caractère justifiable et légitime, la personne concernée peut avoir besoin de connaître les détails du processus ayant abouti à la décision ou à l'action en question, ainsi que les facteurs qui auraient pu conduire à un autre résultat (exemple : le rejet d'une candidature par un algorithme de recrutement peut être expliqué en montrant que le candidat ne possède pas une compétence ou un diplôme expressément mentionné). La personne concernée doit aussi avoir la possibilité d'introduire un recours lorsqu'elle est en désaccord avec le résultat (exemple : le candidat à un emploi peut fournir un complément d'information ou contester l'exactitude de l'algorithme de recrutement auprès d'un être humain habilité à modifier le résultat).

- *L'équité des responsables de la mise en œuvre* (implementer fairness) peut s'exprimer au moyen de mesures de l'obligation de rendre des comptes, notamment via des processus d'audit et d'évaluation. Les responsables de la mise en œuvre doivent veiller à ce que les systèmes d'IA soient transparents et explicables par ceux qui les utilisent et par ceux qui sont impactés par cette utilisation. Avant et pendant l'utilisation de l'IA, les responsables de la mise en œuvre doivent prendre en compte les effets sociaux, économiques et politiques, en étant attentifs non seulement aux avantages perçus de l'IA, mais aussi à la survenance et au risque de préjudices ainsi qu'aux victimes des éventuels préjudices. Par exemple, la mise en place d'un algorithme de détermination des peines peut se traduire par une grande cohérence judiciaire et/ou une rationalisation de la prise de décision. Toutefois, le même système peut aussi reproduire des décisions discriminatoires, par exemple, dans des pays à majorité blanche, l'imposition de peines plus longues aux personnes de couleur pour des chefs d'accusation analogues, en raison de certaines caractéristiques de la conception du système ou des données auquel il se réfère. Lorsque de tels conflits apparaissent, l'exactitude ou l'efficacité fonctionnelle du système (le cas échéant) doit être totalement remise en question ; la conception de l'algorithme et le modèle de données doivent être évalués avec soin, et une décision doit être prise quant à l'utilisation ou non du système.

- *L'équité sociétale* (societal fairness) est un sujet plus vaste. Les systèmes dont l'utilisation pourrait avoir des répercussions sur les droits et les privilèges des personnes ou des groupes et/ou sur l'orientation de la société requièrent une attention particulière de la part des êtres humains, et leur utilisation doit faire l'objet d'un examen délibératif ouvert. Les responsables de l'élaboration des politiques, les chercheurs et les militants ont pour rôle de proposer et d'examiner d'un œil critique des stratégies et des actions destinées à promouvoir le bien-être général et la justice sociale.Lorsque l'IA est utilisée dans le secteur privé ou dans le secteur public (ou dans les deux en raison d'un partenariat public-privé), elle peut contribuer à préserver ou à combattre des systèmes sociaux, économiques et politiques existants. De ce fait, le rôle de l'IA au sein de ces systèmes doit faire l'objet d'une évaluation ouverte et inclusive, et les humains qui participent à la conception et à la mise en œuvre de l'IA devraient être tenus d'expliquer leurs choix. En définitive, le recours à l'IA, comme à tout autre outil, n'est acceptable que s'il favorise l'amélioration des conditions de vie des êtres humains sans causer de dommages.

- **Ensemble de données/Jeu de données (dataset)**: Un ensemble ou jeu de données est un fichier d'informations qui se présente généralement sous la forme d'un recueil de mesures ou d'observations enregistrées dans un ensemble de lignes et de colonnes. Chaque ligne correspond à un individu ou à un objet qui peut être décrit au moyen d'un ensemble de valeurs enregistrées correspondant chacune à une caractéristique, cet ensemble étant représenté par la série de colonnes. Par exemple, le jeu de données suivant représente une série de mesures relatives aux patients d'un cabinet médical fictif, un numéro d'identification unique étant attribué à chaque patient:

N° patient	Âge (intervalle)	Poids (Kg)	Tension artérielle
1883652	26 > 30	71	115/75
1268833	31 > 40	null	139/83
1776436	65 > 70	90	170/90
1557821	41 > 50	72	131/82

  L'exemple ci-dessus ne montre que les quatre premiers patients, et trois paramètres seulement sont enregistrés. Les jeux de données médicales peuvent néanmoins être très volumineux, en nombre de patients, mais aussi en nombre de valeurs possibles enregistrées. Ici, aucune donnée de poids n'est enregistrée pour le patient no 1268833. Les données manquantes posent de grandes difficultés aux systèmes d'apprentissage automatique et peuvent influer sur l'exactitude du modèle qui est élaboré.

