Consultation – Livre blanc sur l'intelligence artificielle

SNCF tient tout d'abord à souligner la qualité du projet de Livre blanc de la Commission Européenne « Intelligence Artificielle une approche européenne axe sur l'excellence et la confiance ». SNCF suit avec intérêt les travaux et réflexions engagés depuis maintenant plusieurs années par la Commission, pour mettre en place un cadre européen unifié pour créer une IA de confiance.

SNCF confirme son intérêt pour coopérer utilement à la mise en place d'une telle IA de confiance. Nous avons à cet effet déjà transmis nos observations sur le premier projet du HLEG, portant sur les lignes directrices en matière éthique pour une IA digne de confiance, et en participant activement au programme de test du questionnaire proposé par le HLEG à travers plusieurs projets concrets (*deep dive*).

Nous présenterons nos observations en suivant le déroulé du Livre blanc pour faciliter la compréhension de nos observations. Le texte du Livre blanc figure en italique souligné et nos observations figurent en dessous

1. TITRE

Observation générale :

La référence dans le titre à une IA de confiance nous semble devoir être modifiée. En effet, la confiance est définie comme les lignes directrices en matière d'éthique pour une IA digne de confiance :

« nous reprenons la définition suivante tirée de la littérature : « la confiance se définit comme : 1) un ensemble de convictions spécifiques portant sur la bienveillance, la compétence, l'intégrité et la prévisibilité (convictions en matière de confiance) ; 2) la volonté d'une partie de dépendre d'une autre partie dans une situation risquée (intention de faire confiance), ou 3) la combinaison de ces éléments » (Siau K WANG Building Trust in Artificial Intelligence, Machine Learning and Robotics). Si la confiance n'est en général pas une propriété que l'on prête aux machines, le présent document vise à souligner l'importance d'être capable de faire confiance non seulement dans le fait que les SIA sont conformes sur le plan juridique, respectent l'éthique et sont robustes, mais aussi que cette confiance peut être accordée à l'ensemble des personnes et des processus impliqués dans le cycle de vie du système d'IA ».

Cette définition nous semble vague et sujette à de nombreuses interprétations et polémiques. Ne faudrait-il pas plutôt parler d'une approche éthique et responsable de l'IA ?

2. INTRODUCTION

Observation générale

L'IA pose clairement la question du choix de société souhaitée et de la place de l'IA dans cette société. Sont évoqués dans le Livre blanc l'état de droit et l'attachement

aux valeurs et droits fondamentaux tels que la liberté, la dignité humaine et la protection de la vie privée, la démocratie, le développement durable. C'est ce choix de société qui devrait guider les décisions européennes en matière d'IA.

La question des usages de l'IA pose également la question cruciale de la place du travail dans les nouveaux modèles de société.

« *C'est une chance pour l'Europe, qui est profondément attachée aux valeurs et à l'état de droit et possède une capacité avérée à fabriquer des produits et fournir des services complexes, sûrs et fiables, dans des secteurs aussi divers que l'aéronautique, l'énergie, l'automobile et les équipements médicaux.* »

Observation :

Pourquoi ne pas évoquer les transports en général plutôt que l'aéronautique et l'automobile ?

« *dans le domaine des services d'intérêt public, par exemple, les coûts de fourniture de services (transports, éducation, énergie et gestion des déchets) seront réduits, la durabilité des produits sera améliorée et les services répressifs disposeront d'outils appropriés pour assurer la sécurité des citoyens, avec des garanties adéquates en matière de respect des droits et des libertés* ».

Observation :

Pour établir un réel bilan des améliorations sociétales générées par les SIA, il faudrait prendre en compte non seulement l'impact de l'IA sur les services mais également celui sur les emplois, car les SIA engendreront de nombreuses suppressions d'emplois dont seulement certaines seront compensées par de postes nouveaux. Une réflexion approfondie sur les emplois qui vont disparaître, une évaluation du nombre et de la nature de ces emplois, ainsi que sur les nouveaux métiers à anticiper dans les services d'intérêt public serait utile.

