LIVRE BLANC
MARS 2024

L’Intelligence Artificielle de Confiance

ACN
Alliance pour la confiance numérique

MARINA FERRARI
Secrétaire d'Etat

EDITO

L'intelligence artificielle (IA) s'est imposée comme une révolution technologique comparable à l'arrivée de l'électricité dans nos sociétés. Elle occupe désormais une place de plus en plus importante dans notre quotidien, alimentant les applications que nous utilisons ou les services numériques auxquels nous avons recours. De son côté, l'émergence de l'IA générative rend ses progrès encore plus visibles pour le citoyen, lui laissant entrevoir un puissant pouvoir de transformation. Les avantages sociétaux et économiques qu'il est possible d'en tirer sont nombreux, tant pour la société civile que pour les entreprises.
Cette puissance suscite également, chez les Français, des craintes, légitimes, qui invitent à penser le cadre de déploiement d'une l'IA « de confiance » : comment nos données sont-elles traitées ? quel impact les algorithmes peuvent-ils avoir sur nos vies ? Pour tirer pleinement parti des opportunités associées à l'IA, l'acceptabilité et la compréhension de son déploiement sont essentiels.
Ainsi, le fonctionnement des algorithmes doit être sûr, explicable et responsable, notamment pour les systèmes qui peuvent engager des vies humaines ou porter atteinte aux droits des individus. L'IA doit être de confiance pour limiter les risques de défaillances, assurer la sécurité des utilisateurs et renforcer la fiabilité des systèmes. La France, dont le message universaliste a traversé les siècles et les crises, doit être à la pointe sur ces sujets et montrer l'exemple. Il ne s'agit pas simplement d'une question de valeurs, c'est également une question d'opportunité : nous ne pourrons tirer les dividendes de l'IA sans confiance.

L'IA de confiance est ainsi cruciale pour de nombreux secteurs industriels pour lesquels la fiabilité est primordiale : l'automobile, le ferroviaire, l'aéronautique, ou l'industrie pharmaceutique. **Les industriels français en ont d'ailleurs fait une priorité** comme l'illustre le Manifeste pour une IA au service de l'industrie rassemblant de grands acteurs industriels s'engageant sur le sujet. Dans le même mouvement que les acteurs privés, **l'Etat français investit massivement dans le secteur de l'intelligence artificielle dans le cadre de France 2030 :**

Ce sont plus de 2,2 milliards d'euros qui sont investis dans notre économie mais aussi notre formation en IA. L'IA de confiance y tient une place de premier plan. C'est le cas par exemple dans le soutien à des plateformes partenariales ouvertes mettant à la disposition des entreprises des outils pour valider et certifier leurs systèmes d'IA, ainsi que pour expérimenter et faciliter l'intégration d'algorithmes dans leurs processus opérationnels contraints. Ces plateformes visent à poursuivre les développements effectués dans le programme Confiance.ai, le plus important projet industriel européen sur la thématique. Grâce à ce programme, un écosystème riche d'acteurs tant grands groupes, startups et PME qu'académiques et chercheurs ont collaboré pour établir les fondements d'une IA de confiance.

Impulsée par le président Emmanuel Macron, la stratégie française s'inscrit également dans une dynamique européenne de diffusion d'une IA digne de confiance avec le financement d'infrastructures de tests et d'évaluation des systèmes d'IA dans quatre secteurs stratégiques que sont l'industrie manufacturière, la santé, l'agro-agri et la mobilité. Aujourd'hui, nos actions et investissements dans **la fiabilisation et la sécurisation des systèmes d'IA posent les jalons de notre compétitivité et de notre autonomie stratégique futures.**

Le développement de l'IA de confiance requiert également un cadre réglementaire clair et adapté. Le Règlement européen sur l'Intelligence Artificielle, dont la France a notamment présidé les débats lors de Présidence française du Conseil de l'UE, fixe ainsi des standards élevés pour limiter les risques, tout promouvant une IA explicable et digne de confiance. Son application doit désormais maintenir un juste équilibre entre innovation et protection.

Ce sont toutes ces actions qui nous permettront de dessiner le paysage de l'IA de confiance de demain, en France et en Europe, au service de notre souveraineté. Dans ce contexte, je souhaite saluer la mobilisation et la qualité des travaux de l'Alliance pour la Confiance Numérique pour la mise en valeur de ce sujet stratégique pour notre économie.

SOMMAIRE

INTRODUCTION

La transformation numérique initiée depuis plusieurs années a engendré des innovations de plus en plus nombreuses. Aujourd'hui, force est de constater que l'une de ces innovations prend une place prépondérante dans les espaces publics et privés ainsi que dans les débats qui animent nos concitoyens : l'intelligence artificielle (IA). Cette dernière est en effet sous le feu des projecteurs, la propulsant au cœur des débats politiques et sociétaux. Le plus souvent, l'IA y est réduite à une dystopie de science-fiction ou à Chat GPT, IA conversationnelle et générative, qui est la première de son genre à bénéficier d'un écho aussi important en aussi peu de temps. Lancé en novembre 2022 par Open AI, cet agent conversationnel se définit comme « *conçu pour comprendre et générer du texte en réponse à une grande variété de questions et de demandes », qui « tire [ses] connaissances d'une vaste gamme de sources en ligne et capable d'aider avec des sujets allant d'informations générales aux conseils spécifiques*[1] » . Capable de compiler un très grand nombre d'informations pour les résumer en une réponse unique, cet exemple d'utilisation d'une IA générative exerce depuis des mois une véritable fascination dans le monde entier et suscite de nombreuses interrogations et spéculations sur l'impact que le perfectionnement de tels outils pourrait avoir sur notre quotidien et sur nos sociétés.

Pour l'heure, l'IA générative est consciente de ses propres limites : lorsque la question « *Es-tu doté de défauts ?* » est posée à cet agent conversationnel, celui-ci répond de la manière suivante : « *En tant que programme informatique, je n'ai pas de sentiments, de personnalité ou de conscience de moi-même, donc je ne peux pas avoir de défauts au sens humain du terme. Cependant, il est important de noter que je ne suis pas parfait et que je peux parfois générer des réponses incorrectes, incohérentes ou incomplètes. Je peux également être influencé par les biais et les limitations des données sur lesquelles j'ai été formé. Il est donc toujours conseillé de vérifier les informations que je fournis et de consulter des sources supplémentaires pour obtenir une confirmation ou une perspective plus complète* ».

Si l'IA générative et les agents conversationnels ont créé un immense débat mondial, les questions qu'ils soulèvent sont loin de se résumer à cette forme particulière d'IA et ont vocation, au contraire, à s'appliquer à la technologie de l'IA en général, quel que soit son usage. En effet, c'est bien de la place de l'IA dans la société dont il est question et surtout de ses critères d'acceptabilité, et ce, quel qu'en soit l'usage considéré.

[1] Extrait d'un échange avec Chat GPT en juin 2023. Disponible sur le lien suivant : https://chat.openai.com/

L'IA est un domaine extrêmement technique, multiforme, polysémique et complexe, dont la compréhension globale n'est accessible qu'à très peu de spécialistes mais dont les manifestations, les effets ou les promesses sont visibles de tous. En ce sens, l'IA fait partie de ces sujets fortement clivants entre ses admirateurs béats et ses opposants farouches. Pour certains, l'IA peut à la fois incarner le futur de la santé et la promesse d'une médecine augmentée qui permettrait d'améliorer massivement notre espérance de vie, alors que pour d'autres, elle constitue la menace la plus importante pour l'Homme qu'elle serait tentée de supplanter ou d'éliminer dans un futur relativement proche. Sans pencher pour l'utopie ni pour la dystopie, beaucoup s'interrogent face aux résultats de premiers cas d'usages expérimentés et notamment sur le constat que l'IA semble accentuer des biais et discriminations de nos sociétés lorsqu'elle est nourrie des données de celles-ci. C'est en effet ce que relève l'étude *Gender Shade*[2] de la chercheuse du *MIT Media Lab* Joy Buolamwini de 2018. Cette dernière démontre que les systèmes de reconnaissance faciale utilisés par IBM et Microsoft aux Etats-Unis reconnaissent plus facilement un homme blanc (99,7% pour IBM et 100% pour Microsoft de taux de réussite) qu'une femme noire (respectivement 65,3% et 79,2%). Ici, le résultat fourni par la mise en œuvre de l'IA se comporte comme un révélateur des biais présents dans les données qui ont servi à entrainer l'algorithme : l'amplification, ou même simplement le report de ces biais peut entrainer des conséquences préjudiciables selon l'usage de l'IA considéré.

Les risques liés aux sujets des biais ne sont, là aussi, qu'un exemple de questions soulevées par l'IA. Pour les résoudre nous avons à nouveau sollicité un agent conversationnel : lorsqu'on lui demande les moyens existants permettant de corriger ses biais, il conclut qu'« *en fin de compte, une approche collective et continue est nécessaire pour faire progresser la lutte contre les biais dans les modèles de langage* ».

L'IA ne laisse plus personne indifférent et tout le monde a désormais un avis sur le sujet, y compris l'IA elle-même. L'objectif de ce livre blanc n'est pas d'ajouter une opinion supplémentaire dans le maelström ambiant mais plutôt de formuler une vision partagée par les entreprises qui mettent quotidiennement en œuvre l'IA dans différents aspects et différents usages souvent méconnus et loin de l'image d'Epinal d'une l'IA réduite aux robots conversationnels. Loin de vouloir éluder le débat, l'Alliance pour la Confiance Numérique (ACN) souhaite au contraire mettre l'expertise de ses membres au service du collectif pour essayer de réconcilier les promesses immenses de cette nouvelle technologie avec les risques réels ou perçus que son utilisation, sans encadrement satisfaisant, laisse présager. Au cœur de notre réflexion sur l'IA, se place la confiance. De la même manière que pour l'ensemble des technologies numériques, la confiance est ici la clef de voûte d'une appropriation des technologies. L'enjeu du groupe de travail de l'ACN dédié à l'IA est donc de contribuer à poser les premiers éléments de définition de ce que pourrait être une IA dite « *de Confiance* ».

[2] Extrait de l'étude Gender Shade. Disponible sur le lien suivant : http://gendershades.org/overview.html

En effet, considérant que cette confiance est un élément indispensable pour le développement des usages de l'IA, les entreprises du domaine souhaitent que les critères qui peuvent caractériser cette confiance soient débattus et partagés de sorte à pouvoir développer des usages dans un cadre bienveillant et sécurisé à la fois pour les entreprises qui le conçoivent et à la fois pour les utilisateurs.

Le postulat de départ de l'ACN est que la confiance implique de définir et, par la suite d'évaluer, un ensemble de caractéristiques, de principes et de pratiques qui assurent que l'IA est fiable, éthique, sécurisée, transparente et respectueuse des valeurs et des intérêts humains.
Afin de brosser une première esquisse de ces débats, l'ACN s'est attachée à aborder l'IA sous les angles juridique, technique, mais aussi éthique. Considérant que pour qu'une IA puisse être qualifiée comme étant de Confiance, elle devra répondre à des exigences dans ces 3 domaines distincts.

L'Union européenne s'est emparée de ce sujet en formulant, le 21 avril 2021, une proposition de règlement, l'*AI Act*. Il aura pour vocation d'apporter des réponses concrètes juridiques et éthiques en adaptant le cadre réglementaire et législatif existant autour de l'IA et en réduisant les incertitudes qui résultent aujourd'hui de ce cadre pouvant apparaitre comme inadapté au fil du temps. Ce faisant, l'Union européenne fait figure de pionnière et cette initiative doit être saluée. En adoptant une approche fondée sur les risques et non sur la technologie *per se* elle adopte une grille de lecture constructive et susceptible d'apporter de nombreuses réponses mais aussi une certaine confiance en fixant un cadre plus précis à l'utilisation de l'IA.

Au-delà de ce nécessaire cadre juridique, l'*AI Act* sera également l'occasion de replacer les principes et les valeurs fondamentales de l'Europe au cœur du développement de cette technologie, créant ainsi une différenciation avec les approches suivies dans d'autres régions du monde aux valeurs différentes de celles qui constituent le socle de l'Union européenne. L'ACN espère que cette initiative sera complétée par des référentiels techniques et des normes harmonisées, à élaborer, et à travers lesquels le consensus des experts pourra définir à la fois l'état de l'art de la technologie, mais aussi les moyens techniques d'attester de la sécurité et de la confiance que le volet légal appellera de ces vœux.
Ces moyens techniques pourront ainsi être mis en œuvre dès la conception des systèmes recourant à l'IA et apporter des réponses au sein des problématiques afférentes telles que la transparence, l'explicabilité, etc.