- **Explicabilité (explainability)**: L'explicabilité d'un système d'IA est étroitement liée à la transparence. Elle indique dans quelle mesure les processus et la logique qui sont à l'origine des résultats produits par le système peuvent être compris par des utilisateurs humains. L'explicabilité peut permettre de savoir dans quelle mesure les rouages internes du modèle peuvent être exprimés en langage simple afin d'améliorer la prise de décision et d'accroître la confiance.

- **Généralisabilité (generalisability)**: Un modèle est dit généralisable s'il est efficace pour un large éventail de données d'entrée reflétant le monde réel et de contextes opérationnels. Les modèles qui ne sont pas suffisamment entraînés sur des données représentatives affichent souvent une généralisabilité limitée lorsqu'ils sont mis en service dans le monde réel.

- **Modèle (model)**: On entend par modèle le résultat final de l'application d'un algorithme à un ensemble de données d'entrée (ou « variables ») dans le but d'obtenir une valeur de sortie de prédiction ou d'information.En règle générale, un modèle est une fonction de mise en correspondance (mathématique) formelle dont le but est de représenter les processus sous-jacents et les interactions entre ces processus, lesquels sont supposés faire apparaître une relation entre les données d'entrée observées et la sortie de l'algorithme. Par exemple, le modèle simple suivant pourrait exprimer la relation entre un ensemble de variables d'entrée, telles que la superficie d'un bien immobilier (x1), le nombre de chambres (x2) et l'âge du bien (x3), et la variable de sortie (y), qui représente le prix. Ici, les coefficients ou paramètres des variables x sont utilisés comme des facteurs de pondération qui indiquent l'importance relative de chacune des variables d'entrée en fonction de son influence sur y. Dans le cas d'espèce, la tâche de l'algorithme d'apprentissage serait de trouver les valeurs des paramètres qui permettent de prédire précisément le prix réel de la maison pour toutes les données d'entraînement. Le modèle ainsi défini pourrait ensuite être utilisé pour estimer le prix de maisons ne figurant pas dans le jeu de données d'origine.

- **Obligation de rendre des comptes (accountability)**: L'obligation de rendre des comptes peut être décomposée en deux éléments : l'obligation de s'expliquer (answerability) et l'auditabilité (auditability).L'obligation de s'expliquer désigne l'établissement d'une chaîne continue de responsabilités humaines sur la totalité du flux de travaux d'exécution du projet d'IA. Elle exige que des autorités humaines compétentes expliquent et justifient le contenu des décisions assistées par algorithme et des processus qui sous-tendent leur production dans un langage simple, compréhensible et cohérent. L'auditabilité répond à la question « comment les concepteurs et les personnes chargées de la mise en œuvre des systèmes d'IA doivent-ils répondre de leurs actes ? ». Cet aspect de l'obligation de rendre des comptes concerne la mise en évidence à la fois des responsabilités quant aux pratiques de conception et d'utilisation et du caractère justifiable des résultats.

- **Propriété intellectuelle (intellectual property)**: La propriété intellectuelle désigne la détention légale des produits d'un travail de création. Parmi les formes courantes de propriété intellectuelle figurent les droits d'auteur, les brevets, les marques déposées et les secrets commerciaux. Les droits d'auteur sont une forme de propriété intellectuelle qui protège le droit du créateur de tirer avantage de la paternité d'une œuvre originale comme un roman, une composition musicale ou un tableau. Un brevet est une autorisation exclusive, mais limitée dans le temps, de tirer avantage de l'invention et de la découverte de processus, machines, articles manufacturés ou compositions de matière nouveaux et utiles, par exemple de nouveaux médicaments et des technologies de voiture sans chauffeur. Une marque commerciale permet à une entité commerciale de réserver l'usage d'un mot, d'un nom, d'un symbole ou d'une figure, ou de toute combinaison de ces éléments, qui identifie ses produits et les distingue des produits de ses concurrents. On peut citer le nom « Twitter » et ses logos associés, qui identifient de manière unique une grande plate-forme de médias sociaux et la distinguent des autres plates-formes. Le secret commercial désigne toute information pouvant être utilisée dans les activités d'une entité commerciale ou de toute autre entreprise et qui est suffisamment précieuse et secrète pour procurer un avantage économique réel ou potentiel sur les concurrents, par exemple la recette du Coca-Cola.

