

# DÉVELOPPER des systèmes IA conformes aux valeurs éthiques DE L'ENTREPRISE

Les différentes tentatives destinées à fixer un cadre juridique à l'IA prennent soin de rappeler que le développement d'un projet d'IA doit s'accompagner d'une réflexion sur l'éthique, afin d'y intégrer, dès sa conception, les valeurs humaines fondamentales.

Le Comité Économique et Social Européen a, dans son avis du 31 mai 2017, préconisé *“l’instauration d’un code de déontologie pour le développement, le déploiement et l’utilisation de l’intelligence artificielle afin que les systèmes d’IA demeurent, tout au long de leur processus d’exploitation, compatibles avec les principes de dignité humaine, d’intégrité, de liberté, de respect de la vie privée, de diversité culturelle et d’égalité entre hommes et femmes, ainsi qu’avec les droits fondamentaux”*.

De nombreuses entreprises se sont ainsi dotées de chartes éthiques dont l'ambition est de répondre à la nécessité d'intégrer, dès la naissance d'un projet d'IA, et, à chaque étape de sa construction, des considérations éthiques.

Le groupe de travail “Souveraineté, Éthique et Juridique” de la Cité de l'IA a souhaité partager le fruit de ses réflexions sur les considérations éthiques devant, selon lui, être prises en compte par les structures portant des projets d'IA et pouvant être formalisées, le cas échéant, dans une charte éthique.

Il s'agit de rappeler les valeurs et principes identifiés comme importants par l'entreprise dans le développement d'un système d'intelligence artificielle. Ce type de charte, ou tout autre moyen de formalisation des réflexions menées, a vocation à s'appliquer à l'ensemble des collaborateurs impliqués dans la conception, le développement, et/ou l'exploitation de l'IA. Il s'agit de s'assurer qu'ils comprennent les risques et se sentent responsables du succès de l'IA et de son éthique.

Un tel instrument doit être porté au plus haut niveau de l'entreprise. S'il est destiné aux créateurs d'IA, le rappel de la ligne de conduite éthique de l'entreprise est dicté par la prise en compte des utilisateurs ou personnes “visées” par l'outil développé.

L'adoption d'une « Charte éthique » peut s'accompagner d'une procédure de contrôle interne permettant à l'entreprise de s'assurer que les engagements et obligations qui y sont mentionnés sont respectés.

# DÉVELOPPER des systèmes IA conformes aux valeurs éthiques DE L'ENTREPRISE

## Le respect des droits humains fondamentaux

Les droits humains fondamentaux issus, notamment, de la Charte des droits fondamentaux de l'Union européenne correspondent, sur un territoire et à une époque donnés, aux standards que les européens ont identifiés comme le socle de leur identité fondamentale.

Cette identité est généralement articulée autour des principes suivants : droit à la dignité, à l'intégrité mentale et physique, à la liberté, à l'égalité, à l'équité de traitement, à l'intimité et au respect de la vie privée.

L'adoption de règles, dès la conception d'un système intégrant de l'IA, permettant d'identifier et de mettre en œuvre ces principes peut ainsi garantir la connaissance et le respect, par toutes les personnes impliquées dans le projet, de ces garde-fous fondamentaux.

S'il est envisageable qu'un arbitrage soit nécessaire entre la somme des avantages associés au système d'IA d'une part et les risques individuels prévisibles d'autre part, il faut néanmoins conserver à l'esprit que certains droits fondamentaux sont absolus et ne peuvent dépendre d'arbitrages au niveau de l'entreprise.

## La prévention des discriminations

Il s'agit d'identifier, de prévenir et d'éliminer l'apparition ou la reproduction de biais discriminatoires pendant les phases de test, lors du passage à l'échelle, dans le cadre de ré-entraînements réguliers, etc..

Ce risque existe en particulier pour les IA basées sur le machine learning.

Le principe de non-discrimination doit être appliqué au niveau de la donnée, produit d'un arbitrage humain, influencé par des considérations politiques, culturelles ou temporelles notamment, voire des préjugés.

La prévention de ce risque peut être effectuée, tout d'abord, grâce à une évaluation préalable, associant si possible des regards (genre, origine ethnique...) et compétences variés, et destinée à identifier les facteurs susceptibles de créer des biais.

Les équipes doivent ainsi être sensibilisées à ce risque et invitées à apporter, le cas échéant, des correctifs. Ceux-ci peuvent être statistiques (y compris en introduisant des données d'entraînement artificielles) ou algorithmiques (faisant de la non-discrimination une variable de premier rang).

# DÉVELOPPER des systèmes IA conformes aux valeurs éthiques DE L'ENTREPRISE

## L'importance de proposer une IA de confiance

Travaillant à partir de grandes quantités de données, dont certaines à caractère personnel, le système d'IA peut avoir des conséquences importantes sur les droits et libertés des personnes concernées.

Au delà du respect des règles issues du Règlement Général sur la Protection des Données (RGPD), il s'agit également de vérifier que des mesures suffisantes seront prises pour informer les personnes et ainsi s'assurer, par exemple, qu'elles ont compris qu'elles interagissent avec une machine.

Une supervision par l'humain est nécessaire et doit être prise en compte dès la conception de l'IA lorsqu'une décision est exclusivement fondée sur un traitement automatisé pouvant avoir des effets juridiques ou affecter une personne.

La confiance dans l'IA passe, en outre, par l'explicabilité du système, permettant aux personnes de comprendre pourquoi tel résultat a été généré, telle décision prise. Cette exigence se traduit aussi par la traçabilité des processus de collecte, d'étiquetage des données, de la méthode d'entraînement de l'algorithme, qui doit être sérieusement documentée.

Il est également fondamental de s'assurer que la finalité de l'application d'IA (c'est-à-dire la raison pour laquelle elle est développée) soit clairement définie, respectée et explicitée auprès des personnes concernées.

Le système doit, de surcroît, sous réserve des droits de propriété intellectuelle et du secret des affaires, être auditable.

L'analyse des conséquences de l'IA sur, par exemple, la santé mentale, le comportement, les choix des utilisateurs ou encore sur un risque de harcèlement ou d'addiction, pourrait, par ailleurs, être menée dans le cadre d'une démarche éthique.

## La prise en compte de l'impact environnemental

Compte-tenu des impacts environnementaux élevés (consommation d'énergie, émissions de GES, utilisation de ressources, déchets et biodiversité sur le cycle de vie), l'IA devrait s'inscrire dans une logique d'éco-conception.

Il s'agit ainsi d'intégrer les standards écologiques dans la conception, la fabrication et l'utilisation d'un système d'IA, des centres de données associés, voire des capteurs utilisés.

# DÉVELOPPER des systèmes IA conformes aux valeurs éthiques DE L'ENTREPRISE

Dans cette logique, une évaluation environnementale du projet peut être envisagée, couvrant l'utilisation des équipements pour la collecte des données, le fonctionnement des centres de données, l'utilisation d'énergies renouvelables aux étapes les plus intenses en énergie, l'extraction des matières premières et la fin de vie des équipements.

Des enjeux éthiques peuvent également être identifiés relativement à (i) la localisation, dans le monde, des centres de données pour éviter le détournement de ressources locales ou (ii) à la gestion prévisionnelle des métaux rares (accès raisonné, gestion des déchets, économie circulaire...).

Le développement attendu de certifications relatives à la performance énergétique du système d'IA, permettant de prendre en compte tout le cycle de vie des systèmes et l'infrastructure à laquelle ils sont liés, doit aussi faire l'objet d'une attention particulière.

## La préservation de la place centrale de l'humain

L'IA doit être considérée comme un outil au service de l'homme, destinée à en augmenter les capacités et la fiabilité.

La liberté de l'utilisateur de l'IA doit néanmoins pouvoir être préservée, en conférant à l'IA un rôle de facilitateur de décision, en n'imposant par exemple pas de suivre les recommandations fournies et en conservant la place de la responsabilité humaine dans le processus.

Il peut également être intéressant de s'interroger sur la préservation de l'aspect irrationnel, sensible et aléatoire inhérent à l'humain. Il semble aussi essentiel de veiller à éviter le traitement déshumanisé de situations par l'IA et de distinguer les interactions avec un humain et avec une machine.

Plus largement, le questionnement sur la légitimité des usages de l'IA peut être abordé dans la charte éthique, ainsi que l'articulation entre les valeurs collectives et individuelles portées par l'entreprise.

En conclusion : toute charte éthique a vocation à évoluer dans le temps, à s'adapter aux nouveaux enjeux sociétaux ou réglementaires. Une gouvernance pluridisciplinaire est ainsi pertinente pour la mise en œuvre de ces principes éthiques qui s'inscrivent en complémentarité des contraintes règlementaires relatives à tout projet d'IA.