En France, la CNIL définit l'intelligence artificielle comme étant « tout outil utilisé par une machine afin de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité [...] incluant les comportements dépassant les capacités humaines[3] ».

[3] CNIL, 25 mars 2022, « Intelligence artificielle, de quoi parle-t-on ? ». Disponible sur le lien suivant : https://www.cnil.fr/fr/intelligence-artificielle/intelligence-artificielle-de-quoi-parle-t-on

Quant à la notion de confiance, elle résulte d'un ensemble de critères qu'il est nécessaire de définir précisément. Ces critères doivent également faire l'objet d'un consensus pour que cette confiance soit atteignable. C'est dans ce débat, visant à définir les critères permettant, par leur application, de caractériser une IA de Confiance et, par leur reconnaissance générale.

Ce livre blanc tentera d'apporter une contribution à ce débat en abordant dans un premier temps l'angle juridique, présentant les limites du cadre européen et français existant qui n'a pas été pensé pour l'IA actuelle, ses usages, ses applications et son contrôle. De nombreux acteurs économiques étant concernés par l'émergence des systèmes d'IA, ces limites induisent nécessairement de repenser un cadre cohérent, et concerté. Dans un second temps, les critères pouvant se rattacher à une IA de Confiance seront abordés sous un angle technique. La technologie d'IA se doit d'être transparente, interprétable et explicable. Elle doit également être compréhensible, entrainée sur des bases de données les moins biaisées possible et nettoyées des résultats erronés qu'elle pourrait proposer.

Enfin, dans une troisième partie, ce livre blanc abordera l'acceptabilité sociale et l'éthique de l'IA qui sont des conditions *sine qua non* d'une IA de Confiance. La primauté de l'humain et le respect des valeurs fondamentales de l'Europe, dont le respect des libertés publiques et individuelles et le respect de la vie privée, sont autant d'éléments cardinaux sur lesquels les entreprises de la filière de la confiance numérique entendent fonder une vision française et européenne. L'IA, au service de notre société, constituera alors un outil majeur pour à la fois défendre et porter nos valeurs et principes fondamentaux dans le monde.

PARTIE 1

LA NÉCESSAIRE ÉVOLUTION D'UN CADRE JURIDIQUE MAL ADAPTÉ

Il est coutume de constater que les technologies vont plus vite que le droit. Cet adage est particulièrement vrai en ce qui concerne l'IA et ses usages potentiels. La multiplication de ces derniers ainsi que la nature profondément disruptive des évolutions que l'IA laisse entrevoir, met à mal le cadre juridique général qui s'avère mal adapté pour adresser les nouvelles problématiques ainsi soulevées. L'application de dispositifs juridiques pensés pour encadrer l'espace numérique avant l'irruption de l'IA conduit à apporter des réponses inappropriées qui peuvent freiner l'innovation et la maîtrise de l'IA.
Le caractère singulier de l'IA réside dans le fait qu'en l'absence de maîtrise de cette technologie, c'est l'ensemble de notre avenir numérique qu'il deviendra impossible de protéger. Adapter notre corpus législatif afin qu'il permette le développement de l'IA est donc une condition pour pouvoir maîtriser notre avenir numérique et apporter à nos concitoyens une protection adaptée de leurs droits, notamment fondamentaux, dans l'espace numérique.

Plusieurs facteurs rendent difficile la relation entre l'IA et le cadre juridique général. Tout d'abord l'évolution technologique rapide de l'IA fait que les lois et les réglementations peuvent rapidement devenir obsolètes. Les législations existantes ne sont souvent pas conçues pour anticiper les développements futurs de l'IA. Par ailleurs, l'IA peut s'avérer complexe et opérer de manière opaque, ce qui rend difficile la création de règles juridiques spécifiques pour chaque situation.
Il peut être difficile pour les législateurs de comprendre pleinement le fonctionnement interne de certaines technologies d'IA et particulièrement s'ils ne sont pas accompagnés d'experts. Ainsi, les systèmes d'IA peuvent involontairement perpétuer, à travers leurs résultats, les biais existants au sein des bases de données sur lesquelles ils sont formés.

Les lois actuelles ne sont souvent pas équipées pour traiter ces questions de biais et de discrimination. Plus généralement, l'attribution de la responsabilité en cas de préjudice causé par une IA peut être complexe. Ces mêmes questions de responsabilité sont d'autant plus compliquées lorsque plusieurs acteurs (concepteurs, fournisseurs, utilisateurs) sont impliqués dans le cycle de vie de l'IA.

En raison de ces défis, il nous apparaît urgent de mettre à jour le cadre juridique européen et français pour mieux tenir compte de l'IA. Cela peut inclure la création de lois spécifiques à l'IA, l'élaboration de lignes directrices, de principes éthiques et la recherche de solutions pour traiter des questions telles que la détermination de la responsabilité, la transparence et la protection des données.

Il apparait donc nécessaire d'apporter une attention toute particulière à son cadre à venir afin que les défauts de la règlementation actuelle soient pris en compte et ses lacunes identifiées et comblées. Cet encadrement doit avoir comme objectif de trouver un équilibre entre la réglementation, les valeurs et droits humains et l'innovation dans le domaine de l'IA.

UNE PLURALITÉ DES CADRES DÉJÀ APPLICABLES À L'IA

Loin d'être absente des règlementations européenne et nationale, l'IA y est, au contraire, mentionnée à de nombreuses reprises. Elle se retrouve dans des textes dédiés à des thématiques variées allant du domaine de l'achat sur internet[4] à l'aviation[5]. Cependant, aucun texte à part entière aujourd'hui en vigueur ne s'est saisi de la question de l'IA de manière holistique, ce qui rend son appréhension par les fabricants, les utilisateurs, mais aussi les régulateurs, parfois complexe. Cette complexité contribue à contraindre le développement des systèmes d'IA en Union européenne et en France et laisse toujours plus de place à l'usage de systèmes d'IA non-européens venant se heurter au respect de nos valeurs fondamentales.

a) L'IA, sa définition et ses enjeux saisis par le droit

La protection juridique des données, personnelles et non personnelles, aujourd'hui en vigueur, appréhende difficilement la notion d'IA. Or, les systèmes d'IA ont pour but de traiter des données et d'en calculer un résultat afin de faciliter la prise de décision finale effectuée par l'Homme.
Si les données traitées sont à caractère personnel, les traitements sont couverts par le Règlement Général pour la Protection des Données (RGPD) de 2016[6], sinon, c'est le cadre européen de la donnée non personnelle qui s'applique à travers la directive *e-privacy* de 2002 ainsi que le règlement sur le libre flux des données à caractère non personnel dans l'UE de 2018[7].

[4] Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques. Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AOJ.L_.2022.277.01.0001.01.FRA&toc=OJ%3AL%3A2022%3A277%3ATOC
[5] Règlement (UE) 2018/1139 du Parlement européen et du Conseil du 4 juillet 2018 concernant des règles communes dans le domaine de l'aviation civile et instituant une Agence de l'Union européenne pour la sécurité aérienne. Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32018R1139
[6] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données). Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679
[7] Règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne. Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32018R1807

Le RGPD, conduit par le principe *d'accountability*, responsabilise d'une part chaque acteur qui traite des données à caractère personnel proportionnellement à son intervention et d'autre part, en fonction des cas d'usages, les fabricants, les fournisseurs et les utilisateurs finaux des systèmes d'IA (ci-après « l'ensemble des acteurs impliqués dans un système d'IA ») qui pourraient apparaître alternativement comme sous-traitant ou responsable de traitement en fonction de ces deux paramètres. Cependant, au moment de l'élaboration du RGPD, les systèmes d'IA étaient encore envisagés comme de « nouvelles technologies ».

A son article 35, le RGPD retient que pour un traitement, notamment par « le recours à de nouvelles technologies et compte tenu de la nature, de la portée, du contexte et des finalités du traitement, est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques, le responsable du traitement effectue, avant le traitement, une analyse de l'impact des opérations de traitement envisagées sur la protection des données à caractère personnel » .
Par conséquent, dès lors qu'un traitement de données à caractère personnel est exécuté par de nouvelles technologies au sens du RGPD et qu'il rencontre un des critères établis par le Groupe 29[9], ces traitements doivent nécessairement être notifiés à une autorité de contrôle, la Commission Nationale de l'Informatique et des Libertés (CNIL) en France.
Le recours à ces nouvelles technologies pour un traitement de données personnelles doit seulement faire l'objet d'une analyse d'impact relative à la protection des données (AIPD) afin d'évaluer la gravité et la vraisemblance de la survenance des risques (vol de données, modification de données par un accès illégitime, etc). Les AIPD étant souvent coûteuses et chronophages, elles constituent un frein supplémentaire au développement de cette technologie.

Les systèmes d'IA font pourtant bien partie du quotidien de certaines personnes et sont largement déployés / utilisés par de nombreuses entreprises comme en témoigne le groupe de travail sur l'IA de Confiance de l'Alliance pour la Confiance Numérique (ACN). Ainsi, l'IA n'est donc plus à considérer comme une nouvelle technologie au sens du RGPD.

[8] Article 35 du Règlement Général sur la Protection des Données (RGPD). Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679
[9] Lignes directives concernant l'analyse d'impact relative à la protection des données 5AIPD) et la manière de déterminer si le traitement est « susceptible d'engendre un risque élevé » aux fins du règlements (UE) 2016, 679, 4 octobre 2017. Disponible sur le lien suivant : https://www.cnil.fr/sites/cnil/files/atoms/files/wp248_rev.01_fr.pdf
et compte tenu de la nature, de la portée, du contexte et des finalités
du traitement, est susceptible d'engendrer un risque élevé pour les droits et libertés des
personnes physiques, le responsable du traitement effectue, avant le traitement, une
analyse de l'impact des opérations de traitement envisagées sur la protection des données
à caractère personnel »

De plus, l'exercice des droits des personnes concernées par les traitements au sens du RGPD et les algorithmes d'IA sont parfois complexes à accorder. Par exemple, il est difficile de faire valoir le droit à l'opposition d'une personne dans le cadre d'un traitement de données personnelles captées par des caméras de vidéosurveillance. Il est techniquement complexe de s'opposer à un traitement continu et en temps réel d'une caméra dont le périmètre est étendu, sans entraver la réussite des finalités poursuivies.

Toutes ces problématiques se rapportent également au champ des données à caractère non personnel, encadré par le règlement sur le libre flux des données à caractère non personnel dans l'UE de 2018. La protection de la vie privée est un objectif affiché de ces textes toutefois, ils ne mentionnent pas les systèmes d'IA en tant que tels : il en résulte un flou juridique important pour ces systèmes qui, sans être explicitement autorisés, ne sont pas pour autant interdits. Dans le même temps, la question de la responsabilité en matière d'IA doit également être traitée.

b) L'encadrement de la responsabilité : un cadre à adapter aux systèmes d'IA

Comme vu précédemment, les systèmes d'IA n'ayant pas encore fait l'objet d'un encadrement précis, ils sont traités par le cadre juridique déjà existant qui, national comme européen, n'appréhende pas toute leur réalité technique. Une place est cependant laissée à son expérimentation en France dans le cadre des Jeux Olympiques et Paralympiques de 2024 (JOP 2024)[10]. Son article 10 prévoit que la vidéosurveillance en temps réel soit autorisée, sous conditions, dans le cadre de cet événement de grande ampleur, sous la responsabilité de l'Etat. Cette expérimentation, qui reste limitée dans le temps pourrait être l'occasion de faire évoluer le cadre juridique de l'IA en France.

En droit commun et de manière synthétique, on peut dissocier la responsabilité subjective, qui implique la notion de faute réalisée, de son lien de causalité et du préjudice en découlant, de la responsabilité objective qui se rattache plus aux faits en cause. Ces principes juridiques permettent d'appréhender la responsabilité d'une action ou non-action d'une personne physique ou morale que ce soit dans une relation contractuelle ou non. Si une IA produit des calculs et propose une décision à l'Homme (comme un logiciel d'évaluation qui calcule le risque de récidive criminelle) la question se pose de savoir si elle peut être considérée comme responsable de ses calculs ? Ou bien devons-nous considérer que son fournisseur en est responsable ? Ou encore son développeur ? Ce sont des questions qui restent encore en suspens que les autorités de contrôle compétentes devront traiter.