- **Proportionnalité (proportionality):** La proportionnalité est un principe juridique qui fait référence à l'idée de fournir un résultat juste en employant des moyens qui sont proportionnés au coût, à la complexité et aux ressources disponibles. Dans le même ordre d'esprit, la proportionnalité peut aussi être utilisée comme une notion évaluative, comme dans le cas du principe de protection des données, qui énonce que seules sont recueillies les données à caractère personnel nécessaires et suffisantes aux fins de la tâche considérée.

- **Représentativité (representativeness):** Les données utilisées dans l'algorithme reflètent le monde réel. On peut donc se demander si l'échantillon choisi reflète les caractéristiques observées dans la population générale. Le fait que les plus grandes bases de données d'images soient créées par des personnes d'un petit nombre de pays est un bon exemple de non-représentativité. Ainsi, une recherche des « robes de mariée » dans une base de données d'images courante ne renverra peut-être pas les images figurant des tenues de mariage de nombreuses cultures non occidentales.

- **Système de décision automatisée (automated decision system)**: Un système de décision automatisée (SDA) complète ou remplace la prise de décision par un être humain en utilisant des processus de calcul pour produire des réponses à des questions, soit sous forme de catégories discrètes (oui/non, masculin/féminin/non binaire, malin/bénin, etc.) soit sous forme de notes (degré de solvabilité, risque de survenance d'un délit, prédiction de croissance d'une tumeur, etc.). La plupart des SDA produisent des prédictions concernant des personnes ou des situations au moyen de l'apprentissage automatique ou d'une autre logique computationnelle, en calculant la probabilité qu'une condition donnée soit remplie.

  Le plus souvent, les systèmes de décision automatisée sont « entraînés » sur des données du passé, dans lesquelles ils recherchent des schémas relationnels (par exemple, la relation entre les relevés d'un baromètre, la température ambiante et la chute de neige). La décision automatisée est prise en comparant des schémas connus avec des données d'entrée existantes pour évaluer leur degré de correspondance (exemple : prévision météorologique reposant sur la similitude entre les relevés météorologiques du jour et l'historique des relevés). Exemples de système SDA : algorithme calculant des cotes de solvabilité, systèmes de reconnaissance biométrique qui cherchent à identifier des individus à partir de leurs caractéristiques physiques comme les traits du visage.

- **Système sociotechnique (socio-technical system)**: On qualifie de sociotechnique un système qui établit un couplage entre le comportement humain (ou social) et le fonctionnement d'un système technique, donnant ainsi naissance à des fonctions nouvelles (et émergentes) qui ne sont pas réductibles aux seuls éléments humains ou techniques. En intervenant dans les attitudes et comportements humains ou leurs relations au monde, le système technique restructure le comportement humain. Dans la perspective sociotechnique, on considère les souhaits ou objectifs humains qu'une technologie réalise ou est censée réaliser.

  Par exemple, les systèmes de recommandation fondés sur l'IA que l'on trouve couramment dans les sites de vente au détail, les sites de vidéos et les réseaux sociaux sont sociotechniques, car leur but est d'inciter les usagers à adopter des comportements voulus par les opérateurs de ces plates-formes, comme l'allongement du temps passé sur le site et/ou l'achat de produits. Les algorithmes d'apprentissage automatique des sites de partage de vidéos analysent le comportement de milliers, voire de millions d'utilisateurs et leur recommandent des contenus en fonction de leur ressemblance avec tel ou tel sous-groupe d'utilisateurs. Ces sites sont dits sociotechniques, car

ils ont besoin d'informations sur les usagers et parce que l'objectif de leur analyse est de fidéliser ces derniers afin de générer des recettes publicitaires.

On peut aussi qualifier de sociotechniques les systèmes dont l'existence, la mise en œuvre ou les effets mêmes supposent des relations politiques, économiques ou sociales humaines. Par exemple, les systèmes de surveillance adoptés par les forces de l'ordre sont sociotechniques, car leur adoption et leur utilisation ont des dimensions politiques ; les personnes ciblées par la surveillance policière sont impactées plus que les autres par l'utilisation des technologies de surveillance en raison de choix passés faits par des agents de l'État ou de la force publique. De ce point de vue sociotechnique, les technologies de surveillance interviennent dans les relations entre les personnes et les centres de pouvoir de la société.