« *L'utilisation des systèmes d'IA peut jouer un rôle considérable dans la réalisation des objectifs de développement durable, et en ce qui concerne le soutien du processus démocratique et des droits sociaux* ».

Observation :

Sans contrôle, les SIA n'auront pas forcément que des aspects positifs dans la vie démocratique, car ils sont susceptibles de faciliter le profilage, les bulles d'enfermement et le *nudge* qui permettent d'orienter les choix politiques des citoyens en réduisant la sphère de débats et de contradictions, indispensables à l'exercice d'une pensée critique et éclairée.

« *Avec les propositions qu'elle a formulées récemment dans le pacte vert pour l'Europe*[6]*, l'UE montre l'exemple pour ce qui est de relever les défis climatiques et ceux liés à l'environnement. Les technologies*

numériques telles que l'IA sont des facteurs déterminants pour atteindre les objectifs fixés dans le cadre du pacte vert. L'importance de l'IA ne cessant de croître, il faut dûment tenir compte de l'incidence environnementale des systèmes d'IA tout au long de leur cycle de vie et sur l'ensemble de la chaîne d'approvisionnement, c'est-à-dire en ce qui concerne l'utilisation des ressources pour l'entraînement des algorithmes et le stockage des données ».

Observation : il nous semble utile de rajouter ici l'impact environnemental de toute la chaîne d'IA, notamment l'évaluation du cycle de vie des capteurs (Life Cycle Assessment) utilisés pour la collecte des données. Ils peuvent être composés de métaux rares et consomment une quantité d'énergie non négligeable. Le stockage et le traitement de données en masse pose aussi des questions au niveau du développement durable visé.

« *La stratégie européenne pour les données, qui accompagne le présent Livre blanc, vise à permettre à l'Europe de devenir l'économie tirant parti de ses données la plus attrayante, la plus sûre et la plus dynamique au monde, en mettant à sa disposition des données qui contribueront à améliorer, d'une part, les processus de décision et, d'autre part, la qualité de vie de tous ses citoyens* ».

Observation :

La notion de « qualité de vie » semble vague et cette notion varie d'un État membre à l'autre. Il faudrait utiliser des critères de mesure précis.

3. Il faut par ailleurs faire attention à la fuite des données vers des pays où elles ne sont pas protégées. Le partage des données ne doit pas se faire au détriment de l'économie européenne.

3.SAISIR LES OPPORTUNITÉS FUTURES: PRENDRE LA NOUVELLE VAGUE DE DONNÉES

Observation :

Des investissements dans le domaine des calculateurs quantiques devraient également être proposés car ils permettent une puissance de calcul très importante ce qui constitue un fort enjeu concurrentiel au regard des augmentations de volume de données. Il faudra néanmoins que son utilisation soit nécessaire et proportionnelle au besoin.

4. UN ÉCOSYSTÈME D'EXCELLENCE

A. COOPÉRATION AVEC LES ÉTATS MEMBRES

a. Action 1: en tenant compte des résultats de la consultation publique sur le Livre blanc, la Commission proposera aux États

membres une révision du plan coordonné en vue d'une adoption d'ici à la fin 2020.

« Le financement au niveau de l'UE dans le domaine de l'IA devrait permettre d'attirer et de mutualiser les investissements dans des domaines où l'action requise va au-delà de ce qu'un État membre peut réaliser seul. L'objectif est d'attirer, dans l'UE, un montant total de plus de 20 milliards d'euros[18] d'investissements par an dans l'IA au cours de la prochaine décennie. Pour stimuler les investissements publics et privés, l'UE mettra à disposition des fonds au titre du programme pour une Europe numérique, du programme Horizon Europe et des Fonds structurels et d'investissement européens afin de répondre aux besoins des régions moins développées et des zones rurales ».

Observation : il faudra veiller à ce que ces fonds soient affectés à des projets dans le cadre de cahiers des charges précis, avec une publicité européenne, dans des délais d'attribution proportionnels à l'évolution des technologies. Le financement devra être versé en plusieurs tranches, en fonction des résultats précis attendus qui puissent faire l'objet d'une évaluation objective, non seulement « ex post » mais également « en continu » sur la base d'indicateurs objectifs, mesurables et consensuels. Ces fonds devront être bien répartis en fonction des pays, des domaines et des applications, de sorte à ce qu'ils soient attribués de façon équitable. Il faudra prévoir des engagements contractuels des bénéficiaires de fonds européens de développer leurs SIA dans l'Union Européenne. Ils devront respecter l'intégralité de la règlementation et des principes éthiques européens, et s'engager à les faire respecter par leurs sous-traitants et partenaires établis en dehors de l'UE.

« Action 2: la Commission facilitera la création de centres d'essai et d'excellence faisant appel à des investissements européens, nationaux et privés, y compris éventuellement un nouvel instrument juridique ».

Observation :

Le cadre juridique de ces centres d'essai sera déterminant pour clarifier leurs rôles, les attentes, les bénéficiaires qui devraient être européens etc. Il devra également définir les règles applicables qui devraient être allégées de sorte à ne pas entraver la recherche et à l'avancée scientifique et conserver l'attractivité des chercheurs européens.

« Action 3: mettre en place et soutenir, par l'intermédiaire du pilier « compétences avancées » du programme pour une Europe numérique, des réseaux d'universités et d'établissements d'enseignement supérieur de premier plan pour attirer les meilleurs professeurs et chercheurs et proposer des programmes de masters de classe internationale dans le domaine de l'IA. »

Consultation – Livre blanc sur l'intelligence artificielle

Observation :

Il faudra veiller à ce que cela se fasse de concert avec les entreprises, notamment par un système d'alternance, pour que les travaux de recherches ne soient pas déconnectés des besoins des entreprises.

« Les domaines de la santé et des transports, notamment, feront l'objet d'une attention particulière car, dans ces secteurs, la technologie est parvenue à la maturité nécessaire à un déploiement à grande échelle ».

Observation :

Le terme « maturité » est vague. Des exemples concrets seraient utiles.

5. UN ÉCOSYSTÈME DE CONFIANCE : UN CADRE RÉGLEMENTAIRE POUR L'IA

B. ADPATATIONS POSSIBLES DU CADRE LEGISLATIF EXISTANT DE L'UE POUR TENIR COMPTE DE L'IA

« Les principaux risques liés à l'utilisation de l'IA concernent l'application des règles visant à protéger les droits fondamentaux (notamment la question des données à caractère personnel, le respect de la vie privée et la non-discrimination, ainsi que les questions liées à la sécurité et à la responsabilité ».

Observations

Toutes ces questions font l'objet déjà de règlementations multiples et complexes comme il est rappelé dans le Livre blanc. Ainsi, le Livre blanc donne comme exemple de secteur à haut risque qui devrait faire l'objet d'une nouvelle règlementation, l'identification des personnes par reconnaissance faciale. Or, le droit européen permet déjà d'encadrer très largement ces cas d'usage, qu'il s'agisse de l'application de la convention européenne des droits de l'homme, de la charte des droits fondamentaux de l'Union Européenne, du règlement général européen sur la protection des données et de la directive police justice.

Est-il pertinent de règlementer une technologie mouvante qui avec le temps ne cessera de se développer, et non les résultats de cette technologie susceptibles de générer un préjudice ? Le Livre blanc évoque les IA boîtes noires et le machine learning, mais les techniques évoluant si rapidement, peut-être sera-t-il possible demain de rendre explicable ce qui ne l'était pas aujourd'hui ? La répartition de la responsabilité entre les différents acteurs de la chaîne d'approvisionnement ou de production du service est une question importante, mais n'est-ce pas le rôle de la jurisprudence d'adapter les principes de responsabilité déjà existants ?

Une approche de droit souple (sous la forme de « soft law ») semble plus adaptée à cette technologie évolutive. Elle permettrait d'élargir son application au-delà des frontières de l'Union Européenne et éviterait les contraintes génératrices de distorsions de concurrence.

Consultation – Livre blanc sur l'intelligence artificielle

L'important n'est-il est que les praticiens de l'IA, les chercheurs, les ingénieurs, les techniciens, tous les opérationnels de l'IA et du développement des cas d'usage s'approprient les grands principes européens rappelés dans la charte des droits fondamentaux de l'Union Européenne et dans les lignes en matière éthique pour une IA digne de confiance pour une intelligence artificielle ? Une conformité imposée par le biais de normes nécessitera de lourds budgets en cabinets de conseils, extérieurs aux entreprises, ce qui ne permettra pas une acculturation à l'éthique et défavorisera les entreprises les plus fragiles. A partir des grands principes déjà bâtis dans les documents précités, les acteurs pourraient établir entre eux des codes de conduite, notamment pour chacun des cas d'usage jugés les plus à risque. Ces codes de conduite auraient une valeur « forte » puisqu'ils seraient adossés à la règlementation souple évoquée plus haut. Ils devraient être évidemment compatibles avec elle, et avoir reçu l'engagement de tous les acteurs.

La règlementation ne doit pas être un frein au développement, à l'innovation, à la recherche et à la compétitivité des développements européens en matière d'intelligence artificielle. Il ne faut pas briser l'élan de l'innovation, surtout au moment où les principaux concurrents de l'Union Européenne (Chine, États-Unis, etc.) investissent massivement.

Une approche européenne et si possible mondiale semble nécessaire pour permettre une libre circulation des marchandises et services et éviter des distorsions de concurrence.

C. CHAMP D'APPLICATION D'UN FUTUR CADRE RÉGLEMENTAIRE DE L'UE

- *« Selon la Commission, une application d'IA devrait généralement être considérée comme étant à haut risque en fonction de ce qui est en jeu, en examinant si des risques importants sont associés à la fois au secteur et à l'utilisation envisagée, notamment du point de vue de la protection de la sécurité, des droits des consommateurs et des droits fondamentaux. En particulier, une application d'IA devrait être considérée comme étant à haut risque si elle remplit cumulativement les deux critères suivants:*

  premièrement, l'application d'IA est employée dans un secteur où, compte tenu des caractéristiques des activités normalement menées, des risques importants sont à prévoir. Ce premier critère garantit que l'intervention réglementaire cible les domaines dans lesquels, d'une manière générale, les risques sont réputés les plus probables. Le nouveau cadre réglementaire devrait comprendre une liste précise et complète des secteurs concernés. Par exemple, les soins de santé ; les transports ; l'énergie et certains pans du secteur public[50]*. La liste devrait être réexaminée à intervalles réguliers et modifiée, le cas échéant, en fonction des réalités ;*

  deuxièmement, l'application d'IA dans le secteur en question est,

de surcroît, utilisée de façon telle que des risques importants sont susceptibles d'apparaître. Ce second critère prend en considération le fait que toutes les utilisations de l'IA dans les secteurs sélectionnés n'impliquent pas nécessairement des risques importants. Par exemple, si les soins de santé peuvent, d'une manière générale, faire partie des secteurs concernés, une défaillance du système de planification des rendez-vous dans un hôpital ne présentera normalement pas de risques tels qu'ils justifient une intervention législative. L'appréciation du niveau de risque d'une utilisation donnée pourrait reposer sur ses conséquences pour les parties concernées. Par exemple, les utilisations d'applications d'IA qui produisent des effets juridiques sur les droits d'une personne physique ou d'une entreprise, ou l'affectent de manière significative de façon similaire ; qui occasionnent un risque de blessure, de décès ou de dommage matériel ou immatériel important ; dont les effets ne peuvent être évités par les personnes physiques ou morales.

Grâce à l'application cumulative des deux critères, la portée du cadre réglementaire serait circonscrite avec précision et garantirait la sécurité juridique. Les exigences obligatoires contenues dans le nouveau cadre réglementaire sur l'IA (voir la section D) ne s'appliqueraient en principe qu'aux applications désignées comme étant à haut risque selon ces deux critères cumulatifs ».

Observation :

Cette approche par un critère sectoriel semble floue, subjective, difficile à mettre en œuvre, et risquée. Elle nous semble devoir être abandonnée.

Le second critère pourrait s'appliquer à la plupart des secteurs. Il est cependant flou et laisse une marge d'appréciation importante aux entreprises et aux autorités nationales. Ceci peut être une source d'insécurité juridique. L'appréciation de ce critère nécessiterait la réunion de comités d'éthique qui n'existent que dans les grands groupes. Par ailleurs, les autorités de contrôle nationales auront-elles les moyens de contrôler efficacement le respect de ces critères ?

- *« Nonobstant les considérations qui précèdent, il peut également exister des cas exceptionnels dans lesquels, compte tenu des risques, l'utilisation d'applications d'IA à certaines fins devrait être considérée comme étant à haut risque en soi, c'est-à-dire indépendamment du secteur concerné, et resterait soumise aux exigences ci-dessous*[51]*. À titre d'illustration, il pourrait s'agir notamment des applications suivantes :*

compte tenu de son importance pour les particuliers et de l'acquis de l'UE sur l'égalité en matière d'emploi, l'utilisation d'applications d'IA dans les procédures de recrutement et dans des situations ayant une incidence sur les droits des travailleurs serait toujours considérée comme étant «à haut risque …»

Observation : L'expression « *des situations ayant une incidence sur les droits des travailleurs* » est générale. Elle fait entrer dans son champ beaucoup de situations qui peuvent pourtant ne pas être à haut risque. On donnera l'exemple de la formation obligatoire en réalité virtuelle qui pourrait, en cas d'irrespect, faire l'objet de sanctions directes ou indirectes (baisse de prime etc.) et qui ne présente pourtant aucun risque en elle-même.

- *« l'utilisation d'applications d'IA à des fins d'identification biométrique à distance*[52] *et pour d'autres technologies de surveillance intrusive serait toujours considérée comme étant « à haut risque », et les exigences ci-dessous s'appliqueraient dès lors à tout moment ».*

  Observation :

  Nous avons déjà mentionné que selon nous l'identification par reconnaissance faciale était déjà encadrée. Les remarques formulées ci-dessous ont pour unique but de montrer comme il est difficile de définir de façon formelle et figée, notamment dans un texte règlementaire, les usages à haut risque et ceux qui ne le sont pas**.**

  En effet, de nombreuses fonctions de la reconnaissance faciale nous semblent à haut risque et pas uniquement les fonctions d'identification. Par exemple, les fonctions de catégorisation. Elles peuvent permettre d'identifier des catégories de personnes selon leur origine ethnique.

  Un des risques de cette technologie est sa combinaison avec d'autres bases de données qui permettront l'identification des individus dans de très nombreux domaines, permettant un profilage massif des individus et portant une très forte atteinte aux libertés et droits fondamentaux. La détection des émotions, à supposer que nos connaissances en neurosciences nous autorisent à considérer qu'elle soit véritablement possible, génère elle aussi une forte intrusion dans l'intimité d'une personne ; elle porte atteinte à sa vie privée et à sa dignité ; elle nous semble donc être à haut risque. Il faut ajouter à cela les risques d'erreur d'interprétation dont les conclusions erronées pourraient avoir des conséquences négatives même dans des domaines où l'usage pourrait sembler justifié, comme dans le domaine médical ou de la sécurité. Le risque de détournement d'usage ou de hacking est également à prendre sérieusement en

compte.

Même le recours à la reconnaissance faciale pour permettre à une personne de s'authentifier peut être également à haut risque. Il y a en premier lieu le risque éthique de faux négatif si la personne n'est pas reconnue, ce qui pourrait laisser croire qu'elle n'est pas en règle et porter ainsi atteinte à sa dignité. Or, comme le présent Livre blanc le souligne l'on sait que les faux négatifs sont plus fréquents chez les personnes de couleur, ce qui peut générer une forme de discrimination. La reconnaissance faciale peut également être utilisée par les entreprises pour contrôler l'accès à des ordinateurs ou à des applications. La personne risque de ne guère avoir le choix que d'accepter de recourir à la reconnaissance faciale pour avoir accès aux services, même si, dans un premier temps, une alternative pourrait lui être proposée. Même si le système reposait sur le consentement des individus, comment penser que celui-ci soit réellement libre quand il existe un rapport de forces inégales ? Le recours à la reconnaissance faciale dans ces cas-là peut très vite donner l'impression aux personnes d'être épiées dans leurs comportements, dans leurs horaires, ou dans leur assiduité, ce qui peut générer un sentiment de surveillance, d'atteinte à la vie privée et aux libertés individuelles.

La collecte de données biométriques qui constituent un des attributs du caractère unique d'une personne peut être vécue comme une atteinte à la dignité. Plus le système se généralisera, plus il y aura un risque de sentiment de perte d'individualité ; le visage qui exprime les émotions et la sensibilité d'une personne prendra la dimension d'un simple outil parmi d'autres outils, générant ainsi un sentiment de « dé personnification » et de déshumanisation. La matière première de cette technologie n'est rien moins que nos visages. Peut-on considérer que le visage d'un utilisateur constitue une « data » comme les autres ?

De plus, un mauvais usage ou un détournement d'usage peut avoir des conséquences graves sur les droits et libertés des personnes : usurpation d'identité, diffusion d'images sur les réseaux sociaux, chantage, harcèlement etc.

Comme le souligne la CNIL, « les dispositifs de reconnaissance faciale sont particulièrement intrusifs et présentent des risques majeurs d'atteinte à la vie privée et aux libertés individuelles des personnes concernées. Ils sont par ailleurs de nature à créer un sentiment de surveillance renforcée ».

Par ailleurs, sans qu'il ne s'agisse ni de reconnaissance faciale ni de surveillance intrusive, tous les systèmes de suivi de personnes devraient être considérés à haut risque.

Consultation – Livre blanc sur l'intelligence artificielle

Tous les SIA qui permettent une « lecture » des émotions ou des pensées devraient être considérés comme à haut risque ainsi que tous les systèmes portant atteinte à la capacité de jugement et à la liberté de pensée.

D. TYPES D'EXIGENCES

« Compte tenu des lignes directrices élaborées par le groupe d'experts de haut niveau et des éléments exposés ci-dessus, les exigences imposées aux applications d'IA à haut risque pourraient porter sur les éléments essentiels suivants, détaillés dans les points ci-dessous:

- *données d'entraînement ;*
- *conservation des données et des dossiers ;*
- *informations à fournir ;*
- *robustesse et précision ;*
- *contrôle humain ;*
- *des exigences spécifiques pour les applications d'IA utilisées à des fins données, telles que l'identification biométrique à distance ».*

Observation : il semblerait plus simple et plus cohérent de reprendre ici les 7 points considérés comme majeurs par le HLEG (a. facteur humain et contrôle humain, b. robustesse technique et sécurité, c. Respect de la vie privée et gouvernance des données, d. Transparence, e. Diversité, non-discrimination et équité, f. Bien-être sociétal et environnemental, et g Responsabilisation), qui sont d'ailleurs plus ou moins repris dans les développements consacrés aux points ci-dessus. Ces différents points nous semblent poser des grands principes qui permettent d'élaborer un droit souple, applicable au-delà de l'Union Européenne.

« Des exigences garantissant que des mesures raisonnables sont prises pour veiller à ce que toute utilisation ultérieure des systèmes d'IA ne donne pas lieu à des cas de discrimination interdite. En particulier, ces exigences pourraient comporter des obligations d'utiliser des ensembles de données suffisamment ***représentatifs****, notamment pour garantir la prise en compte, dans ces ensembles de données, de tous les aspects pertinents du genre, de l'appartenance ethnique et d'autres motifs possibles de discrimination interdit ».*

Observation :

Comment juge-t-on que cette représentativité est suffisante ou non ?

« les résultats du système d'IA ne deviennent effectifs que s'ils ont été préalablement réexaminés et validés par un être humain (ex.: la décision de

rejeter une demande de prestations de sécurité sociale ne peut être prise que par un être humain) » ;

Observation :

Il faut s'assurer que les humains sont en réelle capacité de valider préalablement les décisions prises par l'IA (nombre d'humains et exercice réel de leur esprit critique).