[10] Loi n°2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions. Disponible sur le lien suivant : https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000047561989

Outre des règles juridiques contraignantes pouvant s'appliquer à l'IA mais ne la visant pas de manière spécifique, il existe des règles dites de « droit souple » (« *soft law* »). L'attente du secteur de l'IA en pleine émergence au niveau de la *soft law* est assez forte car elle permettrait de répondre aux problématiques laissées de côté par le droit ou non encore abordées. En ce sens, la CNIL en France, et le Comité européen de la protection des données en UE édictent quotidiennement des lignes directrices afin d'accompagner l'ensemble des acteurs impliqués dans un système d'IA dans l'interprétation des règlements et des directives européennes et du droit interne.

D'autres institutions ont adopté des recommandations éthiques au sujet de l'IA. Par exemple, l'UNESCO a adopté le 23 novembre 2021 des recommandations sur l'éthique dans l'IA où de nombreuses valeurs sont défendues : le respect, la protection et la promotion des droits de l'Homme, les libertés fondamentales et la dignité humaine. L'UNESCO y rappelle l'importance de prendre en compte ces valeurs dès la conception de solutions d'intelligence artificielle afin de limiter les potentielles dérives. En ce sens, le droit souple a son importance en ce qu'il permet de se saisir des nouveaux enjeux induits par l'IA comme la reconnaissance faciale ou la reconnaissance des émotions mais également de préparer les personnes concernées à l'adoption de l'IA. Ill ne suffit que d'une mention de ces écrits dans un texte de loi ou par un juge pour qu'ils obtiennent une force contraignante, notamment pour les autorités régulatrices.

c) Le rôle déterminant des entités régulatrices à pour un encadrement effectif (CNIL, CEPD, ...)

Les régulateurs de l'IA sont les autorités de contrôle qui peuvent émettre des recommandations thématiques, notamment sur l'IA mais aussi des référentiels, des lignes directrices ou encore des avis. L'ensemble des acteurs impliqués dans un système d'IA sont fortement encouragés à suivre et à appliquer ces recommandations.

Par exemple, en France, la CNIL a rendu public son avis sur la proposition d'*AI Act* (qui sera traité ci-dessous), ou encore sur la reconnaissance faciale et son cadre juridique d'expérimentation. La CNIL a notamment publié une position sur les caméras augmentées, à l'occasion de laquelle l'ACN a été consultée et a ensuite publié sa position[11].

[11] ACN, 14 mars 2023, Position sur la consultation publique portant sur le projet de position de la CNIL relative aux conditions de déploiements des caméras dites « intelligentes » ou « augmentées » dans les espaces publics. Disponible sur le lien suivant : https://www.confiance-numerique.fr/position-acn-sur-la-consultation-publique-portant-sur-le-projet-de-position-de-la-cnil-relative-aux-conditions-de-deploiement-des-cameras-dites-intelligentes-ou-augmentees

Par ailleurs, la CNIL a créé, le 23 janvier 2023 un service de l'intelligence artificielle (SIA) en son sein composé de juristes et d'ingénieurs spécialisés afin de « renforcer son expertise sur ces systèmes et sa compréhension des risques pour la vie privée »[12]. Ce service permettra notamment d'identifier les enjeux juridiques et éthiques des systèmes d'IA d'actualité comme l'IA générative ou les modèles de fondation.

Le Conseil d'Etat tend également à devenir un régulateur majeur pour l'IA, puisqu'il a publié le 31 août 2022 une étude souhaitant que la France s'engage dans le domaine de l'intelligence artificielle pour un meilleur service public[13] afin de l'assister dans ses tâches et d'améliorer sa qualité. Il expose au sein de son étude « la mise en œuvre d'une politique de déploiement de l'intelligence artificielle résolument volontariste, au service de l'intérêt général de la performance publique ». Il plaide également en faveur d'une stratégie de l'IA qui créera les « conditions de la confiance » à ce sujet afin de créer une IA publique de confiance. Enfin, le Conseil d'Etat se projette sur la proposition d'*AI Act*, en exposant son souhait que la CNIL devienne l'autorité nationale de contrôle des systèmes d'IA.

Au niveau européen, le Comité Européen de la Protection des Données (CEPD) émet également des lignes directrices quant à l'application des règles de droit sur les données, sur lesquelles se base la CNIL. En juin 2021, le Comité et le Contrôleur européen de la Protection des Données soulignent dans un communiqué de presse « la nécessité de clarifier explicitement que la législation européenne existante en matière de protection des données s'applique à tout traitement de données à caractère personnel relevant du champ d'application du projet de règlement sur l'IA »[14].

Les événements de grande ampleur comme la Coupe du monde de Rugby 2023 et les Jeux Olympiques et Paralympiques de 2024 sont l'occasion pour ces régulateurs d'accompagner les acteurs de l'IA dans leurs activités. La loi relative à ces Jeux, promulguée le 19 mai 2023 a permis de rendre concrète l'utilisation de certaines solutions d'IA.

En effet, l'article 7 prévoit et encadre l'expérimentation de la vidéosurveillance algorithmique pour la garantie de la sécurité des individus lors de cet événement. Ces événements seront également l'occasion pour l'UE de se saisir des enjeux de l'IA et d'adapter les textes en cours d'élaboration aux expérimentations françaises.

[12] CNIL, 23 janvier 2023, « Création d'un service de l'intelligence artificielle à la CNIL et lancement de travaux sur les bases de données d'apprentissages ». Disponible sur le lien suivant : https://www.cnil.fr/fr/creation-dun-service-de-lintelligence-artificielle-la-cnil-et-lancement-des-travaux-sur-les-bases-de

[13] Etude du Conseil d'Etat, 31 août 2022, Intelligence artificielle et action publique : construire la confiance, servir la performance. Disponible sur le lien suivant : https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance

[14] CEPD, 21 juin 2021, Communiqué de presse. Disponible sur le lien suivant : https://edpb.europa.eu/news/news/2021/edpb-edps-call-ban-use-ai-automated-recognition-human-features-publicly-accessible_fr

II. DES NORMES À PENSER ET À ADAPTER AU CONTEXTE TECHNIQUE DE L'IA

Depuis 2021, l'Union européenne tente de pallier les manquements juridiques mentionnés plus haut et de réguler ce secteur qu'est celui de l'IA. Plusieurs textes sont en cours de production et de discussion au sein des instances européennes afin d'élaborer un cadre autour de l'IA. La concertation au niveau régional est alors la garantie d'un référentiel cohérent, qui pourra par la suite être porté à l'échelle internationale et posséder la légitimité d'un réel outil d'influence.

a) Un cadre réglementaire européen en construction

En premier lieu, la Commission européenne a présenté, le 21 avril 2021, un projet de règlement harmonisant les règles concernant l'intelligence : l'*AI Act*. L'objectif n'est pas de réguler la technologie en elle-même, mais plutôt ses usages.

Il définit un système d''intelligence artificielle à son article 3 comme « *un logiciel développé à l'aide d'une ou plusieurs techniques et approches énumérées à l'annexe I de la proposition et capable, pour un ensemble donné d'objectifs définis par l'Homme, de générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels ils interagissent* » Le règlement établi une classification en trois catégories de risques afin d'appréhender les usages de systèmes d'IA :

- Les systèmes d'IA interdit,
- Les systèmes d'IA à haut risque et
- Les autres systèmes d'IA.

La proposition d'*AI Act* se saisit des nouveaux enjeux induits par l'utilisation de l'IA qui couvre aujourd'hui de nombreux aspects du quotidien. Ce règlement est complété par les propositions de directives (au nombre de 2) du Parlement européen et du Conseil sur la responsabilité en matière d'IA du 28 septembre 2022.

Un premier texte vient réviser la directive existante sur la responsabilité du fait des produits défectueux de 1985 en y incluant les systèmes d'IA. Le second texte, relatif à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'IA, vise à garantir que les victimes des dommages provoqués par des systèmes d'IA soient protégées.

Elle prévoit notamment d'alléger le niveau de charge de la preuve pour les victimes par l'introduction de la présomption de causalité réfragable, c'est-à-dire qu'il suffira à la victime de démontrer un lien « *raisonnablement probable* » entre l'erreur et le dommage. Elle vise également à assurer une transparence des acteurs afin que la victime puisse identifier les responsables plus facilement et accéder « *aux éléments de preuves pertinents* »[15], mais seulement pour les systèmes d'IA à haut risque. Selon la Commission, ces deux garanties devraient permettre d'instaurer plus de confiance dans les systèmes d'IA comme ils ne seront désormais plus abordés comme des « *boîtes noires* » qui seraient intouchables.

Les systèmes d'IA pourraient également être soumis à la proposition de *Data Act* publiée le 23 février 2022 qui tente de faciliter le transfert des données non personnelles entre entreprises, et du *Data Governance Act* entré en vigueur le 23 juin 2022 qui facilite le partage de données entre les secteurs d'activités et entre les Etats membres afin de permettre une meilleure exploitation des données dans un cadre de confiance.

L'UE pourrait également s'inspirer des travaux déjà existants au sein des Etats membres. Il pourrait notamment capitaliser sur la proposition de loi française visant à encadrer l'IA par le droit d'auteur qui viendrait compléter le code de la propriété intellectuelle et pourrait alimenter sa définition européenne.

Les textes européens en cours d'élaboration et touchant au sujet des systèmes d'IA sont donc nombreux. Une surveillance des articulations de ces textes est alors à exercer de manière constante afin de s'assurer qu'ils ne contredisent ou ne superposent pas afin d'obtenir un référentiel cohérent et concerté en matière d'IA.

[15] Proposition de directive du Parlement européen et du Conseil, 22 septembre 2022, article 3. Disponible sur le lien suivant :
https://eur-lex.europa.eu/legal- content/FR/TXT/HTML/?uri=CELEX:52022PC0496
[16] Proposition de loi n°1630 visant à encadrer l'intelligence artificielle par le droit d'auteur, 12 septembre 2023. Disponible sur le lien suivant : https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi

b) La nécessité d'un référentiel cohérent et concerté

Dans le cadre de la proposition d'*AI Act*, certaines limites du texte sont relevées par divers écosystèmes. Ainsi, certaines analyses soutiennent que la proposition d'*AI Act* doit permettre de garantir que les fournisseurs d'IA prennent en compte les dommages individuels que peut causer cette technologie, mais également les dommages sociétaux.
Ce pourrait être le cas, par exemple, d'applications pouvant influencer le comportement des votes des citoyens et influencer le résultat final d'une élection. Le préjudice ne s'apprécierait ici pas seulement au niveau individuel, mais également au niveau collectif (l'ensemble de la société dans l'exemple précité).

Une autre analyse suggère quant à elle de permettre de proposer des modifications de la liste des systèmes restreints à haut risque afin que la classification de la proposition d'*AI Act* suive et s'adapte aux évolutions technologiques et à leurs usages.
Dans le domaine industriel, certaines organisations proposent que les PME participent activement à l'élaboration de ces normes afin de pouvoir plus facilement les appliquer. Ces analyses permettent alors de remonter les réalités techniques de l'IA et d'assurer la concertation. Par conséquent, la Commission européenne, le Conseil de l'UE et le Parlement européen doivent être en mesure de prendre en compte les remarques émanant des écosystèmes concernés afin d'assurer la cohérence de ce riche corpus de ce texte.

Il est également nécessaire de s'assurer que les textes ne se chevauchent et ne se contredisent pas et pour cela, il faut s'assurer d'un vocabulaire commun entre ces différents textes. Le Conseil d'Etat dans son étude publiée le 30 août 2022 mentionnée plus haut souligne « la très forte adhérence entre la régulation des systèmes d'IA et celles des données, en particulier des données à caractère personnel ». Dans sa première partie, le Conseil d'Etat souligne l'importance de construire un langage commun et intelligible de l'IA. L'ACN salue la cohérence linguistique entre la proposition d'*AI Act* et les directives sur la responsabilité en matière d'IA qui mentionne tous des systèmes d'IA et distinguent les systèmes aux risques inacceptables à hauts risques les autres systèmes.

Un référentiel cohérent et concerté constituera un socle solide pour une réglementation européenne pertinente. Au niveau international, il est d'ores et déjà possible de se tourner vers la nouvelle norme ISO/IEC 42001 « *Technologie de l'information, intelligence artificielle, système de management* » conçue pour permettre aux organismes proposant ou utilisant des produits et des services faisant appel à une IA de veiller au développement et à l'utilisation des systèmes d'IA de manière responsable. Elle propose une méthodologie structurée pour gérer les risques et les opportunités associés à l'IA.