- **Transparence (transparency)**: La transparence des systèmes d'IA peut désigner plusieurs caractéristiques de leurs mécanismes et comportements internes et des systèmes et processus qui les sous-tendent. Un système d'IA est dit transparent lorsqu'il est possible de déterminer comment il a été conçu, développé et déployé. Cela suppose, entre autres, de disposer d'un enregistrement des données utilisées pour entraîner le système ou des paramètres du modèle chargé de transformer l'information d'entrée (une image par exemple) en une information de sortie (une description des objets contenus dans l'image). Toutefois, la transparence d'un système d'IA peut aussi concerner des processus plus généraux, par exemple la présence ou non d'obstacles juridiques empêchant les personnes d'accéder aux informations nécessaires pour comprendre pleinement comment le système fonctionne (restrictions en matière de propriété intellectuelle par exemple).

APPENDIX 2: TRAVAUX DU CONSEIL DE L'EUROPE ET AUTRES TRAVAUX AFFÉRENTS DANS LE DOMAINE DE L'IA ET SES DOMAINES CONNEXES : ÉTAT DES LIEUX

Les références figurant dans la présente annexe synthétisent et complètent le chapitre 4 de l'*Étude de faisabilité.* La numérotation des paragraphes correspond à celle de l'*Étude*.

4.1. Protection des données à caractère personnel

- Convention 108/108+ (1981/2018)
  - Le traitement des données sensibles n'est autorisé que s'il existe des lignes directrices appropriées.
  - Tout individu a le droit de connaître la finalité du traitement de ses données. De plus, il a un droit de rectification et d'information lorsque ses données sont traitées en violation de la convention.
  - Cette convention institue la transparence, la proportionnalité, l'obligation de rendre des comptes, les analyses d'impact et le respect de la vie privée dès la conception.
  - Les personnes ne doivent pas être soumises à des décisions prises uniquement sur le fondement d'un traitement automatisé des données, sans que leur point de vue soit pris en compte.
  - « Le cadre juridique construit autour de la convention reste pleinement applicable à la technologie d'IA, dès lors que les données traitées relèvent du champ de ce traité. »
  - Convention modernisée 108+ adoptée en 2018 ; les Lignes directrices sur « La protection des données personnelles des enfants dans un cadre éducatif » ont été adoptées par la Convention en novembre 2020:
    - Elles énoncent « les principes fondamentaux des droits de l'enfant dans le cadre éducatif et aident les législateurs et décideurs politiques, mais aussi les responsables de traitement des données et l'industrie à respecter ces droits. »

4.2. Cybercriminalité

- Convention sur la cybercriminalité (« Convention de Budapest ») (2001)
  - « L'incrimination des infractions commises contre des systèmes informatiques et au moyen de systèmes informatiques, et la mise en œuvre des pouvoirs de procédure pour mener des investigations en matière de cybercriminalité et protéger les éléments de preuve électroniques. »
    - Les crimes comprennent, entre autres, les atteintes à la propriété intellectuelle, la fraude informatique, la pornographie enfantine et les violations des réseaux de sécurité.
    - Les enquêtes comprennent une série de pouvoirs et de procédures, notamment en matière d'interception et de fouille des réseaux informatiques.
  - L'objectif principal est de « mener une politique pénale commune destinée à protéger la société de la cybercriminalité, notamment par une législation appropriée et la coopération internationale. »
  - Étant donné que les réseaux numériques ignorent les frontières, un effort international concerté s'impose pour faire face à l'utilisation abusive des technologies.
  - Trois objectifs de la Convention :
    - « Harmoniser les éléments des infractions ayant trait au droit pénal matériel national et les dispositions connexes en matière de cybercriminalité. »
    - « Fournir au droit pénal procédural national les pouvoirs nécessaires à l'instruction et à la poursuite d'infractions de ce type et d'autres infractions commises au moyen d'un système informatique ou dont des preuves se présentent sous forme électronique. »

        - « Mettre en place un régime rapide et efficace de coopération internationale. »

4.3. Travaux dans le domaine des systèmes algorithmiques

- Déclaration sur les capacités de manipulation des processus algorithmiques (2019)
    - Nombreux sont ceux qui n'ont pas conscience des dangers de l'exploitation des données.
    - Les dispositifs informatiques transforment des formes existantes de discrimination en classant les personnes en catégories.
    - Le Comité des Ministres attire l'attention sur « la menace grandissante qui émane des technologies numériques et qui remet en cause le droit des êtres humains à se forger une opinion et à prendre des décisions indépendamment des systèmes automatisés. »
    - Les principales menaces concernent le micro-ciblage, la détection des vulnérabilités et la reconfiguration des environnements sociaux.
    - Le Comité émet plusieurs recommandations pour remédier à ces menaces. Il recommande notamment d'étudier la nécessité de cadres protecteurs supplémentaires visant à lutter contre les effets de l'utilisation ciblée des technologies, de lancer des débats publics ouverts, éclairés et inclusifs sur la limite entre la persuasion admissible et la manipulation inacceptable, et de donner aux utilisateurs les moyens d'agir en sensibilisant le public et en encourageant la maîtrise des outils numériques.
- Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'homme (2020)
    - Il est recommandé aux États membres de revoir leurs cadres législatifs, leurs politiques et leurs propres pratiques afin de s'assurer que l'acquisition, la conception et le développement des systèmes algorithmiques ne sont pas contraires au cadre de protection des droits de l'homme.
    - « Les droits de l'homme qui sont souvent bafoués du fait de la dépendance à l'égard des systèmes algorithmiques comprennent notamment le droit à un procès équitable, le droit au respect de la vie privée et à la protection des données, le droit à la liberté de pensée, de conscience et de religion, le droit à la liberté d'expression, le droit à la liberté de réunion, le droit à l'égalité de traitement et les droits économiques et sociaux. »
    - De plus, il est recommandé aux États membres d'entreprendre des consultations régulières, inclusives et transparentes avec les parties prenantes concernées, en accordant une attention particulière aux voix des groupes vulnérables.
    - Cette Recommandation comprend diverses obligations des États à l'égard de la protection et de la promotion des droits de l'homme et des libertés fondamentales dans le contexte des systèmes algorithmiques. Ces obligations concernent notamment la législation, la transparence, l'obligation de rendre des comptes et les recours effectifs ainsi que les mesures de précaution.
- MSI-AUT – Responsabilité et IA : Étude sur les incidences des technologies numériques avancées (dont l'intelligence artificielle) sur la notion de responsabilité, sous l'angle des droits humains (2019)
    - Cette étude présente ce qu'est l'IA et comment fonctionnent les technologies appliquées à des tâches spécifiques. Elle décrit aussi les menaces et les dommages associés aux technologies numériques évoluées ainsi que plusieurs « modèles de responsabilité » applicables aux effets négatifs des systèmes d'IA.

- Principales recommandations de cette étude : mise en place de « mécanismes effectifs et légitimes visant à prévenir et empêcher les atteintes aux droits de l'homme », choix politiques concernant les modèles de responsabilité applicables aux systèmes d'IA, soutien des travaux de recherche technique touchant à la protection des droits de l'homme et à l'« audit algorithmique », et existence d'instruments de gouvernance légitimes pour la protection des droits de l'homme à l'ère du numérique.
- Une responsabilité incombe aux personnes qui développent et mettent en œuvre les technologies numériques. Elles doivent rendre des comptes en cas d'effets préjudiciables.

4.4. Travaux dans le domaine de la justice

- Charte éthique européenne d'utilisation de l'intelligence artificielle (IA) dans les systèmes judiciaires et leur environnement (2018)
  - Cette charte énonce cinq grands principes : respect des droits fondamentaux, non-discrimination, qualité et sécurité, transparence, neutralité et intégrité intellectuelle, et maîtrise par l'utilisateur.
  - Il a été observé que la plupart des applications de l'IA dans le domaine de la justice se trouvent dans le secteur privé : « initiatives commerciales destinées à des compagnies d'assurance, des services juridiques, des avocats et des particuliers ».
  - Parmi les usages possibles de l'IA dans un contexte judiciaire figurent la valorisation du patrimoine jurisprudentiel, l'accès au droit et la création de nouveaux outils de pilotage.
  - Parmi les autres points à envisager qui nécessitent une extrême précaution méthodologique, on peut citer la création de barèmes, l'appui à des mesures alternatives de règlement de litiges en matière civile, le règlement des litiges en ligne avant un recours en justice (lorsqu'un recours ultérieur au juge reste possible) ou l'identification des lieux de commission d'infractions.