De plus, cette harmonisation entre normes européennes et internationales des systèmes d'IA permettrait de donner une force contraignante aux textes techniques déjà développés, qui n'auront force contraignante qu'à partir du moment où ils auront été mentionnés par des textes contraignants. Une approche collective internationale permettrait surtout de réduire davantage les risques liés aux systèmes d'IA et de garantir le respect des droits et des libertés fondamentales. Afin que le référentiel européen sur les systèmes d'IA puisse être élevé au rang de norme internationale, il convient de définir ce qui compose et garantit l'intelligence artificielle de confiance.

La définition juridique de l'IA posée plus haut permet d'appréhender les logiques de la conception et de l'encadrement de l'IA. Le cadre européen de la donnée tout comme le cadre national de la responsabilité sont mal adaptés pour appréhender les traitements de systèmes d'IA. Les traitements engendrés n'entrent pas dans les catégories existantes, ce qui signifient qu'elles doivent être repensées tout comme les modalités d'exercice de droits. Certains droits comme le droit d'opposition doivent en effet être réadaptés afin qu'ils puissent pleinement s'appliquer au contexte de l'IA. Le cadre français de la responsabilité peut quant à lui continuer à s'appliquer à condition que la question de la responsabilité des systèmes d'IA soit tranchée par les régulateurs.

Par conséquent, les enseignements tirés de l'analyse des textes encadrant l'IA doivent être intégrés dans la réglementation à venir en matière d'IA au risque de voir l'industrie française et européenne de l'IA désavantagée dans un contexte de concurrence internationale. Pour cela, le nouveau cadre doit être construit en concertation de l'ensemble des acteurs concernés afin d'assurer sa cohérence avec la réalité pratique et qu'il puisse constituer les bases solides d'un encadrement pertinent, protecteur des valeurs et libertés fondamentales et propice à l'innovation et à la maîtrise des technologies d'IA par les acteurs européens. Dans cette logique de cohérence, la définition juridique de l'IA doit être complétée par une définition technique, qui s'entend ici comme les objectifs auxquels les composants technologiques doivent répondre. En d'autres termes, l'IA de Confiance dépend tout autant de son cadre juridique que de sa composition technologique qui doit nécessairement devenir compréhensible pour être admise.

PARTIE 2

L'APPLICATION TECHNIQUE DE LA CONFIANCE DANS L'IA : UNE TECHNOLOGIE TRANSPARENTE ET COMPRÉHENSIBLE

Permettre le développement des systèmes d'IA par un cadre juridique favorable est le premier des critères de la confiance. Mais il doit être complété par des critères éthiques objectifs définis afin de garantir que ces systèmes d'IA ont été développés conformément à nos valeurs européennes. Ces approches juridique et éthique doivent néanmoins être confortées par une approche technique afin de rendre les systèmes d'IA lisibles et compréhensibles, tout en offrant des garanties à la fois de sécurité et de confiance aux utilisateurs.

Ce livre blanc considère que la confiance peut être rattachée aux systèmes d'IA à condition qu'ils offrent la transparence nécessaire à la compréhension de la conception du système intelligent de sorte que les résultats qu'il proposent soient interprétables et surtout explicables. Aussi, ces systèmes doivent être entrainés dans des conditions réelles, c'est-à-dire avec des bases d'entrainement comportant le moins de biais possible pour assurer la précision des résultats. Enfin, la confiance est également garantie par l'application d'exigences de cybersécurité dès la conception du système d'IA et durant ses phases d'entrainement et opérationnelle.

UNE NÉCESSITÉ DE TRANSPARENCE, D'INTERPRÉTABILITÉ ET D'EXPLICABILITÉ

La mise en œuvre d'une intelligence artificielle dite de confiance passe par la nécessité d'une approche éthique globale allant de sa conception jusqu'à son utilisation en passant par son développement, une indispensable robustesse technologique accompagnée d'un principe de transparence de plus en plus décisif. Cette transparence contribue indéniablement à la concrétisation de la confiance, cela sur l'ensemble du cycle de vie de la solution d'IA : de la création à la mise sur le marché, du choix des données d'entraînement jusqu'au déploiement en conditions réelles et à grande échelle.

On envisage alors des sous-principes de loyauté, d'équité, d'interprétabilité ou d'explicabilité. La connaissance seule des algorithmes ne suffit pas à caractériser la transparence, qui fait également intervenir des notions d'accessibilité universelle, d'information, de conscience des interactions avec la technologie, de compréhension des mécanismes de prise de décisions et de maîtrise totale des fonctionnalités.

a) Apporter la connaissance et l'accompagnement sur l'utilisation des solutions d'intelligence artificielle

La notion de transparence, lorsque l'on parle d'algorithmes, passe par l'usage mais doit aussi se faire par la vulgarisation et l'appropriation de la technologie. Mettre à disposition, lorsque le droit de la propriété intellectuelle et le secret des affaires l'autorisent, un code source, probablement indéchiffrable pour une grande majorité d'entre nous, ne suffit pas. C'est par la compréhension du fonctionnement intrinsèque de l'algorithme que passera l'acceptation de l'IA. La peur de l'inconnu, tout à fait humaine et compréhensible, est très probablement le premier facteur responsable de nombreux freins au déploiement de l'Intelligence Artificielle.

Les fournisseurs d'IA de Confiance se donnent pour mission de rendre accessible et compréhensible la construction de leurs modèles. De cette manière, les utilisateurs finaux et personnes concernées seront par eux-mêmes capables d'appréhender les risques liés à l'utilisation de ces mêmes algorithmes. On envisage ici la construction technique mais aussi les possibilités de configuration et bien sûr les restrictions apportées volontairement par les concepteurs pour protéger et sécuriser les utilisateurs et les personnes concernées.

Il existe différentes manières ou stratégies techniques pour construire un algorithme avec une aspiration fonctionnelle équivalente. Les fournisseurs d'IA de Confiance ont pour objectif, dès cette phase de conception, de prendre en compte les différentes problématiques de respect des droits et libertés fondamentales et ainsi faire en sorte que leurs algorithmes soient pensés pour ne pas les entraver. On parle alors de *privacy by design & by default*. Ces deux notions fondamentales à la réussite d'une IA de Confiance, déjà présentes dans la réglementation RGPD, doivent être guidées par une éthique bienveillante de ses concepteurs.

Pour illustrer notre propos, prenons par exemple deux algorithmes distincts de *computer vision* qui serviront tous les deux à compter des personnes présentes à un endroit déterminé :

- L'un peut le faire simplement en considérant les silhouettes d'humains comme des objets passant dans l'image, et ensuite agréger à chaque passage un compteur anonymisé.

- Quand l'autre peut le faire grâce à la reconnaissance de visages menant à la sauvegarde de résultats et d'images contenant des données à caractère personnel superflues.

Ces deux modèles poursuivent le même objectif de comptage de personnes, seulement leur conception éthique entraine des conséquences très différentes tant pour les concepteurs que pour les utilisateurs. Dans ce cas, il est pourtant possible d'atteindre les mêmes finalités avec la même performance tout en assurant le respect de nos valeurs fondamentales, ici, le droit au respect de la vie privée grâce à des moyens proportionnés à la poursuite de l'objectif recherché. La proportionnalité du moyen utilisé permet une interprétation équilibrée du résultat donné, qui permet par la suite sa compréhension et donc sa capacité à être expliqué.

De ce fait, communiquer et donner à voir la complexité et les nuances, aujourd'hui masquées par la technique, est la clé d'une transparence qui permettra à tous d'envisager l'IA sous un angle serein et propice à l'innovation. Pour cela, un travail de sensibilisation et de pédagogie autour de cette technologie est à mettre en œuvre et à entretenir afin que la perception de l'IA ne soit pas diabolisée et qu'il soit clairement identifiable que les moyens utilisés sont proportionnés à son utilité, réflexion que ce livre blanc initie.

b) IA « for good », lutter contre la diabolisation et la désinformation

La lutte contre la désinformation commence par la sensibilisation, et pour cela, il faut donc que les systèmes d'intelligence artificielle puissent être expliqués et vulgarisés à la fois aux partenaires commerciaux et aux utilisateurs. La notion d'explicabilité est centrale en ce qu'elle permet d'exprimer le choix des facteurs sélectionnés de manière compréhensible par les humains. Chaque facteur est interprété dans son contexte d'utilisation permettant d'assurer qu'il soit compris et donc transparent.
Si l'on reprend l'exemple cité plus haut, le facteur choisi pour le comptage de personnes dans un lieu prédéterminé sera la reconnaissance d'une silhouette humaine afin de reconnaitre une personne. Si l'on choisit de reconnaitre les visages, alors l'interprétation du choix de ce facteur principal dans le contexte du comptage ne permet pas d'expliquer les données relevées. Sa transparence n'est donc pas garantie car le moyen utilisé est disproportionné au but poursuivi ce qui peut laisser sous-entendre qu'un autre objectif que celui annoncé est ou pourrait être poursuivi.

Comme tout outil, il est de la responsabilité de ses créateurs et de ses utilisateurs de prendre en compte les enjeux de ses performances et d'en faire un outil altruiste, « *pour le bien* ». L'objectif d'une IA de Confiance est d'utiliser les capacités de cette technologie pour aider, soulager, augmenter l'humain dans des tâches complexes ou ingrates, sans le remplacer, et surtout sans porter préjudice aux personnes concernées par son utilisation.

Pour un éditeur de logiciel cela passe par de nombreuses phases de pré-conception issues des besoins exprimées par les clients (finalités d'usages) : discussions éthiques entre les acteurs du développement, choix déontologique des orientations stratégiques et des partenaires, travail conjoint entre les équipes produits et juridiques. Ces phases sont indispensables à l'émergence de cas d'usages réfléchis et cohérents avec les principes fondateurs de la confiance et de la transparence.

Cette démarche mène à la création d'intelligences artificielles dites « pour le bien », altruistes, et donc beaucoup plus à même d'être acceptées et transparentes dans l'utilisation qui en sera faite.

D'ailleurs, il est à noter que cette construction du bien se fait avec le filtre culturel, moral et sociétal des concepteurs. Il est alors d'autant plus important de porter le développement et l'affirmation d'une intelligence artificielle souveraine et garantir que ces systèmes d'IA respectent les valeurs fondamentales européennes contrairement à d'autres plus éloignés géographiquement et éthiquement de ces valeurs.

Ainsi, l'intelligence artificielle est aujourd'hui déployée pour de nombreux cas d'usages, autres que sécuritaires, dans des domaines aussi variés que le développement durable, la santé, prévention des risques climatiques, etc...

C'est pourquoi, il est primordial d'améliorer notre connaissance collective de l'intelligence artificielle, et de faire connaitre ses multiples usages et leurs apports.

Toutefois, au-delà de cet effort de transparence et d'explicabilité, une exigence particulière doit être portée à la phase d'apprentissage des algorithmes, dans la mesure où cette phase détermine les résultats que produira l'intelligence artificielle concernée. Il est notamment indispensable que les modèles d'apprentissage utilisés soient fidèles à la réalité du terrain sur lequel elles auront vocation à fonctionner. L'entrainement par des bases de données d'apprentissages fiables apparait donc comme une donnée *sine qua non* de sa fiabilité, de son attribut de confiance.

II. LA PRÉCISION ET LA FIABILITÉ DES RÉSULTATS DE L'IA

Afin d'obtenir des systèmes d'IA les plus performants possibles, il est nécessaire, de même que pour le cerveau humain, de les entrainer avec de grands volumes de données. Pour que le résultat fourni par les systèmes d'IA soit le plus fiable possible, il faut que ces données reflètent la réalité du contexte dans lequel ces IA vont opérer. Ces données sont à la fois issues de l'humain et relevées par ses soins, le plus souvent par ce qu'on appelle l'annotation. Ces données comportent alors des défauts. Afin d'assurer la confiance, il est nécessaire de réduire les risques liés à l'apprentissage des défauts et biais contenus dans les bases de données qui servent à entrainer les systèmes d'intelligence artificielle afin de limiter les résultats erronés.

a) Un nécessaire entrainement sur des bases de données fiables

Préalablement à sa mise en circulation, un système d'IA doit suivre plusieurs étapes afin de s'assurer qu'il ne présente pas de risques. Pour que les algorithmes soient entrainés, ils doivent traiter de larges volumes de données. En ce sens, deux méthodes existent :

- Les systèmes d'IA supervisés,
- Les systèmes d'IA non supervisés.

Une fois ces bases de données construites pour l'entrainement, la validation et le test d'un système d'IA, elles sont confiées aux algorithmes afin qu'ils apprennent à reconnaitre les données et fournir des prédictions. Pour cela, ces données sont étiquetées lorsqu'elles sont supervisées, c'est-à-dire qu'on leur donne un titre afin qu'elles soient classées, rangées dans des catégories enregistrées par l'algorithme. Plus l'algorithme est entrainé avec un panorama de données variées et rejouables et plus sa précision sera garantie.

L'étiquetage des données est une étape primordiale. Les bases d'apprentissage doivent donc être quantitatives, qualitatives et transparentes afin que les résultats produits par les algorithmes puissent être expliqués et interprétés.

Il est à souligner que l'entretien de la conformité après l'acquisition de tels volumes de données peut parfois être techniquement compliqué. Il est souvent nécessaire de faire appel à un fournisseur pour obtenir des ensembles de données, ce qui peut être un frein à leur développement, notamment à cause de leur prix.

Une fois les bases de données créées et l'apprentissage accomplis, une phase de test où de nouvelles données sont injectées afin d'évaluer l'entrainement de l'algorithme est lancée. Cette évaluation est annotée puis corrigée afin d'obtenir un résultat le plus précis possible. Les anomalies sont identifiées pour améliorer et limiter les potentiels impacts négatifs sur les résultats proposés par l'algorithme.

L'efficacité de l'IA réside principalement dans sa capacité d'entrainement. Il est donc extrêmement important d'être en capacité de certifier les bases d'entrainement. Les technologies d'apprentissage IA sont très dépendantes des bases qui leur servent d'apprentissage. A minima, il faut créer des standards et des processus de qualification de ces bases d'apprentissages qui permettraient d'assurer la fiabilité des bases d'apprentissage en matière d'intégrité et de confidentialité et donc de réduire le risque qu'un attaquant empoisonne ou altère ces jeux de données à des fins de détournement. Ces standards permettraient aussi d'en assurer l'éthique afin de garantir que des biais d'apprentissage discriminants (parité, couleur de peau, ...) ne sont pas introduits dans ces apprentissages. Pourtant, ce sont bien ces bases de données discriminantes et les données sensibles qui permettent de détecter ces biais et de rééquilibrer les bases de données. Il est donc opportun de contrôler la création de bases d'apprentissage en produisant des standards et des processus de qualification et s'assurer qu'elles sont protégées. Il apparaît également nécessaire que l'Europe puisse se doter de son propre organisme de certification des algorithmes biométriques, d'un point de vue sécuritaire mais aussi éthique. Aujourd'hui, seul le NIST (*National Institute of Standards and Technology*) américain est reconnu au niveau international.

b) Une attention particulière à porter aux audits pour garantir l'équité et la non-discrimination

Selon plusieurs études, dans la majorité des cas, les craintes exprimées par le public au sujet de l'IA sont liées au thème de la protection de la vie privée, suivi par le thème des biais algorithmiques (environ 30% des cas) puis le manque d'explicabilité (14%)[17]. Il faut comprendre le biais algorithmique par un défaut dans la collecte, l'analyse, l'interprétation, la publication ou la revue des données d'entrainement, de validation et/ou de test induisant des résultats erronés.

[17] CapAI, 25 mars 2022, "A procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act". Disponible sur le lien suivant : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091

Le biais peut être introduit dans un algorithme par l'acquisition de données et ne pas représenter une partie de la population, manquer de diversité ou bien reporter les discriminations historiques de nos sociétés en son sein. Le biais peut également être introduit via ses interactions avec son utilisateur qui peut lui-même être influencé par autrui, par la manière dont la situation est présentée ou encore par le temps.

Les conséquences de ces éventuels biais sont d'autant plus importantes que le niveau de risque de l'application considérée, au sens de la proposition d'*AI Act* de la Commission européenne, augmente. Par exemple, si un système d'IA pour une application triant des *Curriculum Vitae* à des fins de recrutement comporte des biais au niveau du genre en ne classant les profils féminins qu'à des postes d'assistantes de direction, une partie des personnes concernées par ce système d'IA, ici les femmes, est discriminée. Un autre exemple, souligné par une étude, évoque les disparités de performances des systèmes de reconnaissance faciale étudiés dans ce document qui s'avéraient plus précis lorsqu'ils identifient des hommes blancs alors qu'ils avaient plus de difficultés à reconnaitre des femmes noires[18]. Cette discrimination est du au fait que ce système d'IA a été principalement entrainé sur des visages d'hommes blancs.

Cependant, qualifier les biais d'algorithmes conduit à nommer l'algorithme comme responsable en cas d'erreur. En réalité, l'algorithme ne dépend que de son programmeur et des données utilisées pour l'entrainement de son algorithme.

Il est souvent impossible de produire un jeu de données idéal exempt de tout biais (comme une répartition parfaite quel que soit le genre, la religion, les orientations sexuelles, ...) comme ce n'est pas le cas dans notre société. De fait, les biais déjà existants dans la société sont alors reportés dans nos algorithmes qui peuvent alors se transformer en catalyseur de discriminations.

Afin de résoudre ces problématiques, il est nécessaire de nettoyer les données afin d'éliminer autant que possible les discriminations qu'elles transportent. Les modèles sont également à tester afin de vérifier leur « justesse ». Une autre réponse peut être l'utilisation de données synthétiques. Celles-ci permettent de contrôler la génération du contenu en solutionnant par exemple les déséquilibres entre certains groupes de population. Par ailleurs, c'est une solution qui semble être moins coûteuse que de longues tâches manuelles de préparation de données réelles. Toutefois, les données réelles restent nécessaires, notamment pour la phase de validation, les données synthétiques n'apportant pas une réponse totale au sujet des biais. En effet, ces bases de données sont représentatives d'une situation donnée mais ne pourront pas l'être de l'une à l'autre.

[18] Broutonlab, « Comment les données synthétiques peuvent-elles résoudre le problème du bais de l'IA ». Disponible sur le lien suivant :
https://broutonlab.com/blog/ai-bias-solved-with-synthetic-data-generation

c) Des systèmes d'IA protégés des cyberattaques

Pour assurer la fiabilité des systèmes d'IA et particulièrement celle de leurs bases de données d'entrainement, il faut qu'ils puissent être protégés de cyberattaques malveillantes à leur encontre. En effet, les bases de données d'entrainement peuvent faire l'objet d'empoissonnement ce qui modifierait le comportement d'un système d'IA. Dans ce cas, l'attaquant est capable d'injecter ou d'altérer des données dans la base d'entrainement et l'a donc corrompue afin que les résultats fournis soient erronés. Plus largement, le modèle d'un système d'IA peut aussi faire l'objet d'une exfiltration, que ce soit sur les paramètres du modèle (poids) ou sur les données ayant servi à entrainer ce modèle. En effet, le modèle et ses paramètres, qui sont les éléments différenciants et à forte valeur dans un système d'IA, peuvent être récupérés et ainsi améliorer la connaissance du système par un attaquant.

En complément de la phase d'entrainement, il est également important de protéger la phase opérationnelle, dans laquelle un modèle entrainé est déployé en environnement de production. Les entrées et sorties de ces environnements de protection doivent donc être contrôlées et les injections de données surveillées. Pour cela, il est essentiel d'appliquer des mesures de cybersécurité.

Les premières d'entre elles sont les mesures d'hygiènes développées par l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI)[19]. Ces règles constituent les premiers gestes à appliquer sur ses systèmes d'information qui trouvent une application aux systèmes d'IA. Plus spécifiquement, il est recommandé, pour un système d'IA, de s'assurer de la protection des environnements de développement, d'entrainement et de production, d'encadrer le contrôle des droits d'accès ou encore de favoriser l'usage de langages de développement sécurisés et de modules de logiciels tiers évalués et validés.

Dans le contexte particulier des IA génératives, il est également nécessaire de s'assurer du contrôle de la non-innocuité des entrées et des sorties des utilisateurs et de leurs interactions avec le système d'IA ou avec d'autres applications bureautiques. Par conséquent, il est nécessaire de procéder à une analyse de risques des systèmes d'IA le plus en avance de phase possible afin d'identifier la menace potentielle, les vulnérabilités des systèmes et les mesures adéquates à appliquer pour réduire le plus possible les risques d'attaques.

[19] ANSSI, Guide d'hygiène informatique, 23 janvier 2027. Disponible sur le lien suivant : https://cyber.gouv.fr/publications/guide-dhygiene-informatique

Nous l'avons vu, l'intelligibilité et la fiabilité d'un système d'IA garantissent sa transparence. En étant portées à la connaissance du public, elles permettent de mener un débat public plus éclairé autour des enjeux de l'IA comme le fait la CNIL en publiant ses fiches pratiques sur la constitution de bases de données d'apprentissage[20]. Ce débat éclairé et nécessaire autour des sujets sociétaux importants que soulève l'IA doit être mené dans nos sociétés, afin que cette technologie et ses utilisations puissent être socialement acceptées. Pour ce faire, l'éthique doit être au cœur de tous les développements et usages de l'IA.

[20] CNIL, Quel est le périmètre des fiches pratiques sur l'IA ?, 11 octobre 2023. Disponible sur le lien suivant : https://www.cnil.fr/fr/quel-est-le-perimetre-des-fiches-pratiques-sur-lia

PARTIE 3

L'ACCEPTABILITÉ SOCIALE ET L'ÉTHIQUE DE L'IA, PARENTS DE LA CONFIANCE DANS LA TECHNOLOGIE

Comme vu précédemment, l'IA est aujourd'hui un champ largement investi par les entreprises françaises et les cas d'usages proposés par ces acteurs rencontrent, par la valeur ajoutée qu'ils créent, de nombreux succès commerciaux. Le secteur est composé de diverses entreprises, grands groupes, PME et start up extrêmement innovantes, agiles, et positionnées à la pointe de l'état de l'art technologique.

Toutefois, le développement du secteur de l'IA est freiné sous le double effet de l'absence de cadre juridique adapté et d'une perception, par le grand public, souvent réductrice, ou déformée de cette nouvelle technologie, généralement liée à des exemples issus d'autres pays ne partageant pas nécessairement les valeurs fondamentales européennes.

Pour tenter de répondre à cette problématique, l'UNESCO (précurseur en la matière) ou encore la Commission européenne pour l'efficacité de la justice (CEPEJ) du Conseil de l'Europe a adopté le premier texte européen énonçant des principes éthiques relatifs à l'utilisation de l'IA dans les systèmes judiciaires. La Charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires et leur environnement adoptée par la CEPEJ les 3 et 4 décembre 2018 prévoit que la mise en œuvre de l'IA doit se faire de manière responsable en conformité avec les droits fondamentaux garantis notamment par la Convention européenne des droits de l'Homme (CEDH) et la Convention du Conseil de l'Europe pour la protection des données à caractère personnel. L'IA doit rester un outil au service de l'intérêt général. La CEPEJ identifie 5 principes éthiques à respecter en matière d'IA[20] :

- Le principe de respect des droits fondamentaux dès la conception des systèmes d'IA ;
- Le principe de non-discrimination ;
- Le principe de qualité et de sécurité, soit d'utiliser des sources certifiées et des données intangibles dans un environnement technologique sécurisé ;
- Le principe de transparence, neutralité et intégrité intellectuelle en rendant accessible et compréhensible les méthodologies utilisées ;
- Le principe de maitrise par l'utilisateur en bannissant l'approche prescriptive et permettre un choix maitrisé et éclairé.

[20] CEPEJ, 3-4 décembre 2018, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement. Disponible sur le lien suivant: https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b

Cette charte qui, pour avoir une force contraignante, doit être mentionnée dans un texte a portée législative ou réglementaire, vient compléter l'édifice de *soft law* évoqué précédemment.

A ces cinq principes énoncés par la CEPEJ, l'ACN propose d'ajouter quelques notions supplémentaires, considérant que l'acceptabilité sociale de l'IA doit être recherchée en couvrant l'ensemble du scope des préoccupations légitimes de notre société.

LES PRINCIPES SUPPLÉMENTAIRES À VALORISER

L'ACN propose de considérer également, en plus des principes énoncés plus haut, les principes de la primauté de l'humain, la nécessité de performance ainsi que le principe de respect de l'environnement.

a) L'humain au cœur de l'IA

L'IA doit avant tout être pensée et conçue pour fournir à l'Homme les éléments utiles à la prise de décision éclairée et non pas dans un objectif de prise de décision finale et de manière automatisée sans intervention humaine. En ce sens, l'ACN est en accord avec les propos de la Charte[21] en ce que :

- L'autonomie de l'utilisateur doit être renforcée et ne pas être restreinte par l'utilisation d'outils et de services d'intelligence artificielle ;
- Toutes prédictions par l'algorithme doivent pouvoir être remontées, c'est-à-dire qu'il est nécessaire qu'elles puissent être expliquées et comprises mais également interprétables. Dans le même sens, les possibilités de s'écarter de la décision proposée doit toujours être permise ;
- L'utilisateur doit pouvoir être informé dans un langage clair et compréhensible du caractère contraignant ou non des solutions proposées pour les outils d'IA et des différentes options possibles. Les personnes concernées doivent également être informées de tout traitement par une intelligence artificielle de leurs données et être en mesure de s'y opposer ;

Pour autant, cela ne signifie pas qu'il ne peut pas y avoir de décisions automatisées, qui reste très utiles à l'Homme dans ses tâches les plus répétitives. Cela signifie qu'un contrôle par l'humain doit cependant toujours être exercé sur ces outils afin de limiter les potentielles discriminations qui pourraient en découler.

[21] CEPEJ, 3-4 décembre 2018, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, p12. Disponible sur le lien suivant: https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b

De manière générale, une sensibilisation des utilisateurs doit être entreprise par les pouvoirs publics sur l'usage de l'intelligence artificielle. La multiplication des usages des systèmes d'IA conduit à l'augmentation des risques perçus de manière homothétique aux nouveaux usages de ces technologies. Cette sensibilisation est essentielle dans la mesure où l'IA est de plus en plus présente dans notre vie quotidienne, que ce soit à travers les médias sociaux les soins de santé, les véhicules autonomes ou d'autres domaines. Sensibiliser le public à l'IA permet de mieux comprendre comment cette technologie fonctionne, ce qu'elle peut accomplir et quelles sont ses limites. L'objectif est de permettre à chacun de prendre des décisions éclairées concernant l'utilisation de l'IA. Cela inclut des décisions sur la protection de la vie privée, la sécurité en ligne, l'adoption de produits ou de services basés sur l'IA.

Sans une compréhension approfondie de l'IA, la confiance du public dans cette technologie ne pourra pas se développer. Le climat actuel de méfiance généralisée envers l'IA pourrait contribuer à entraver son adoption, même pour les usages pour lesquels ses avantages potentiels sont évidents. L'IA est souvent mal comprise ou perçue comme mystérieuse. La sensibilisation peut aider à démystifier cette technologie, montrant qu'elle repose sur des concepts accessibles et compréhensibles mais bien encadrée et maitrisée par des experts.

Une sensibilisation accrue du public aux implications éthiques de l'IA, permettrait par ailleurs d'encourager la réflexion et le débat public sur les questions telles que les biais algorithmiques, la discrimination, la vie privée et la responsabilité en cas d'erreur ou de préjudice causé par une IA évoqués précédemment. Une compréhension plus approfondie de l'IA pourrait également contribuer à aider les citoyens à détecter les informations trompeuses ou biaisées générées par des systèmes d'IA, contribuant ainsi à lutter contre la désinformation.

En résumé, mettre l'humain au cœur de l'IA signifie adopter une approche éthique, transparente et responsable dans la conception, le déploiement et l'utilisation de l'IA, en veillant à ce qu'elle serve les intérêts et les valeurs humaines tout en minimisant les risques et les impacts négatifs. Cela nécessite une collaboration entre tous les acteurs (gouvernements, entreprises, chercheurs, société civile, ...) pour élaborer des politiques et des pratiques qui favorisent une IA éthique et centrée sur l'humain.

b) La nécessité de performance

Dans un second temps, une IA de Confiance doit être tournée vers un objectif de performance. En d'autres termes, cela signifie que les moyens utilisés par l'IA doivent être proportionnés à l'objectif qu'elle cherche à atteindre tout en permettant de réduire les risques. Il ne doit donc pas exister de moyens plus simples et moins risqués, pour les droits fondamentaux des personnes, d'atteindre son objectif. Pour reprendre l'exemple mentionné plus haut, le comptage de personnes dans un espace public peut se faire par un système de vidéosurveillance ou simplement par une personne physique.

Dans les 2 cas, l'objectif est atteint. La question à poser ici est celle de l'intérêt : la vidéosurveillance me permet-elle d'avoir des résultats plus précis ? Son installation est-elle proportionnée ? Le recours à la vidéosurveillance peut apparaître disproportionné par rapport à un simple objectif de comptage qui nécessite des moyens bien plus conséquents et intrusifs qu'une personne physique. A l'inverse, si cette tâche semble trop complexe pour une personne physique et permet surtout de meilleurs résultats, alors l'usage de l'IA parait approprié.

Plus globalement, les droits européen et français considèrent tous deux les principes de performance et de proportionnalité. Les juges doivent contrôler que l'atteinte qui a été portée à un droit fondamental n'est pas disproportionnée. Pour cela, il importe de vérifier d'abord si le but poursuivi est légitime, atteint, et si une autre mesure, moins intrusive, aurait pu être utilisée de manière tout aussi efficace. Il s'agit d'opérer une sorte de balance des intérêts en présence entre les libertés ou les droits susceptibles d'être atteints et de l'objectif poursuivi.

En matière d'intelligence artificielle, ce contrôle doit également être exercé. Le principe de performance doit être mis en œuvre dès la conception de systèmes d'IA qui doivent démontrer l'apport que cette technologie entraîne au regard du but poursuivi.

c) Une technologie environnementalement acceptable

L'impact du numérique sur l'environnement s'impose comme une thématique majeure qu'il convient également d'intégrer dans la réflexion autour de l'IA. En effet, l'Agence de l'environnement et de la maîtrise de l'énergie (Ademe) en collaboration avec l'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (Arcep) attribue, dans un rapport sur l'impact environnemental du numérique en France, 2,5% des émissions de gaz à effet de serre en France au secteur du numérique[22]. L'IA n'en représente qu'une partie, certes appelée à croître, proportionnellement au développement des usages d'IA qui devront donc intégrer *ab initio* cette donnée et minimiser, autant que possible leur impact environnemental qui peut s'appréhender sous différents aspects. Du point de vue de la consommation d'énergie, les modèles d'IA, en particulier les réseaux neuronaux profonds (*Deep Learning*), sont souvent très gourmands en énergie dû au très grand nombre de données qu'ils nécessitent.

L'entraînement de grands modèles sur des ordinateurs puissants nécessite d'importantes quantités d'électricité, ce qui peut contribuer aux émissions de gaz à effet de serre. Par ailleurs, la construction d'infrastructures pour soutenir la recherche et le développement en IA, y compris les centres de données et les serveurs de calcul, nécessite d'énormes quantités de ressources matérielles, notamment des métaux rares et de l'eau pour le refroidissement. De la même manière, les systèmes d'IA peuvent nécessiter la collecte et l'analyse de grandes quantités de données, ce qui peut avoir des implications environnementales si la collecte de données n'est pas gérée de manière durable.

C'est pourquoi, une attention particulière doit être portée, dès la conception des systèmes d'IA, à la minimisation de leur impact environnemental sur ces différents pans du développement de l'IA. C'est le cas notamment des IA dites « frugales » conçues pour limiter leur consommation d'énergie. Pour cela, le *transfert learning*, une méthode d'apprentissage par transfert de connaissance déjà acquises peut être une solution adaptée, car cela permettrait de bénéficier de l'apprentissage des algorithmes déjà existants par un simple transfert de connaissance. Au-delà de la réduction de consommation d'énergie ainsi obtenue, cette solution permet également un gain de temps considérable.

[22] L'info durable, 28 février 2022, « L'intelligence artificielle au service de l'écologie ». Disponible sur le lien suivant : https://www.linfodurable.fr/technomedias/lintelligence-artificielle-au-service-de-lecologie-31068

Toutefois, en contrepoids de son impact environnemental, il est pertinent de considérer les apports des solutions d'IA en tant qu'outils au service du développement durable. Par exemple, un projet de La Banque des territoires en partenariat avec Enedis et GRDF (gaz réseau distribution France) propose un service en ligne dédié à la rénovation thermique des bâtiments. Elle a pour objectif de faciliter les décisions des collectivités quant à la rénovation thermique de leurs bâtiments en analysant les données de consommation recueillies[23] grâce à des solutions d'IA et permettant d'élaborer les scenarii de réduction de consommation les plus efficaces. L'université de Grenoble-Alpes propose également un projet utilisant l'intelligence artificielle afin de simuler le recul à venir des glaciers des Alpes et d'anticiper les conséquences du réchauffement climatique[24].

Il apparaît donc, en matière de développement durable, que l'IA doit être perçue sous l'angle de son propre impact, qui doit être réduit autant que possible, mais aussi sous l'angle des nombreux apports, actuels et futurs, à cet enjeu mondial. S'il apparaît nécessaire de prendre des mesures telles que l'adoption de technologies plus économes en énergie, l'utilisation de l'IA pour optimiser la gestion des ressources, la sensibilisation aux implications environnementales de l'IA et la promotion de pratiques de développement durable dans le domaine de l'IA. Il est également important de développer des réglementations et des normes pour encourager une utilisation responsable de l'IA en tenant compte de son impact environnemental.

Au-delà des principes de primauté humaine, de performance et de respect de l'environnement ainsi que ceux énoncés par la charte éthique du Conseil de l'Europe, il est également nécessaire de mener des actions concrètes afin d'avoir un impact concret au sein de la société.

[23] L'info durable, [28] février [2022], « L'intelligence artificielle au service de l'écologie ». Disponible sur le lien suivant : https://www.linfodurable.fr/technomedias/lintelligence-artificielle-au-service-de-lecologie-[31068]

[24] L'info durable, [28] février [2022], « L'intelligence artificielle au service de l'écologie ». Disponible sur le lien suivant : https://www.linfodurable.fr/technomedias/lintelligence-artificielle-au-service-de-lecologie-[31068]

II. DES ACTIONS CONCRÈTES POUR PERMETTRE UNE BONNE APPRÉHSION DE L'IA DANS NOS SOCIÉTÉS

Les principes auxquels l'IA doit se conformer pour permettre de créer la confiance risquent toutefois de se révéler insuffisants pour permettre une appropriation globale de ces technologies par notre société. C'est à cette seule condition que les effets et les nouveaux usages de l'IA pourront être déployés au bénéfice de tous dans des domaines aussi variés que la santé, le développement durable, la compréhension et l'organisation des flux et réseaux ou encore la sécurité. L'appropriation de cet outil par la suite ne pourra s'opérer qu'au terme d'un travail de sensibilisation et d'explication auprès de l'ensemble des acteurs de l'IA, des pouvoirs publics et des utilisateurs afin que chacun perçoive clairement les possibilités, les limites, mais aussi les risques éventuels liés à ces usages. Ces différentes notions doivent impérativement faire l'objet d'un débat public, le plus large possible, mais également être éclairé et se fonder sur la réalité que recoupe ces technologies.

En parallèle, il est essentiel de poursuivre activement et d'encourager de manière volontariste la recherche et l'innovation dans ce domaine afin de permettre l'amélioration des connaissances et surtout de préserver des compétences précieuses dans notre pays pour développer ces solutions et maîtriser cette technologie essentielle à notre souveraineté, à notre avenir numérique ainsi qu'à la défense de nos valeurs fondamentales dans un espace numérique où nous devons rester en capacité de jouer un rôle actif.

Enfin, pour démonter toute l'utilité et l'innocuité des solutions d'IA de Confiance conçues et développées par les entreprises de la filière française de l'IA, il est également urgent de permettre à leurs concepteurs de mener des expérimentations dans des conditions réelles dans un cadre temporel et spatial défini. Une telle expérimentation, comme la prévoit la loi JOP 2024, permettra de tirer des enseignements et d'apporter les éventuelles améliorations nécessaires au regard des résultats obtenus.

a) La nécessité de la pédagogie

Comme expliqué ci-dessus, l'intelligence artificielle fait aujourd'hui partie du quotidien de ses utilisateurs et prend une place de plus en plus importante. Dès lors, la perception de l'IA est primordiale à évaluer afin de mesurer leur niveau de compréhension de cette technologie et niveau de maturité quant à son usage afin de réduire les risques que ces usages pourraient créer. En 2021, l'institut d'études *YouGov*[25] a mené une étude sur la perception de l'IA et de l'automatisation par la population et relève que 37% des interrogés en France se montrent « sceptiques » vis-à-vis de l'IA. A l'international, ce chiffre tombe à 27%. Toujours dans cette étude, les Français interrogés sont 35% à penser que l'IA et l'automatisation ont un impact négatif sur la société contre seulement 12% pour un impact positif.

Cet écart de perception entre la France et le reste du monde interroge et rend nécessaire de permettre à nos concitoyens d'avoir une meilleure compréhension du sujet, et de rendre accessibles toutes les informations nécessaires concernant ces technologies.

Toutefois, dès lors que l'on s'éloigne d'une appréciation générale sur l'IA, et que l'on s'intéresse à des secteurs plus spécifiques, la défiance relative constatée semble s'amenuiser. C'est le cas notamment dans le domaine de la santé où le baromètre de la *healthTech Pulselife*[26] relève que « si une grande majorité [des professionnels de la santé] pensent que la qualité de soins se détériore en France, ils sont encore plus nombreux à estimer que l'IA peut les aider » dans leurs tâches professionnelles quotidiennes. Le baromètre précise également qu'alors que « les connaissances scientifiques doublaient tous les 3 ans en 2010, elles sont désormais multipliées par 2 tous les 72 jours » complexifiant l'acquisition de ces nouvelles connaissances par l'Homme. Toujours au sein de cette étude, 81% des soignants et 86% des médecins pensent que l'IA peut les aider au quotidien.

Cet écart de perception peut être relié au caractère plus précis du champs considéré : ici, la santé. En effet, les effets positifs de l'IA apparaissent clairement l'emporter sur d'autres considérations si l'on considère un domaine particulier dans lequel cet apport est visible et compréhensible pour les praticiens.

[25] L'ADN, 6 décembre 2021, « Perception de l'intelligence artificielle et de l'automatisation ». Disponible sur le lien suivant : https://business.ladn.eu/news-business/actualites-annonceurs/etude-perception-intelligence-artificielle-automatisation-francais/

[26] ActuIA, 20 avril 2023, « Baromètre PulseLife : des soignants débordés mais confiants dans l'IA ». Disponible sur le lien suivant : https://www.actuia.com/actualite/barometre-pulselife-des-soignants-debordes-mais-confiants-dans-lia/?mc_cid=71518f2383&mc_eid=8cbeef909b

C'est pourquoi, beaucoup d'auteurs choisissent d'évoquer le « s » IA « s », pour souligner que cette technologie doit être appréhendée au regard de chacun de ses usages spécifiques et que les apports/risques qui y sont liés doivent aussi être appréhendés à ce même niveau. En effet, les variations entre les différents usages de l'IA sont si importantes que toute appréhension de l'IA comme concept général devient difficile voire est susceptible de conduire à des raisonnements et des conclusions erronés.

Aussi, l'ACN propose que les actions de pédagogie, sensibilisation, explication du rôle et du fonctionnement de l'IA adopte cette logique de différenciation en opérant systématiquement une différenciation entre les cas d'usages proposés. Cette vision nuancée permettrait de mener un débat public apaisé, en séparant les usages que la société considère comme inacceptables de ceux qu'il est nécessaire d'encourager et de développer en toute confiance afin que notre société puisse tirer le meilleur parti de l'évolution technologique.

L'objectif premier de cet effort pédagogique sera donc de renforcer la confiance des utilisateurs dans cette technologie, notamment en ses garde-fous technologiques et réglementaires, élaborés et consentis collectivement. Parallèlement, il semble plus que jamais nécessaire, d'accélérer et d'amplifier les efforts de recherche et d'innovation dans le domaine.

b) L'importance de la recherche

La recherche est une des clés de voute de l'évolution des sociétés. Le sujet de l'Intelligence Artificielle est, à cet égard, particulièrement crucial. En effet, maîtriser ces technologies, et donc animer une recherche de pointe sur ces domaines est un enjeu majeur pour notre pays, pour plusieurs raisons.

En premier lieu, l'IA est aujourd'hui à la pointe de la plupart des avancées technologiques dans le monde entier. En investissant dans la recherche en IA, la France peut rester compétitive sur le plan technologique, ce qui est essentiel pour sa position dans l'économie mondiale. Ensuite, l'IA est un moteur majeur de l'innovation, et permet de développer de nouvelles applications et de résoudre des problèmes complexes dans divers domaines, tels que la santé, la finance, l'industrie, l'agriculture, ….

En soutenant la recherche en IA, la France peut stimuler l'innovation et créer de nouvelles opportunités économiques. Par ailleurs, les entreprises françaises peuvent bénéficier de l'IA pour améliorer leur efficacité, leur productivité et leur compétitivité sur le marché mondial. La recherche en IA contribue à développer des solutions technologiques qui peuvent être utilisées par les entreprises pour rester compétitives. Mais l'IA peut également être utilisée pour relever des défis sociétaux importants, tels que la lutte contre le changement climatique, l'amélioration des soins de santé, la gestion des ressources naturelles, etc. Enfin, la recherche de pointe en IA contribue aussi au rayonnement international de la nation dans le domaine de la science et de la technologie.

Au-delà, la recherche en IA permettra aussi à notre pays d'être en capacité de maintenir notre souveraineté numérique, mais aussi de pouvoir défendre nos valeurs fondamentales dans l'espace numérique. Il existe encore peu d'universités en France proposant des formations dédiées à l'IA ou à l'apprentissage du *Machine Learning, du Deep Learning*, ... ce qui conduit à une pénurie de développeurs spécialisés en la matière. Cependant de plus en plus d'initiatives tentent tout de même de répondre à ces problématiques.

Ainsi, la stratégie nationale pour l'IA lancée en 2018 par le gouvernement, a pour objectif de développer un modèle éthique de l'IA qui « s'intègre au mieux dans la société et que chacun puisse se l'approprier pour en faire le meilleur usage »[27]. Le premier axe de cette stratégie consiste à développer un écosystème de talents. L'Etat souhaite retenir les meilleurs profils autour de l'IA et attirer les meilleurs centres de recherches privés. Ainsi, le programme national de recherche, coordonné par l'Institut national de recherche en informatique et en automatique (INRIA), s'est fixé 3 objectifs : développer l'un des meilleurs écosystèmes de recherche en France, améliorer les liens entre l'industrie et le secteur de la recherche publique et doubler le nombre d'étudiants formés à l'IA.

La création d'un service d'IA au sein de la CNIL est également la manifestation de l'importance de la recherche dans ce domaine. Il a pour objectif de renforcer son expertise et sa compréhension des risques pour la protection de la vie privée induit par l'IA ainsi que de se préparer à l'entrée en application de la proposition d'*AI Act*. Ses principales missions sont de :

[27] Ministère de l'Economie, des Finances et de la souveraineté industrielle et numérique, « La stratégie nationale pour l'IA ». Disponible sur le lien suivant : https://www.entreprises.gouv.fr/fr/numerique/enjeux/la-strategie-nationale-pour-l-ia

- Faciliter la compréhension du fonctionnement des systèmes d'IA par la CNIL mais aussi pour les professionnels et les particuliers,
- Consolider l'expertise de la CNIL dans la connaissance et la prévention des risques pour la vie privée liées à la mise en œuvre de ces systèmes,
- Préparer l'entrée en application de l'*AI Act*,
- Développer des relations avec les acteurs de l'écosystème.

En effet, pour être à même de comprendre les besoins en matière d'IA, il faut également inclure les acteurs qui la produisent dans les discussions qui auront lieu dans cette instance et valoriser leur expertise.

c) Favoriser l'expérimentation – création de valeur pour l'homme

Les solutions et technologies d'IA sont encore relativement récentes. Quelles que soient les promesses de ces nouvelles technologies, leur expérimentation à grande échelle est aujourd'hui une priorité.

En effet, les expérimentations en intelligence artificielle sont cruciales pour évaluer, améliorer et valider les systèmes d'IA. Elles jouent un rôle central dans le développement de cette technologie, en contribuant à son efficacité, à son éthique et à son impact dans une pluralité de secteurs.

Les expérimentations doivent permettre de valider les concepts théoriques en IA pour déterminer, par exemple, si une approche ou un algorithme fonctionne tel qu'on le souhaite, en utilisant des données fiables. Elles ont aussi pour objectif d'itérer et d'améliorer les modèles d'IA. En mesurant les performances d'un modèle, on peut identifier les lacunes, les domaines où des améliorations sont nécessaires et ajuster les paramètres des modèles d'IA pour obtenir les meilleures performances. Cela peut inclure l'optimisation des hyperparamètres, tels que la vitesse d'apprentissage ou la profondeur d'un réseau neuronal.

Expérimenter, c'est aussi valider l'efficacité d'une solution : dans des domaines tels que la médecine ou la finance, les expérimentations sont cruciales pour valider l'efficacité des systèmes d'IA avant leur déploiement dans des situations à grande échelle. Cela peut soutenir la garantie que les orientations choisies par l'IA sont fiables, à identifier des problématiques éthiques et les biais et donc conduire à prendre des mesures pour correctives et garantir le respect de l'éthique de l'IA.

Enfin, l'adaptation aux données fiables représente un enjeu majeur pour le bon fonctionnement d'une IA : les données fiables peuvent être complexes et variées. Les expérimentations permettent de tester comment les modèles d'IA se comportent à grande échelle, en prenant en compte les imprévus et les variations.

L'expérimentation permet de vérifier la véracité d'une hypothèse. En matière d'IA, l'expérimentation se concrétise par le test en conditions réelles d'un algorithme entrainé par des données d'apprentissages avec des données fiables et de la documentation sur cette expérimentation qui permettent de démontrer sa valeur sociétale visée établie dès sa conception.

En pratique, l'expérimentation permet donc de vérifier la performance d'un algorithme à la sortie de son laboratoire dans des conditions réelles ou au moins proches du réel. Sur le moyen terme, elle permet également de vérifier que l'évolution de ses usages reste compatible avec son intention initiale. Les Jeux Olympiques et Paralympiques de 2024 (JOP24) offrent une immense opportunité en France aux concepteurs d'IA. En effet, la loi du 19 mai 2023 relative au JOP24 autorise l'expérimentation de la vidéosurveillance intelligente pour assurer la sécurité des manifestations sportives, récréatives ou culturelles particulièrement exposées à des risques jusqu'au 31 mars 2025. Les expérimentations possibles qui en découlent permettront à de nombreux systèmes d'IA et à leurs concepteurs de tester leurs performances et de s'adapter aux besoins et aux contraintes de la réalité relevés.

CONCLUSION

La démocratisation de l'intelligence artificielle

Les entreprises membres de l'Alliance pour la Confiance Numérique (ACN) revendiquent une vision française et européenne de la confiance numérique. Cette vision s'inscrit pleinement au service des valeurs fondamentales que sont notamment les droits de l'Homme, les libertés fondamentales, la démocratie, l'état de droit et la souveraineté nationale et européenne. En effet, loin de s'opposer, la sécurité et les droits et libertés fondamentales sont consubstantiels, et leur imbrication est la clef de voûte de notre modèle de société français et européen. Les entreprises du secteur de la confiance numérique considèrent que cette conception garantit l'expression et la promotion des valeurs fondamentales européennes et constitue un atout majeur dans la compétition avec des acteurs d'autres régions du monde.

La France dispose d'atouts très forts dans ce domaine technologique et bénéficie notamment d'un tissu industriel complet doté de compétences de pointe reconnues, composé de leaders mondiaux mais aussi de nombreuses PME et start up très innovantes, reposant sur une recherche publique et privée performante.

Ces technologies sont diverses et peuvent être utilisées en vue de différentes finalités (sanitaire, amélioration de la gestion des flux, sécuritaire, organisation des grands évènements, *safe and smart city*, ...), et dans des cas d'usage très variés. En outre, elles disposent de nombreuses configurations permettant tout à la fois d'atteindre la finalité souhaitée tout en s'inscrivant dans le cadre du respect des valeurs fondamentales.

Pour autant, ces technologies, du fait notamment de raccourcis conceptuels largement répandus et/ou d'exemples tirés de contextes et de régions différentes, suscitent des interrogations légitimes de la part d'une partie de l'opinion publique. C'est pourquoi, l'ACN a souhaité contribuer à l'élaboration de critères et principes permettant de distinguer, de manière précise, auditable et sans ambiguïté, les IA qui pourraient être qualifiées d'IA de confiance.

Ainsi, une IA de confiance est donc une IA qui est transparente, éthique, responsable, sécurisée, traçable, guidée par des recommandations éthiques, soumise à validation indépendante, utilisée de manière éthique et avec le consentement des utilisateurs. Elle doit être conçue pour servir au mieux les intérêts de la société et respecter les valeurs humaines.

Afin de parvenir à cet écosystème de l'Intelligence Artificielle de Confiance que l'ACN appelle de ses vœux, il est nécessaire d'agir sur son cadre juridique qui doit poser ses critères afin d'accompagner et de soutenir son évolution. Les textes actuels étant insuffisamment adaptés, l'incertitude juridique permanente est un frein majeur pour les entreprises d'IA de Confiance. Il est donc urgent de finaliser les travaux engagés au niveau européen pour établir un cadre juridique stable pour l'IA, susceptible de protéger les droits et libertés des citoyens et de permettre l'innovation et le développement dans ce domaine stratégique. Le cadre en construction, composé de nombreux textes touchant aux multiples aspects de l'IA (nature technique, responsabilité, ...) doit être élaboré de manière cohérente et concertée pour constituer par la suite un socle solide international. L'enjeu crucial pour le législateur, est de proposer des réglementations qui parviennent à éliminer les risques d'atteintes aux valeurs fondamentales de l'Europe, sans pour autant dénoncer ou interdire des technologies dans leur ensemble (*ou per se*), privant de fait notre pays d'outils indispensables à notre souveraineté numérique et à notre autonomie stratégique.

Concernant leurs aspects techniques, les systèmes d'IA doivent poursuivre des objectifs de transparence, d'acceptabilité et d'explicabilité afin de rendre cette technologie accessible et accompagner les utilisateurs dans leurs usages pour que les systèmes développés soient exploités à leur plein potentiel dans des conditions de confiance. Les systèmes d'IA doivent également répondre à des objectifs de fiabilité et de sécurité en étant développés puis opérés dans des environnement de confiance qui respectent les bonnes pratiques de cybersécurité afin de réduire les possibilités d'attaques (injection de données, évasion, ...).

Ces différents objectifs doivent s'appuyer sur des référentiels techniques partagés, reflétant à la fois le consensus et l'état de l'art du secteur, au regard desquels il sera possible de déterminer objectivement si une solution peut, techniquement, être considérée comme de confiance. Les technologies d'IA et leur apprentissage sont dépendants des bases de données disponibles permettant de les entrainer et de les évaluer. A minima, comme pour tout ce qui concerne l'IA, il est nécessaire de créer des standards et des processus de qualification de ces bases d'apprentissage afin d'en assurer l'éthique et s'assurer que des biais d'apprentissage discriminants (parité, couleur de peau, ...) ne sont pas introduits dans ces apprentissages.

Il est donc opportun de contrôler la création de bases d'apprentissage en produisant des standards et des processus de qualification de ces bases pour s'assurer qu'elles sont protégées. Il apparaît également nécessaire que l'Europe puisse se doter de son propre organisme de certification des algorithmes biométriques, d'un point de vue sécuritaire mais aussi éthique. Aujourd'hui, seul le NIST américain (*National Institute of Standards and Technology*) est reconnu au niveau international.

Enfin, en complément des aspects juridiques et techniques, une IA de Confiance doit également remplir des critères éthiques, notamment en respectant les principes de la CEPEJ, les principes de la primauté de l'Homme dans l'IA, la nécessité de performance et la préservation de l'environnement.

La combinaison de ces critères juridiques, techniques et éthiques doit permettre de renouer le lien de confiance entre ces technologies nouvelles et l'ensemble des citoyens. Pour accompagner ce mouvement, la sensibilisation, l'éducation, la recherche et les expérimentations joueront un rôle essentiel.

La tenue d'événements internationaux comme les Jeux Olympiques et Paralympiques 2024 sont une occasion d'expérimenter les solutions françaises d'intelligence artificielle et leurs apports pour des objectifs de performance et de sécurité et ce dans un cadre d'expérimentation délimité et maitrisé par le législateur. L'ACN y voit une opportunité majeure pour créer de la confiance en démontrant l'utilité, l'efficacité de ces outils mais aussi leur parfaite compatibilité avec les droits et libertés fondamentales des citoyens. L'IA de Confiance, loin de s'opposer à ces valeurs fondamentales doit être comprise, au contraire, comme l'outil de leur préservation et de leur promotion.

En effet, l'IA est incontestablement une révolution technologique importante dont seuls les premiers effets sont aujourd'hui visibles. Cette évolution est porteuse à la fois d'opportunités majeures dans tous les domaines mais aussi de risques. Comme le démontre quotidiennement l'actualité, nos démocraties se voient bousculées par l'avènement des IA génératives abordées en introduction de ce livre blanc, et plus largement par des technologies permettant la manipulation de l'opinion à grande échelle, la diffusion de « *fake-news* » ou encore la manipulation cognitive.

Ces dangers ne pourront être combattus que par une maîtrise technique des processus conduisant à la mise en place des IA et par la mise en place des critères permettant de définir le champ de la confiance pour ces technologies.

La démarche que l'ACN porte, à travers ce document, consiste à considérer que cette révolution doit être encadrée, maîtrisée et optimisée pour permettre à nos sociétés de profiter pleinement du potentiel immense de l'IA tout en se prémunissant des dangers et risques réels dont cette même technologie est porteuse. Les entreprises françaises de la confiance numérique, qui portent cette vision, sont autant d'outils, à disposition de notre pays pour lui permettre de conserver la maîtrise de notre avenir numérique et de garantir l'exercice de nos droits et libertés dans l'espace numérique mais aussi physique.

Pour y parvenir, la confiance est la clef de voûte : garantir notre avenir commun, conformément à nos valeurs, passe par la définition de ce qu'est une IA de confiance. Il s'agit là d'un enjeu majeur qui doit intéresser chaque individu à son niveau : il est temps de définir collectivement ce qui est de confiance ou non pour notre futur !

SOURCES

Extrait d'un échange avec Chat GPT en juin 2023. Disponible sur le lien suivant : https://chat.openai.com/

2 Extrait de l'étude Gender Shade. Disponible sur le lien suivant : http://gendershades.org/overview.html

3 CNIL, 25 mars 2022, « Intelligence artificielle, de quoi parle-t-on ? ». Disponible sur le lien suivant :
https://www.cnil.fr/fr/intelligence-artificielle/intelligence-artificielle-de-quoi-parle-t-on

4 Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques. Disponible sur le lien suivant :
https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AOJ.L_.2022.277.01.0001.01.FRA&toc=OJ%3AL%3A2022%3A277%3ATOC

5 Règlement (UE) 2018/1139 du Parlement européen et du Conseil du 4 juillet 2018 concernant des règles communes dans le domaine de l'aviation civile et instituant une Agence de l'Union européenne pour la sécurité aérienne. Disponible sur le lien suivant :
https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32018R1139

6 Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données). Disponible sur le lien suivant :
https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679

7 Règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne. Disponible sur le lien suivant :
https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32018R1807

8 Article 35 du Règlement Général sur la Protection des Données (RGPD). Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679

9 Lignes directives concernant l'analyse d'impact relative à la protection des données 5AIPD) et la manière de déterminer si le traitement est « susceptible d'engendre un risque élevé » aux fins du règlements (UE) 2016, 679, 4 octobre 2017. Disponible sur le lien suivant : https://www.cnil.fr/sites/cnil/files/atoms/files/wp248_rev.01_fr.pdf

10 Loi n°2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions. Disponible sur le lien suivant : https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000047561989

11 ACN, 14 mars 2023, Position sur la consultation publique portant sur le projet de position de la CNIL relative aux conditions de déploiements des caméras dites « intelligentes » ou « augmentées » dans les espaces publics. Disponible sur le lien suivant : https://www.confiance-numerique.fr/position-acn-sur-la-consultation-publique-portant-sur-le-projet-de-position-de-la-cnil-relative-aux-conditions-de-deploiement-des-cameras-dites-intelligentes-ou-augmentees

12 CNIL, 23 janvier 2023, « Création d'un service de l'intelligence artificielle à la CNIL et lancement de travaux sur les bases de données d'apprentissages ». Disponible sur le lien suivant : https://www.cnil.fr/fr/creation-dun-service-de-lintelligence-artificielle-la-cnil-et-lancement-des-travaux-sur-les-bases-de

13 Etude du Conseil d'Etat, 31 août 2022, Intelligence artificielle et action publique : construire la confiance, servir la performance. Disponible sur le lien suivant : https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance

14 CEPD, 21 juin 2021, Communiqué de presse. Disponible sur le lien suivant : https://edpb.europa.eu/news/news/2021/edpb-edps-call-ban-use-ai-automated-recognition-human-features-publicly-accessible_fr

15 Proposition de directive du Parlement européen et du Conseil, 22 septembre 2022, article 3. Disponible sur le lien suivant : https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52022PC0496

16 Proposition de loi n°1630 visant à encadrer l'intelligence artificielle par le droit d'auteur, 12 septembre 2023. Disponible sur le lien suivant : https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi

17 CapAI, 25 mars 2022, "A procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act". Disponible sur le lien suivant : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091

18 Broutonlab, « Comment les données synthétiques peuvent-elles résoudre le problème du bais de l'IA ». Disponible sur le lien suivant : https://broutonlab.com/blog/ai-bias-solved-with-synthetic-data-generation

19 ANSSI, Guide d'hygiène informatique, 23 janvier 2027. Disponible sur le lien suivant : https://cyber.gouv.fr/publications/guide-dhygiene-informatique

20 CNIL, Quel est le périmètre des fiches pratiques sur l'IA ?, 11 octobre 2023. Disponible sur le lien suivant : https://www.cnil.fr/fr/quel-est-le-perimetre-des-fiches-pratiques-sur-lia

21 CEPEJ, 3-4 décembre 2018, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement. Disponible sur le lien suivant : https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b

22 CEPEJ, 3-4 décembre 2018, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, p12. Disponible sur le lien suivant : https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b

23 L'info durable, 28 février 2022, « L'intelligence artificielle au service de l'écologie ». Disponible sur le lien suivant : https://www.linfodurable.fr/technomedias/lintelligence-artificielle-au-service-de-lecologie-31068

24 L'info durable, 28 février 2022, « L'intelligence artificielle au service de l'écologie ». Disponible sur le lien suivant : https://www.linfodurable.fr/technomedias/lintelligence-artificielle-au-service-de-lecologie-31068

25 L'info durable, 28 février 2022, « L'intelligence artificielle au service de l'écologie ». Disponible sur le lien suivant : https://www.linfodurable.fr/technomedias/lintelligence-artificielle-au-service-de-lecologie-31068

26 L'ADN, 6 décembre 2021, « Perception de l'intelligence artificielle et de l'automatisation ». Disponible sur le lien suivant : https://business.ladn.eu/news-business/actualites-annonceurs/etude-perception-intelligence-artificielle-automatisation-francais/

27 ActuIA, 20 avril 2023, « Baromètre PulseLife : des soignants débordés mais confiants dans l'IA ». Disponible sur le lien suivant : https://www.actuia.com/actualite/barometre-pulselife-des-soignants-debordes-mais-confiants-dans-lia/?mc_cid=71518f2383&mc_eid=8cbeef909b

28 Ministère de l'Economie, des Finances et de la souveraineté industrielle et numérique, « La stratégie nationale pour l'IA ». Disponible sur le lien suivant :
https://www.entreprises.gouv.fr/fr/numerique/enjeux/la-strategie-nationale-pour-l-ia

Crédit photos : Pixabay

A PROPOS DE L'ACN

L'Alliance pour la Confiance Numérique (ACN) représente les entreprises (leaders mondiaux, ETI, PME/TPE et start up) du secteur de la Confiance Numérique et notamment celles de l'identité numérique, de la cybersécurité et de l'IA de Confiance. La France dispose dans ce domaine d'un tissu industriel très performant et d'une excellente internationalement reconnue grâce aux différents acteurs dynamiques du secteur.

On dénombre 2 130 entreprises réalisant en France 17,7 milliards d'euros de chiffre d'affaires dans ce secteur en forte croissance (7,6% de croissance annuelle moyenne depuis 2016). Les 106 entreprises de l'ACN, dont 89% de PME/TPE ou ETI, représentent 2/3 du chiffre d'affaires des entreprises françaises de la Confiance Numérique dans le monde (fabricants de matériel, éditeurs de logiciels, intégrateurs, services, laboratoires d'évaluation de sécurité, recherche, ...).

L'ACN est membre de la FIEEC (Fédération des Industries Electriques, Electroniques et de Communication), est membre associé du Campus cyber et participe activement aux travaux du CSF (Comité Stratégique de Filière) des Industries de Sécurité.

Par ailleurs, l'ACN est également membre fondateur de l'association représentant l'écosystème européen de la cybersécurité : ECSO (*European Cybersecurity Organisation*).

ECS
EUROPEAN CYBER SECURITY ORGANISATION

FIEEC
FÉDÉRATION DES INDUSTRIES ÉLECTRIQUES, ÉLECTRONIQUES ET DE COMMUNICATION

ACN
Alliance pour la confiance numérique

SITE :

https://www.confiance-numerique.fr/

@ACN_SecNum

ACN - Alliance pour la Confiance Numérique

ACN

Alliance pour la confiance numérique

Design Graphique : Iris Dupuy