4.5. Travaux dans le domaine de la bonne gouvernance et des élections

- Comité européen sur la démocratie et la gouvernance (CDDG)
  - Prépare actuellement une étude sur l'impact du passage au numérique sur la démocratie et la gouvernance.
- Commission de Venise : Principes pour un usage conforme aux droits fondamentaux des technologies numériques dans les processus électoraux
  - Souligne la nécessité d'une approche conforme aux droits de l'homme consistant en huit principes mettant en jeu l'utilisation des technologies numériques au cours des élections.
  - Ces huit principes, énumérés ci-dessous, sont directement tirés du document d'origine, qui en donne une description plus détaillée :
    1. « Les principes de la liberté d'expression impliquant un débat public solide doivent être traduits dans l'environnement numérique, en particulier en période électorale. »
    2. « Pendant les campagnes électorales, un organe d'administration électorale (EMB) ou un organe judiciaire impartial et compétent devrait être habilité à exiger des entreprises privées qu'elles retirent de l'internet des contenus de tiers clairement définis, sur la base des lois électorales et conformément aux normes internationales. »
    3. « Pendant les périodes électorales, l'internet ouvert et la neutralité du réseau doivent être protégés. »
    4. « Les données personnelles doivent être protégées efficacement, en particulier pendant la période cruciale des élections. »
    5. « L'intégrité électorale doit être préservée grâce à des règles et réglementations périodiquement révisées sur la publicité politique et sur la responsabilité des intermédiaires internet. »

6. « L'intégrité électorale doit être garantie en adaptant les réglementations internationales spécifiques au nouveau contexte technologique et en développant les capacités institutionnelles de lutte contre les cybermenaces. »
7. « Le cadre de coopération internationale et la coopération entre les secteurs public et privé devraient être renforcés. »
8. « L'adoption de mécanismes d'autorégulation devrait être encouragée. »

4.6. Travaux dans le domaine de l'égalité entre les femmes et les hommes et de la non-discrimination

- Recommandation CM/Rec(2019)1 du Comité des Ministres sur la prévention et la lutte contre le sexisme
  - Cette Recommandation dispose que les États doivent prendre des mesures pour prévenir et combattre le sexisme. Elle les invite en outre à intégrer une perspective d'égalité entre les femmes et les hommes dans tous les travaux liés à l'IA tout en trouvant des moyens de contribuer à combler les écarts entre les femmes et les hommes et à éliminer le sexisme.
- Commission européenne contre le racisme et l'intolérance (ECRI) – Discrimination, intelligence artificielle et décisions algorithmiques (2018)
  - Les applications d'IA ont trouvé des moyens d'échapper au droit existant. La plupart des lois antidiscrimination ne s'appliquent qu'à des caractéristiques protégées. Or il existe d'autres formes de discrimination qui, quoique non liées à des caractéristiques protégées, peuvent pourtant renforcer l'inégalité sociale.
  - Les valeurs et les problèmes variant selon les secteurs, l'idée de règles sectorielles pour la protection de l'équité et des droits de l'homme dans le domaine de l'IA est proposée.
  - Pour un secteur donné, l'ECRI propose plusieurs questions auxquelles une réponse doit être apportée:
    - « Quelles règles s'appliquent dans ce secteur, avec quelle logique sous-jacente ? »
    - « Comment est ou pourrait être utilisée la prise de décision automatisée dans le secteur concerné, et moyennant quels risques ? »
    - « Au vu de la logique sous-jacente des règles applicables au secteur concerné, faudrait-il revoir la réglementation pour tenir compte des décisions d'IA ? »

4.7. Travaux dans les domaines de l'éducation et de la culture

- Recommandation CM/Rec(2019)10 du Comité des Ministres visant à développer et à promouvoir l'éducation à la citoyenneté numérique
  - Invite les États membres à adopter des mesures d'orientation réglementaires sur l'éducation à la citoyenneté numérique, à associer toutes les parties prenantes concernées à la conception, la mise en œuvre et l'évaluation de la législation, des politiques et des pratiques en matière d'éducation à la citoyenneté numérique, et à évaluer l'efficacité des nouvelles politiques et pratiques.
  - Souligne l'importance de « donner aux citoyens les moyens d'acquérir des compétences nécessaires à une culture de la démocratie afin qu'ils puissent faire face aux défis et aux risques présentés par l'environnement numérique et les nouvelles technologies. »
- Comité directeur pour les politiques et pratiques éducatives (CDPPE)
  - Étudie les répercussions de l'utilisation de l'IA dans les environnements éducatifs.
- Eurimages et le Conseil de l'Europe – Entering the new paradigm of artificial intelligence and series [Le nouveau paradigme de l'intelligence artificielle et des séries] (2019)
  - Étude de l'impact des technologies prédictives et de l'IA sur le secteur audiovisuel.
  - Dans cette étude, l'utilisation de l'intelligence artificielle dans le secteur audiovisuel est taxée de « menace potentielle contre la diversité des contenus et le libre accès à l'information des citoyens des États membres. »

- Cinq recommandations finales sont présentées, parmi lesquelles « demander à Eurimages de développer les connaissances sur les séries », « proposer des conditions commerciales pour la production de séries dans les États membres en s'inspirant des bonnes pratiques internationales et encourager les collaborations » et « sensibiliser à l'impact de l'IA dans le secteur audiovisuel ».
- Il est aussi recommandé que le Conseil de l'Europe envisage la création d'un « organe directeur pour une certification des médias en matière d'IA ».

4.8. Travaux de l'Assemblée parlementaire du Conseil de l'Europe

- La convergence technologique, l'intelligence artificielle et les droits de l'homme (2017)
  - Appelle à la mise en œuvre d'« une véritable gouvernance mondiale de l'internet qui ne dépende pas de groupes d'intérêts privés ni de quelques États. »
  - De plus, l'Assemblée invite le Comité des Ministres à:
    - « Achever la modernisation de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel. »
    - « Définir le cadre de l'utilisation de robots de soins et de technologies d'assistance dans la Stratégie du Conseil de l'Europe sur le handicap 2017-2023. »
  - L'Assemblée rappelle en outre combien il est important que la responsabilité des systèmes d'IA et l'obligation de rendre des comptes à leur sujet incombent aux êtres humains, et propose un certain nombre de lignes directrices, notamment la nécessité d'informer le public sur la production de données à caractère personnel et sur le traitement de ces données, et de reconnaître les droits liés au respect de la vie privée et familiale.
- L'Assemblée parlementaire a adopté sept rapports sur l'IA, qui portent, entre autres, sur la gouvernance démocratique au regard de la discrimination et sur les aspects juridiques concernant les véhicules autonomes.
- La nécessité d'une gouvernance démocratique de l'intelligence artificielle (2020)
  - Ce rapport fait les recommandations suivantes :
    - « L'élaboration d'un instrument juridiquement contraignant sur l'intelligence artificielle [...] »
    - « Garantir qu'un tel instrument juridiquement contraignant soit fondé sur une approche globale, se rapporte à l'ensemble des cycles de vie des systèmes fondés sur l'IA, soit destiné à l'ensemble des parties prenantes et comprenne des mécanismes afin de garantir la mise en œuvre de cet instrument. »

4.9. Travaux du Congrès des pouvoirs locaux et régionaux du Conseil de l'Europe

- Le rapport intitulé *Les villes intelligentes : les défis pour la démocratie* est en cours d'élaboration et sera publié au cours du second semestre 2021.

4.10. Travaux de la Commissaire aux droits de l'homme

- Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme (2018)
  - Il convient de suivre un certain nombre de recommandations pour atténuer ou prévenir les incidences négatives des systèmes d'IA sur les droits de l'homme.
  - Une série de recommandations pratiques sont énoncées dans dix domaines d'action : analyses d'impact sur les droits de l'homme ; consultations publiques ; normes en matière de droits de l'homme dans le secteur privé ; information et transparence ; contrôle indépendant ; non-discrimination et égalité ; protection des données et respect de la vie privée ; liberté d'expression, liberté de réunion et d'association, et droit au travail ; possibilités de recours ; promotion de la connaissance et de la compréhension de l'IA.
  - Une liste de contrôle est fournie pour permettre la réalisation des recommandations figurant dans le document.

4.11. Travaux du Conseil de l'Europe dans le domaine de la jeunesse

- Stratégie 2030 du secteur jeunesse du Conseil de l'Europe (2020)
  - Appelle à une amélioration des réponses institutionnelles aux nouvelles problématiques (notamment l'IA) qui touchent les droits des jeunes et leur passage à l'âge adulte.
  - Les trois grands axes de la stratégie 2030 se déclinent comme suit :
    - « Élargir la participation des jeunes »
    - « Renforcer l'accès des jeunes aux droits »
    - « Approfondir la connaissance de la jeunesse »
  - Parmi les autres priorités thématiques figurent, entre autres, le renforcement de la capacité à faire progresser la démocratie participative, l'application de politiques en associant des groupes diversifiés de jeunes et le renforcement « des capacités, de l'action et du rôle de direction des jeunes en matière de prévention de la violence, de transformation des conflits et d'établissement d'une culture de la paix [...] ».

4.12. Travaux du Comité européen pour les problèmes criminels (CDPC)

- Étude de faisabilité quant à un futur instrument du Conseil de l'Europe sur l'intelligence artificielle et le droit pénal (2020)
  - Le groupe de travail du CDPC a été chargé, en décembre 2019, de « réaliser une étude de faisabilité visant à déterminer la portée et les principaux éléments d'un futur instrument du Conseil de l'Europe, de préférence une convention, sur l'intelligence artificielle et le droit pénal. »
  - Examine dans quelle mesure le Conseil de l'Europe pourrait ouvrir la voie à l'adoption d'un instrument juridique international sur l'IA et le droit pénal, et expose, sur la base des réponses des États membres à un questionnaire sur l'IA et le droit pénal, les grandes lignes d'un instrument international du Conseil de l'Europe sur l'IA et le droit pénal.
  - Quatre objectifs de cet instrument juridique ont été identifiés :
    i. Établir un cadre international pour le développement des législations nationales s'agissant des problématiques du droit pénal en matière d'IA (plus particulièrement concernant la responsabilité pénale dans le cadre de la conduite automatisée) ;
    ii. Inciter les États membres à prendre en compte l'enjeu juridique constaté en matière de droit pénal et d'IA, en traitant la question au niveau législatif, à l'aide de principes normatifs communs ;
    iii. Anticiper les problématiques probatoires et juridiques d'ores et déjà identifiées en matière de responsabilité pénale et d'IA et assurer les principes du procès équitable ainsi qu'une coopération internationale efficace en la matière ;
    iv. Assurer le développement des systèmes d'IA dans le respect des droits fondamentaux protégés par les instruments du Conseil de l'Europe.
  - Cette étude de faisabilité indique en conclusion : « Se mettre d'accord sur des normes communes pour répartir clairement et correctement la responsabilité pénale éventuelle et clarifier les questions de procédure connexes ainsi que les implications possibles pour les droits de l'homme doit relever d'un effort combiné entre la sphère publique et les acteurs privés, encourageant le développement de la technologie dans de bonnes conditions et dans le respect des principes fondateurs de la société civile. »

Il est frappant de constater que les progrès rapides accomplis ces vingt dernières années dans le domaine de l'intelligence artificielle (IA) et des technologies fondées sur les données placent la société contemporaine à un moment charnière où se décide quelle forme prendra le futur de l'humanité. D'un côté, la multiplication des innovations de l'IA bénéfiques pour la société promet de nous aider à lutter contre le changement climatique et la perte de la biodiversité, d'améliorer équitablement les soins médicaux, la qualité de vie, les transports, la production agricole, etc., et de remédier à bon nombre d'injustices sociales et d'inégalités matérielles qui assaillent le monde aujourd'hui. De l'autre, les innovations irresponsables de l'IA qui prolifèrent sont les signes avant-coureurs des problèmes éventuels qui nous attendent si ces technologies poursuivent sur leur lancée inquiétante.

Le présent Guide introductif a pour objet de présenter à un public général et non technique les grands concepts et principes exposés dans *l'Étude de faisabilité* du CAHAI. Il vise aussi à fournir des informations générales sur les domaines de l'innovation en matière d'IA, des normes de protection des droits de l'homme et des mécanismes de conformité, qui entrent dans le champ de cette étude. Conformément à l'engagement du Conseil de l'Europe de mener de larges consultations multipartites et de vastes actions de sensibilisation et de mobilisation, ce guide a été conçu pour faciliter la participation éclairée et pleinement pertinente d'un groupe inclusif de parties prenantes, le CAHAI souhaitant obtenir des retours d'information et des orientations générales sur les questions essentielles soulevées par *l'Étude de faisabilité*.

L'Alan Turing Institute est l'institut national pour la science des données et l'intelligence artificielle, dont le siège se trouve à la British Library. Il vise à faire de grands bonds en avant dans la recherche sur la science des données et l'intelligence artificielle afin de changer le monde en mieux.

www.turing.ac.uk

Le Conseil de l'Europe est la principale organisation de défense des droits de l'homme du continent. Il comprend 47 États membres, dont l'ensemble des membres de l'Union européenne. Tous les États membres du Conseil de l'Europe ont signé la Convention européenne des droits de l'homme, un traité visant à protéger les droits de l'homme, la démocratie et l'État de droit. La Cour européenne des droits de l'homme contrôle la mise en œuvre de la Convention dans les États membres.

www.coe.int

The Alan Turing Institute

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE