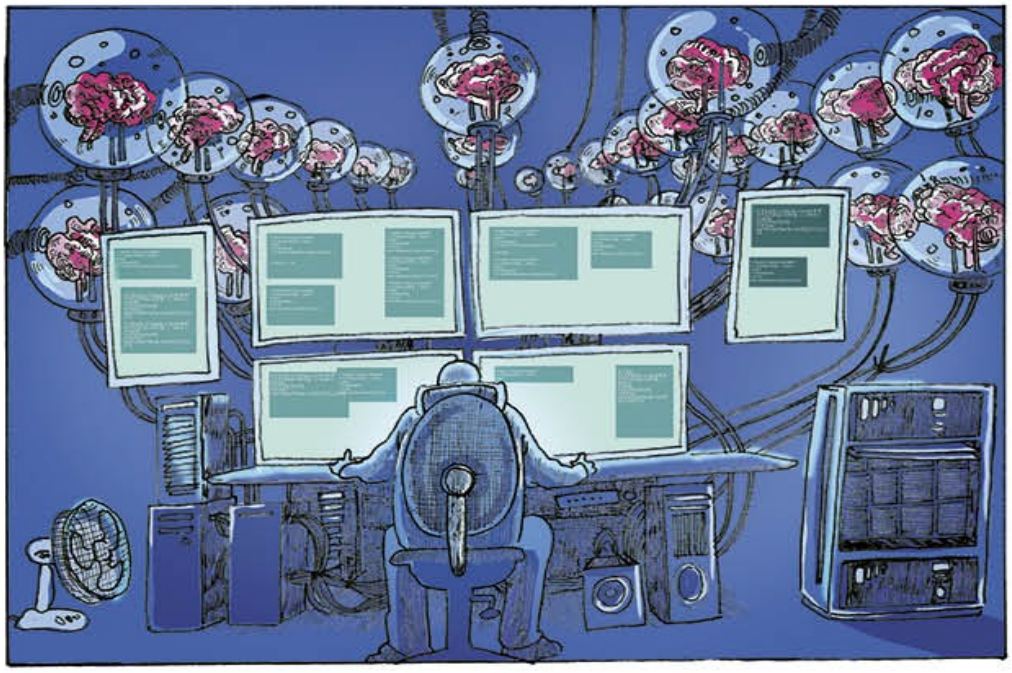
100
QUESTIONS/RÉPONSES



L'INTELLIGENCE ARTIFICIELLE

ET SES APPLICATIONS

Lamia El Badawi

ellipses

Collection « **100** QUESTIONS/RÉPONSES »

L'INTELLIGENCE ARTIFICIELLE ET SES APPLICATIONS

Lamia El Badawi

ellipses

Dans la même collection

Retrouvez tous les livres de la collection et des extraits
sur www.editions-ellipses.fr

ISBN 9782340-100787
Dépôt légal : février 2025

8/10 rue la Quintinie 75015 Paris

www.editions-ellipses.fr

Table des matières

V. L'INTELLIGENCE ARTIFICIELLE DANS L'ENTREPRISE

VI. L'IMPACT DE L'INTELLIGENCE ARTIFICIELLE SUR LE TRAVAIL

VII. L'INTELLIGENCE ARTIFICIELLE ET LA SÉCURITÉ

VIII. L'INTELLIGENCE ARTIFICIELLE ET LES TRANSPORTS

IX. LES ŒUVRES ET LES INVENTIONS DE L'INTELLIGENCE ARTIFICIELLE

Liste des principales abréviations

AMIAD	Agence ministérielle pour l'IA de défense
Art.	Article
BATX	Acronyme de *Baidu*, *Alibaba*, *Tencent*, *Xiaomi*
CAHAI	Comité *ad hoc* sur l'intelligence artificielle du Conseil de l'Europe
Caltech	*California institute of technology*
CCNE	Comité consultatif national d'éthique
CEPD	Comité européen de la protection des données
CEPEJ	Commission européenne pour l'efficacité de la justice
CESE	Comité économique et social européen
CNCDH	Commission nationale consultative des droits de l'homme
CNIL	Commission nationale de l'informatique et des libertés
CNPEN	Comité national pilote d'éthique du numérique
COMPAS	*Correctional Offender Management Profiling for Alternative Sanctions*
CSE	Comité social et économique
DREETS	Direction régionale de l'économie, de l'emploi, du travail et des solidarités
DSA	*Digital Services Act* (règlement sur les services numériques)
FMI	Fonds monétaire international
GAFAM	Acronyme de *Google*, *Amazon*, *Facebook* (désormais *Meta*), *Apple*, *Microsoft*
GPT	*Generative Pre-trained Transformer*
IA	Intelligence artificielle
INRIA	Institut national de recherche en informatique et en automatique ou Institut national de recherche en sciences et technologies du numérique
INRS	Institut national de recherche et de sécurité
IoT	*Internet of Things*
LLM	*Large Langage Models*
MIT	*Massachusetts Institute of Technology*
NBIC	Nanotechnologies, biotechnologies, technologies de l'information et sciences cognitives
NLG	*Naturel Language Generation*
OCDE	Organisation de coopération et de développement économiques
OEB	Office européen des brevets
OIT	Organisation internationale du travail
OIV	Opérateur d'importance vitale
OMS	Organisation mondiale de la santé
RGPD	Règlement général sur la protection des données
RLL	Règlement en ligne des litiges (plateforme)
SAE	*Society of Automotive Engineers*

SALA	Systèmes d’armes létales autonomes
UE	Union européenne
USPTO	*United States Patent and Trademark Office*

Introduction

Présenter l'intelligence artificielle est un exercice plus complexe qu'il n'y paraît, car il est très difficile de définir ce qu'est en soi l'intelligence, et il n'est pas certain que l'on puisse un jour y parvenir. L'ambition de reproduire toute l'intelligence humaine semble alors être une tâche difficilement réalisable, ne serait-ce que parce que nous ne maîtrisons pas suffisamment le fonctionnement du cerveau humain.

De ce fait, fournir une définition précise de ce qu'est véritablement l'intelligence artificielle devient un exercice incertain. Cette dernière n'est pourtant plus uniquement un thème des romans de science-fiction. Elle ne correspond certes pas vraiment à l'image que ceux-ci véhiculent, mais elle devient une réalité qui envahit tous les secteurs d'activité : la sécurité, la justice, les transports, la santé, l'industrie, la création artistique, la finance ou encore le commerce. Elle nous donne l'impression de vivre une nouvelle révolution.

Omniprésente, tous les secteurs de la société semblent touchés par les bienfaits de cette technologie, et les espoirs qu'elle suscite sont immenses. Et cette révolution n'est pas uniquement industrielle et technologique, elle est également sociétale avec le développement du numérique lié à la montée en puissance des ordinateurs et de leur vitesse de calcul. En effet, ces technologies modifient en profondeur la manière de travailler, de se divertir, les interactions sociales, le rapport à la santé, aux transports ou encore à l'information.

L'intelligence artificielle est déjà utilisée dans un très grand nombre d'applications. Il sera bientôt difficile de nommer un produit qui n'intègre pas un système d'intelligence artificielle. En analysant des quantités de données colossales à des vitesses inaccessibles à l'être humain, ces outils d'intelligence artificielle deviennent de plus en plus indispensables et influencent les décisions prises par les humains.

Mais l'intelligence artificielle suscite aussi des interrogations et des craintes, parfois injustifiées. Encore aujourd'hui, beaucoup l'imaginent comme une machine surdouée pouvant dépasser voire détruire l'homme. Cette peur est nourrie par la manière dont ce sujet est traité dans la culture populaire. Les frontières sont parfois floues entre mythe et réalité, entre science-fiction et futurologie. Ces prédictions alarmistes et infondées, concernant des menaces potentielles extrêmes liées au développement de l'intelligence, risquent pourtant d'occulter les dangers immédiats et bien plus probables.

Au-delà de cette vision dystopique, l'IA pose réellement de multiples questions éthiques et juridiques. Les réflexions se concentrent surtout sur la conception des algorithmes qui peuvent avoir des conséquences importantes sur les droits et les libertés fondamentaux des personnes.

L'intelligence artificielle est en effet régulièrement critiquée en raison de l'opacité des algorithmes et de leur effet « boîte noire ». Cette opacité ne concerne pas seulement les résultats mais aussi leur logique de fonctionnement. Du fait de l'utilisation croissante des systèmes de prise de décision automatique, il devient pourtant indispensable de pouvoir expliquer, interpréter et confirmer les résultats obtenus par les algorithmes. S'il est

impossible de comprendre une décision, comment faire confiance à des systèmes pour la conduite des véhicules ou le diagnostic d'une maladie grave ? Ce faisant, la question de la transparence de l'IA, et le principe d'explicabilité du fonctionnement des algorithmes et des décisions automatisées font partie des enjeux éthiques majeurs de l'intelligence artificielle.

Et comme l'intelligence artificielle s'invite dans de nombreux domaines, cet ouvrage est un voyage pour découvrir ses vastes applications, mais aussi ses multiples enjeux.

I.
QU'EST-CE QUE L'INTELLIGENCE ARTIFICIELLE ?

1 Qu'est-ce que l'intelligence ?

L'intelligence caractérise toutes les espèces vivantes, elle est inhérente à la vie. Qu'elle soit humaine ou animale, elle est naturelle. La notion d'intelligence est présente dans toutes les cultures mais il est difficile d'en dégager une définition unanimement admise. Sur le plan étymologique, le terme « intelligence » provient du latin *intellegentia*, dérivé de *intellegere* qui signifie comprendre. Le préfixe *inter* et le radical *legere* ou *ligare* désignent la capacité de faire des liens entre des éléments séparés. L'intelligence s'appréhende donc davantage par ce qu'elle permet de réaliser que par ce qu'elle est. Ce que l'on appelle intelligence recouvre les facultés dont les êtres vivants usent pour réaliser leurs objectifs. Nombre de disciplines ont tenté de définir cette notion sans pour autant parvenir à un consensus. Les définitions de l'intelligence sont le reflet des sociétés. Notre époque accorde sa confiance aux chiffres ; elle s'est donc entichée du QI. Le concept d'intelligence et sa mesure, exprimée par le QI, sont fréquemment utilisés comme des synonymes. En réalité, ce quotient ne mesure pas l'intelligence, il se contente de comparer les performances cognitives d'un individu par rapport à ses pairs du même âge, avec les marges d'erreurs que cela implique. Une clarification conceptuelle est, dès lors, un préalable indispensable à toute discussion sur l'intelligence.

Existe-t-il une forme unique d'intelligence ou plusieurs formes d'intelligence indépendantes chacune d'entre elles, spécifique à un domaine de compétence ? Telle est la question à laquelle de nombreux spécialistes ont tenté de répondre pendant plusieurs décennies.

Au début du XX^{e} siècle, Charles Spearman, psychologue britannique, publie un article qui fera date : « *L'intelligence générale objectivement déterminée et mesurée* ». Spearman affirme qu'un facteur général, qu'il nomme le facteur g, détermine un même niveau de performance pour l'ensemble des capacités intellectuelles, ce qui signifie que l'intelligence serait une seule et même entité. Trente ans plus tard, le psychologue américain Louis Thurstone adopte une approche différente et isole sept aptitudes, dont trois contenus (verbal, numérique et spatial), trois fonctions (mémoire, induction et déduction) et une septième aptitude : la fluidité verbale. Thurstone considérait ces aptitudes primaires comme indépendantes et définissait l'intelligence comme un ensemble hétérogène.

Après des décennies de débats, un large consensus s'est établi autour du modèle hiérarchique de l'Américain John Bissell Carroll qui, en 1993, a synthétisé les travaux existants. Carroll conçoit l'intelligence comme une pyramide à trois niveaux : à la base, on trouve des compétences spécifiques, telles que les capacités de raisonnement, la mémoire visuelle, la fluidité des idées, la facilité à manier les chiffres et le vocabulaire. Au deuxième niveau, ces capacités se regroupent en macrocompétences dont la mémoire, la vitesse de traitement de l'information, etc. Au sommet de la pyramide, se trouve un facteur d'intelligence générale, soit le facteur g.

Le psychologue américain Howard Gardner propose, quant à lui, une théorie multifactorielle. Il identifie sept, puis neuf formes d'intelligence : linguistique, logico-mathématique, spatiale, kinesthésique, musicale, interpersonnelle (faculté à bien

comprendre les autres) et intrapersonnelle (faculté à bien se comprendre soi-même), naturaliste (capacité à reconnaître les animaux, les plantes) et existentielle. Cette théorie connaît un certain succès auprès du grand public, notamment dans les milieux éducatifs. Les psychologues et les neuroscientifiques déplorent, quant à eux, son faible intérêt scientifique, et surtout l'absence d'une tentative de modélisation des relations entre les différentes formes d'intelligence.

D'autres chercheurs ont proposé d'élargir le concept d'intelligence. À partir des années 1990 ont ainsi été étudiées l'intelligence sociale, l'intelligence émotionnelle et l'intelligence pratique. Et des instruments d'évaluation de ces types d'intelligence ont été mis au point.

Il n'est donc pas aisé de définir l'intelligence. On admet aujourd'hui une conception assez large de l'intelligence comme étant l'ensemble des capacités cognitives permettant la compréhension des choses et des faits ainsi que leur analyse. Si définir l'intelligence est déjà difficile, comprendre son fonctionnement l'est encore davantage. Au-delà du débat sur les facteurs généraux ou spécifiques, le problème qui se pose est de savoir comment les différentes composantes de l'intelligence s'organisent et interviennent, de façon dynamique, pour réaliser une activité intellectuelle.

Le projet de l'intelligence artificielle repose sur l'idée qu'une fois observés, compris et décrits, les différents aspects de l'intelligence humaine peuvent être réalisés par des ordinateurs, ce qui est pour le moins discutable. Certaines intelligences artificielles sont en effet conçues de manière à simuler l'intelligence humaine mais il ne s'agit que d'une simulation. L'intelligence artificielle n'est donc qu'une « intelligence », fabriquée grâce à des outils informatiques, et elle est artificielle.

2 Qu’est-ce qu’une intelligence artificielle ?

La notion d’« intelligence artificielle » (IA) est entrée dans le langage courant, mais il n’en existe pas véritablement de définition unanimement partagée. Définir l’intelligence artificielle n’est donc pas chose aisée, car chacun en a une vision différente. La science-fiction et les productions cinématographiques ont notamment contribué à façonner les rapports qu’entretient le grand public avec l’IA et à donner l’illusion qu’elle pourrait accomplir bien davantage que ce qu’elle pourrait raisonnablement réaliser.

L’intelligence artificielle n’est pas une technologie au sens propre, mais plutôt une discipline scientifique pluridisciplinaire, qui réunit un ensemble de théories et de techniques, ayant un objectif ambitieux : comprendre comment fonctionne la cognition humaine et la reproduire en créant des systèmes cognitifs comparables à ceux de l’être humain. Autrement dit, il s’agit de doter les systèmes informatiques de capacités intellectuelles comparables à celles des êtres humains telles que le raisonnement, la planification et la créativité. L’IA serait donc un ensemble de méthodes, d’algorithmes et d’idées permettant de développer des outils qui peuvent accomplir certaines tâches réalisées par les humains : reconnaître les visages et les objets, conduire une voiture, traduire un texte dans une langue étrangère, etc. Tout système mettant en œuvre des mécanismes proches de celui d’un raisonnement humain pourrait ainsi être qualifié d’intelligence artificielle. C’est l’association de plusieurs disciplines qui a rendu possible le projet de créer une intelligence artificielle, celle-ci se situe entre autres à l’intersection de très nombreux domaines dont notamment l’informatique, les mathématiques appliquées, les sciences cognitives et les sciences du langage.

Il faut dire que la notion d’« intelligence artificielle » est assez anthropomorphique tant elle porte à croire que ces technologies seraient capables de reproduire le fonctionnement du cerveau humain. Selon Luc Julia, co-créateur de l’assistant vocal Siri d’Apple, il existe un malentendu autour du nom donné à cette discipline, malgré les dernières avancées technologiques en matière d’IA, car l’intelligence est réservée au vivant, et que l’on ne peut pas parler d’intelligence pour ces systèmes n’ayant pas la capacité d’innover. L’innovation serait en effet l’un des marqueurs de l’intelligence parce qu’elle permet d’aller au-delà de ce que l’on connaît ou du possible. La machine n’a pas cette capacité, elle reste cantonnée à son domaine de spécialisation. La confusion entourant aujourd’hui l’intelligence artificielle serait liée à l’association malheureuse des mots « intelligence » et « artificielle ».

L’adjectif « artificiel » associé au terme intelligence illustre d’ailleurs cet aspect, puisqu’une machine est programmée pour faire semblant de se comporter comme un être humain. En réalité, cette simulation nécessite toujours en amont une intervention humaine ; le programme informatique ne faisant qu’accomplir des tâches pour lesquelles il a été programmé. La programmation informatique peut toutefois reproduire des mécanismes cognitifs humains, tels que la logique déductive.

Il est vrai que ladite IA permet d’amplifier l’intelligence humaine, en analysant des données en quantité et à une vitesse inaccessibles à l’homme afin d’en tirer des conclusions réutilisables. Elle peut donc être considérée comme un outil puissant dans un

domaine particulier augmentant les capacités humaines, sans pouvoir s'y substituer. Logiquement, un outil sert à mieux réaliser la tâche pour laquelle il a été conçu, sans quoi il n'a aucun intérêt. L'IA peut même être supérieure aux meilleurs des humains, que ce soit pour jouer au jeu de go ou pour identifier une tumeur sur une radiographie. Elle est en revanche incapable d'utiliser toutes ses compétences en même temps afin d'analyser une situation ou former un raisonnement. Il n'y a donc pas lieu de développer un tel discours anxiogène vis-à-vis de l'IA, mais il convient de s'intéresser aux nouveaux risques, individuels et collectifs, qu'elles engendrent et d'essayer de poser un cadre clair pour son développement afin d'éviter les éventuelles dérives.

3 L’intelligence peut-elle être artificielle ?

Il existe plusieurs types d’intelligence artificielle, mais il y a une différence fondamentale entre la branche de la programmation qui élabore des solutions pertinentes à des problèmes spécifiques et celle cherchant à modéliser et simuler les fonctions du cerveau humain. C’est ce que l’on appelle l’informatique neuromorphique. De façon littérale, « neuromorphique » signifie qui imite le cerveau.

Le terme « informatique neuromorphique » est né dans les années 1980 avec les travaux de Carver Mead, chercheur en informatique au *California institute of technology* (Caltech). Le domaine de l’informatique neuromorphique est transversal et regroupe la biologie, l’ingénierie électrique, l’informatique et les mathématiques afin de créer des réseaux de neurones artificiels inspirés du système nerveux et du cerveau humain. D’une certaine manière, l’informatique neuromorphique représente une passerelle entre le cerveau humain et l’ordinateur. Bien qu’inventée dans les années 1980, l’informatique neuromorphique n’en est encore qu’à ses balbutiements. Grâce aux avancées des neurosciences, l’idée est venue de concevoir des machines plus intelligentes fonctionnant comme le cerveau, avec des neurones et des synapses artificiels organisés en réseau.

Les neurosciences ont, à ce titre, permis d’estimer que le cerveau humain contient entre 86 et 100 milliards de neurones. Ces chiffres sont constamment révisés au fur et à mesure que de nouvelles recherches sont menées. Il est important de noter que le nombre de neurones n’est pas le seul facteur qui détermine l’intelligence. L’organisation et la complexité des réseaux de neurones jouent également un rôle crucial.

En comparaison, le « cerveau numérique » de l’entreprise américaine Intel, réputé être l’un des plus puissants au monde, ne comporte que 100 millions de neurones, soit le cerveau d’un petit mammifère. Son nouveau système de calcul neuromorphique Hala Point comprend, quant à lui, 1,15 milliard de neurones. Il sera dédié à la recherche. De fait, même l’ordinateur le plus puissant au monde ne peut surpasser, à l’heure actuelle, le cerveau humain dans tous les domaines.

Selon Jean-Louis Dessalles, chercheur en intelligence artificielle à Télécom Paris, les réseaux de neurones, notamment ceux de l’apprentissage profond, ne sont rien de plus qu’une machine à associer. Ce pouvoir associatif permet de diagnostiquer les mélanomes, de maintenir un véhicule sur la route, etc. Ces réseaux acquièrent ainsi des formes d’expertise qui demandent aux humains des années d’études. Mais est-ce pour autant cela être intelligent ? Ce n’est évidemment qu’un aspect de l’intelligence qui correspondrait à l’apprentissage par cœur chez les humains. Il semble difficile de qualifier d’intelligent un individu apprenant par cœur une série de mots ou d’images. On conçoit que des activités faisant appel à la répétition, telles que la reconnaissance faciale, puissent être reproduites par des ordinateurs mais l’art, la spiritualité et, plus généralement, le génie humain leur resteront inaccessibles. Par ailleurs, les méthodes d’apprentissage actuelles ne permettent pas d’accréditer l’idée du cerveau humain fonctionnant comme une machine biologique, postulat issu de la cybernétique. En effet,

nul besoin pour un humain de voir des milliers de photos de chats pour savoir en reconnaître un ; un enfant est capable de distinguer les caractéristiques d'un animal à partir de peu d'exemples.

Certains espèrent pourtant que l'informatique neuromorphique nous rapprochera de ce qui serait une intelligence artificielle générale, capable de reproduire les capacités du cerveau humain, mais personne n'est parvenu à la créer jusqu'à présent. Une intelligence artificielle concurrençant l'intelligence humaine n'existe donc pas. Mais ce rapprochement terminologique entretient une certaine ambiguïté qui favorise parfois la croyance au détriment du savoir. Quelques aspects de l'intelligence humaine peuvent en effet être imités par une machine, on parle alors de simulation et non de création d'une intelligence artificielle *ex nihilo*. L'accroissement de la vitesse de calcul ou la capacité de stockage ne créent pas automatiquement de l'intelligence. L'éducation, les relations sociales et culturelles, ainsi que les influences de l'environnement sont autant de facteurs qui modèlent le cerveau humain tout au long de sa vie. Il n'existe donc pas d'intelligences identiques. La seule intelligence des machines est en réalité celle que nous leur apportons. Au regard des technologies actuelles, il semble improbable que l'on assiste à l'émergence d'une intelligence artificielle comparable à une intelligence humaine.

4 Quels sont les différents types d'intelligence artificielle ?

Les systèmes d'intelligence artificielle deviennent de plus en plus puissants et envahissants. Au sens large, on classe sous ce vocable des systèmes qui sont du domaine de la pure science-fiction et des systèmes déjà opérationnels en capacité d'exécuter des tâches très complexes. Ces technologies sont en réalité classées en fonction de leur capacité à imiter les caractéristiques humaines, de la technologie qu'elles utilisent pour y parvenir et de leurs applications dans le monde réel.

En utilisant ces caractéristiques comme références, tous les systèmes d'intelligence artificielle peuvent être classés en trois catégories : l'intelligence artificielle étroite (*Artificial Narrow Intelligence* ou ANI), également connue sous le nom d'IA étroite ou d'IA faible, qui possède une gamme étroite de capacités ; l'intelligence artificielle générale (*Artificial General Intelligence* ou AGI), également appelée IA forte ou IA profonde, qui est à la hauteur des capacités humaines ; la superintelligence artificielle (*Artificial Super Intelligence* ou ASI), dont les capacités sont supérieures à celles de l'homme.

On appelle IA étroite ou faible (*narrow* ou *weak* en anglais) une intelligence artificielle concentrée sur une tâche précise. Elle est conçue pour accomplir des tâches uniques : par exemple, la reconnaissance faciale, les assistants virtuels, la conduite d'une voiture ou encore la recherche sur internet. Ces applications ne peuvent pas être comparées à l'intelligence humaine, car l'IA faible est destinée à une seule fonction telle que le filtrage des spams. Elle ne peut être dotée ni d'intelligence réelle ni de conscience de soi. En comprenant la parole et le texte en langage naturel, l'IA est programmée pour interagir avec les humains. L'assistant Siri d'Apple, par exemple, est un parfait exemple de ce qu'est l'IA faible. Il est formé pour comprendre certaines requêtes et répondre à un certain nombre de questions posées par l'être humain. Il est ainsi à même de comprendre et d'interpréter le langage, ou du moins une partie de celui-ci. Il s'agit bien d'une véritable prouesse technologique qui reste cependant limitée à un nombre de fonctions prédéfinies. Tous les systèmes actuellement existants sont considérés comme relevant de l'IA faible.

L'IA faible est définie par opposition à l'intelligence artificielle générale ou l'IA forte (*strong*). Une telle IA s'apparente à celles que l'on retrouve dans les films et romans de science-fiction. Ce type d'intelligence artificielle peut penser, comprendre et agir d'une manière indiscernable de celle d'un être humain dans une situation donnée. Pour y parvenir, les chercheurs doivent développer des modèles permettant la création de machines dotées d'esprit, de conscience et de sensibilité, soit un algorithme universel, capable d'apprendre est d'agir dans n'importe quel environnement. Une telle IA ne doit pas se limiter à reproduire ou à simuler, elle doit pouvoir comprendre réellement les humains et penser comme eux. Ces machines auraient ainsi la subtilité du cerveau humain couplée à la performance et la puissance de calcul de l'IA. Le passage d'une IA faible à une IA forte reste peu probable en raison du manque de connaissances approfondies sur le cerveau humain qui reste le modèle de création de l'intelligence générale. Les combinaisons d'IA faibles pourraient toutefois donner l'illusion d'une IA forte. Une amélioration de la capacité des machines à apprendre et à voir, grâce à l'exploitation des

données massives provenant des échanges homme-machine en langage naturel, peut en effet donner l'illusion d'une évolution vers une IA forte. Il est toutefois très difficile de concevoir une machine dotée de capacités cognitives complètes malgré les avancées techniques.

En dépit de ces obstacles, certains prédisent un grand tournant technologique pour l'humanité à travers tout le discours faisant la promotion de la « singularité technologique », relayé en particulier par le technologue Ray Kurzweil ou le philosophe Nick Bostrom. Suivant ce scénario, le futur n'appartiendrait plus à l'homme mais à une nouvelle espèce, soit purement machine, soit hybride d'homme et de machine. Certains auteurs prédisent avec assurance que ce tournant décisif aura lieu au milieu du XXI^{e} siècle.

La superintelligence artificielle ne se contenterait alors plus d'imiter ou de comprendre l'intelligence et le comportement humains, elle parviendrait à les surpasser. Surpasser l'humain sur un plan intellectuel sous-entend que l'IA serait capable de prendre en main son propre destin ainsi que celui de l'humanité.

Cette superintelligence est depuis longtemps un thème privilégié de la science-fiction dystopique. Il nourrit l'imaginaire du grand public avec des œuvres créant un rapport ambivalent aux machines allant du robot qui se révolte contre l'humanité et qui l'asservit jusqu'au robot empathique entretenant un lien émotionnel avec l'homme (*A.I.*, *Wall-E*, etc.).

Longtemps cantonnés à la science-fiction, les promesses et les risques de l'intelligence artificielle alimentent désormais les débats de société. En réalité, la confusion avec les représentations fictionnelles de l'IA provient du fait que le grand public n'établit pas de distinction entre l'intelligence générale et celle dite spécialisée. L'intelligence générale désigne une capacité d'apprentissage et d'adaptation quasi infinie et très rapide. Elle permet de prendre des décisions sans se priver d'une réflexion morale. L'intelligence humaine relève de cette catégorie. Quant à l'intelligence spécialisée, elle désigne la capacité de réaliser des tâches spécifiques ou des objectifs précis. L'intelligence artificielle actuelle relève de cette seconde catégorie. Les exploits de cette technologie sont nombreux et surprenants, mais il est peu probable qu'elle développe un jour une intelligence générale pour différentes raisons, notamment pour son défaut de capacité d'adaptation. Elle peut toutefois nous être supérieure dans des domaines spécialisés, tels que le jeu vidéo *StarCraft* et les jeux de go ou d'échecs.

5 Quelles sont les origines historiques de l'intelligence artificielle ?

L'idée de créer des êtres artificiels, capables d'avoir une conscience similaire à celle de l'homme, a toujours été présente dans l'imaginaire humain. Elle est présente dans le golem de la mythologie juive, la Galatée de Pygmalion, les contes pour enfants comme Pinocchio, ou encore le Frankenstein de Mary Shelley. Très tôt, les philosophes ont rêvé d'une automatisation de la pensée et du raisonnement. Le lien entre la pensée et les mathématiques sera surtout le fait des philosophes du XVII^{e} siècle, tels que Leibniz évoquant l'idée d'un *calculus ratiocinator*, ou d'un calcul pensant.

Si l'idée même d'intelligence artificielle est ancienne, la discipline est particulièrement jeune et n'est apparue qu'après la Seconde Guerre mondiale. Deux actes fondateurs ont posé les bases de cette nouvelle discipline.

L'année 1950 marque un tournant décisif dans l'histoire de l'intelligence artificielle avec la publication de l'article « *Computing Machinery and Intelligence* » (les ordinateurs et l'intelligence) d'Alan Turing, le père de l'informatique moderne ayant permis de déchiffrer le langage crypté de la machine Enigma utilisée par la marine allemande. Son article pose les fondements de l'IA moderne et s'ouvre sur une question qui occupera les chercheurs de toutes disciplines pendant des décennies : « les machines peuvent-elles penser ? ». Pour répondre à cette question, Alan Turing propose un test soumettant l'ordinateur au jeu de l'imitation : il s'agit de questions libres posées par un interrogateur humain à un interlocuteur se trouvant dans une autre pièce. Si l'auteur des questions pense avoir affaire à une personne humaine, la machine réussit alors le test.

L'IA naît alors comme discipline scientifique au milieu des années 1950, dans la continuité de la cybernétique. En réalité, le véritable acte fondateur de l'intelligence artificielle se situe en 1956, lors de la conférence de Dartmouth, organisée par quatre chercheurs américains : John McCarthy, Marvin Minsky, Nathaniel Rochester et Claude Shannon.

Les ambitions de ce groupe de travail étaient d'étudier l'intelligence artificielle et de trouver le moyen de doter les ordinateurs d'une intelligence généraliste comparable à celle de l'homme, et non limitée à certains domaines ou à certaines tâches. Cette conférence marque formellement la naissance de la notion d'intelligence artificielle dans le milieu de la cybernétique.

Dès le début, deux courants de pensée vont s'affronter et rythmer les différentes recherches sur l'IA : le cognitivisme, ou l'approche symbolique, et le connexionnisme.

Qu'est-ce que le cognitivisme et le connexionnisme ?

Le courant dit cognitiviste utilise le raisonnement formel et la logique ; il s'agit d'une conception cartésienne de l'intelligence s'appuyant sur l'idée que nous raisonnons en appliquant des règles logiques (déduction, classification, hiérarchisation…). Les systèmes experts, dont le but est de reproduire le raisonnement et les connaissances d'un expert, sont la forme la plus connue et répandue de ce courant symbolique. Afin d'élaborer ces règles, il convient de reproduire le raisonnement d'un expert, par exemple d'un médecin. Le programmeur doit donc décortiquer et détailler le raisonnement de ce dernier pour en

induire des règles qu'une machine pourra d'appliquer afin de parvenir à un diagnostic. Cette approche s'est avérée très efficace pour résoudre des problèmes complexes dans le domaine des mathématiques et de la science, mais elle a été considérée comme inadaptée au traitement du langage et à la reconnaissance d'objets. En effet, ces systèmes se confrontèrent à une explosion du nombre de règles qui rendait leur programmation et leur maintenance difficiles et coûteuses. Les systèmes experts furent alors progressivement délaissés, ce qui amena à ce que l'on appelle « l'hiver de l'intelligence artificielle » dans les années 1980.

Ce courant, ayant dominé la recherche en IA depuis son origine au milieu des années 1950 jusqu'au début des années 1990, a progressivement laissé sa place aux modèles connexionnistes. Par opposition au cognitivisme, le connexionnisme part de la notion de réseau de neurones. Pour illustrer cette approche, on peut faire une analogie avec les neurones biologiques, qui font circuler de l'information à travers chaque connexion. Ces techniques d'apprentissage utilisant des réseaux de neurones ont longtemps pourtant été moquées et ostracisées par le courant dit cognitiviste. Ce n'est que grâce aux progrès de la puissance de calcul des microprocesseurs, à l'avènement de l'internet et au *big data* ainsi que l'exploration directe du cerveau humain que le courant connexionniste a fini par dominer les recherches et est venu proposer une alternative aux limites du courant symbolique. C'est de cette catégorie que relèvent les techniques d'apprentissage automatique *(machine learning)* ayant le vent en poupe, telles que l'apprentissage profond (*deep learning)*. Selon les connexionnistes, le raisonnement ne serait qu'une part infime de l'intelligence humaine. Ils considèrent que c'est en entraînant la machine à apprendre qu'elle sera en mesure d'agir de manière intelligente, car l'humain apprend par expérience, perception et intuition.

Cette percée spectaculaire du courant connexionniste n'invalide pas pour autant l'approche symbolique. En plus de mémoriser de grandes quantités de phénomènes bruts perçus du réel, notre cerveau a aussi la capacité de les compresser en permanence en théorie. Celles-ci confèrent des capacités de généralisation, de raisonnement et de compréhension qui ne peuvent être développées par la seule compilation de phénomènes perçus et accompagnés de probabilités. Par ailleurs, il est plus aisé de comprendre le processus de prise de décision d'un algorithme lorsque celui-ci s'appuie sur des règles, contrairement à l'approche connexionniste, souvent critiquée pour son opacité. Dès lors que la machine apprend, il est difficile d'expliquer et de retracer précisément tout le processus de décision. C'est la raison pour laquelle certains souhaitent combiner ces deux courants afin de créer une IA hybride. La recherche en intelligence artificielle tend à se diriger vers une combinaison de ces deux différentes approches.

6 Qu'est-ce qu'un algorithme ?

Le terme « algorithme » est souvent associé à l'informatique, pourtant il est issu du nom du mathématicien perse du IX^{e} siècle Al-Khwârizmî, considéré comme le père de l'algèbre. Le domaine qui étudie les algorithmes est appelé, quant à lui, l'algorithmique. La notion d'algorithme est donc très antérieure à la création du premier ordinateur mais ses contours ne sont pas toujours bien définis. Ceux-ci peuvent en effet varier selon le contexte de l'utilisation et du niveau de complexité de l'algorithme. Il existe différentes définitions de la notion d'algorithme. La Commission nationale de l'informatique et des libertés (CNIL) définit l'algorithme comme étant : « *la description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée* ». Il peut aussi être défini comme un « *ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations* » (Larousse).

De manière générale, l'algorithme est une procédure, c'est-à-dire une série d'étapes destinée à trouver une solution à un problème dans un délai raisonnable. Par exemple, on peut considérer comme un algorithme une recette de cuisine spécifiée par des entrées (ingrédients), des instructions à réaliser (frire, rissoler, etc.) et assortie d'une durée dont l'exécution des étapes aboutit à un résultat : un plat. Il en va de même pour une partition de musique, puisqu'il s'agit d'une suite d'étapes permettant d'obtenir un résultat : l'interprétation d'un morceau de musique. Il convient donc de dissocier l'aspect technologique de la notion d'algorithme. L'algorithme relève de l'inconscient mis en œuvre par tout un chacun confronté à la résolution d'un problème. Il apparaît indissociable de la raison humaine et concourt à la formation de la pensée.

Composé de processus et de formules mathématiques, il « *peut être traduit, grâce à un langage de programmation, en un programme exécutable par un ordinateur* » (Larousse). L'algorithme informatique est ainsi composé d'une suite d'instructions permettant d'indiquer à un ordinateur les différentes étapes à réaliser en vue d'exécuter une tâche. Les algorithmes sont en principe destinés à être mis en œuvre sous forme de programme, dans un langage de programmation compréhensible par un ordinateur. L'algorithme permet donc au travers d'un programme informatique de transmettre à l'ordinateur les instructions pour réaliser une tâche. Les ordinateurs ne comprenant pas le langage humain, l'algorithme doit alors être traduit en code écrit dans un langage de programmation. Les développeurs ont le choix entre de nombreux langages qui présentent chacun leur spécificité et des usages spécifiques (par exemple, le langage Python). Autrement dit, pour qu'un ordinateur accomplisse une tâche, il a besoin d'un programme informatique. Ce programme informatique indique à l'ordinateur étape par étape ce qu'il doit faire. Les programmes informatiques sont donc eux-mêmes composés d'algorithmes écrits dans des langages de programmation.

Un algorithme informatique fonctionne par entrée (input) et sortie (output). En d'autres termes, il reçoit une information et applique les instructions à cette entrée pour générer une sortie. Par exemple, un moteur de recherche est un algorithme recevant une requête de recherche en guise d'entrée. Il mène une recherche dans sa base de données pour des éléments correspondant aux mots de la requête et fournit ensuite les résultats (sortie).

Aujourd'hui, les algorithmes sont utilisés dans de nombreux domaines : proposer du contenu à des utilisateurs en fonction de leur comportement, accès à l'enseignement supérieur (Parcoursup), piloter de façon autonome des automobiles, *PageRank* du moteur de recherche Google qui classe les sites web en fonction de leur pertinence par rapport à la requête, *NewsFeed* de Facebook qui détermine les contenus visibles par l'utilisateur, etc.

Les algorithmes et l'intelligence artificielle sont souvent présentés comme étant indissociables. La puissance de calcul dont disposent les ordinateurs a aussi permis l'élaboration de nouveaux algorithmes. Ont notamment été conçus des algorithmes dits « évolutionnistes », à même d'évoluer de manière à résoudre au mieux un problème donné.

En réalité, ils forment la base de tous les systèmes d'intelligence artificielle. En effet, les algorithmes apprenants (*machine learning*) constituent une catégorie spécifique d'algorithmes. Plutôt que de recevoir des instructions spécifiques sur la tâche à effectuer, ils sont capables d'apprendre à partir de données. Il s'agit d'un procédé qui consiste à alimenter la machine avec des données choisies et répétées qui vont lui permettre d'apprendre et de déterminer seule les opérations à accomplir pour effectuer la tâche en question. L'intelligence artificielle reposant sur le machine learning concerne donc des algorithmes dont la particularité est d'être conçus de sorte que leur comportement évolue dans le temps, en fonction des données qui leur sont fournies. Ainsi deux ordinateurs identiques, exécutant une même tâche, pourront aboutir à un résultat distinct si leurs algorithmes sont différents.

Malgré les avantages et les possibilités qu'ils offrent, la fiabilité de ces algorithmes peut être mise en doute. Non seulement les algorithmes peuvent être mal conçus, mais des données de mauvaise qualité ou des choix méthodologiques discutables peuvent conduire à des résultats biaisés, voire discriminatoires.

L'opacité des données et des instructions mises en œuvre par certains procédés algorithmiques soulève également d'autres types d'inquiétudes. Les algorithmes d'intelligence artificielle ont en effet des niveaux d'opacité différents. Selon le type d'algorithme, certains sont plus facilement explicables, d'autres restent opaques. La performance des algorithmes est souvent privilégiée au détriment de leur transparence et de l'explicabilité de leurs résultats. L'effet « boîte noire » des algorithmes, notamment des algorithmes d'apprentissage automatique, est à ce titre souvent souligné.

Or, dans de nombreux domaines, il n'est pas souhaitable que des décisions, ayant un certain impact sur la gestion des affaires humaines, soient prises sans que la personne concernée puisse comprendre la méthode utilisée pour arriver à un tel résultat. Il est en effet important de pouvoir expliquer, interpréter et confirmer les résultats obtenus par les algorithmes. C'est la raison pour laquelle la perspective de la généralisation du recours aux algorithmes suscite des craintes.

7 Qu'est-ce que l'apprentissage automatique (*machine learning*) ?

Parmi les techniques utilisées en intelligence artificielle, certaines relèvent de ce que l'on appelle « l'apprentissage automatique » ou « *machine learning* » en anglais. L'expression *machine learning* a été utilisée, pour la première fois, en 1959, par l'informaticien américain Arthur Samuel, après la création de son programme jouant au jeu de dames et s'améliorant en jouant.

L'apprentissage automatique est un champ d'étude de l'intelligence artificielle qui vise à donner aux machines la capacité d'apprendre à partir de données. Il s'agit de déduire de l'observation de données du passé une règle qui pourrait s'appliquer à des données du présent. Ces données ne servent pas seulement à alimenter l'IA, elles sont aussi le facteur la rendant possible. Pour qu'un système apprenne, il est nécessaire de lui fournir des données. Contrairement à un enfant qui a besoin de voir quelques chats pour apprendre à reconnaître cet animal, une machine a besoin de voir des milliers, voire des millions de photos de chats pour construire son modèle d'apprentissage.

Cette méthode est en effet basée sur des algorithmes auto-apprenants qui se construisent et évoluent en fonction des données qui leur sont fournies. Cela peut être des chiffres, des images, des statistiques, des mots, etc. L'apprentissage automatique passe par une phase d'apprentissage modifiant itérativement l'algorithme qui apprend à partir de données, sans être explicitement programmé, et améliore ses performances au fur et à mesure de son apprentissage, mais aussi au fil des entraînements successifs. Il s'oppose ainsi à la programmation qui permet d'exécuter des instructions prédéterminées. L'intelligence artificielle reposant sur le *machine learning* concerne donc des algorithmes dont la particularité est d'être conçus de sorte que leur comportement évolue dans le temps, en fonction des données qui leur sont fournies.

Les algorithmes classiques sont dits déterministes, leurs critères de fonctionnement sont explicitement définis par ceux qui les mettent en œuvre. Les algorithmes apprenants, au contraire, sont dits probabilistes ; leurs résultats sont mouvants et dépendent de la base d'apprentissage qui leur a été fournie et qui évolue elle-même au fur et à mesure de leur utilisation.

C'est cette approche qui est mise en avant depuis quelques années et qui bénéficie d'une large couverture médiatique. L'essentiel de l'engouement actuel autour de l'intelligence artificielle est en réalité un engouement pour ces méthodes d'apprentissage automatique. Il faut donc garder à l'esprit que l'IA désigne désormais ces algorithmes d'apprentissage automatique, qui se sont considérablement améliorés ces dernières années, mais qui ne constituent qu'une partie de l'intelligence artificielle au sens large.

Différentes formes d'apprentissage sont possibles. On peut distinguer trois grandes catégories d'algorithmes d'apprentissage : l'apprentissage supervisé ; l'apprentissage non supervisé et l'apprentissage par renforcement.

L'apprentissage supervisé utilise des données étiquetées pour former (ou superviser) le modèle et prédire avec précision les résultats. Par exemple, prenons le cas d'une application destinée à reconnaître les spams de manière automatique. En lui présentant des courriels étiquetés comme « spams » ou « désirables », l'algorithme comprend alors

quelles sont les caractéristiques qui permettent de classer ces courriels dans chacune des catégories. Cette approche est similaire à l'apprentissage humain sous la supervision d'un enseignant. Ce dernier fournit à l'élève de bons exemples à mémoriser, et l'élève en déduit des règles générales. En utilisant ces données étiquetées, le modèle apprend alors de chaque exemple avec pour but d'être capable de généraliser son apprentissage à de nouveaux cas.

Les applications de l'apprentissage supervisé sont généralement divisées en deux catégories : la classification et la régression. La classification consiste à attribuer une classe (une catégorie) à des objets tandis que la régression peut servir à déterminer une valeur réelle, calculée, telle que le prix ou le poids.

Par exemple, si l'on souhaite prédire le prix d'une maison dans la région lyonnaise, l'utilisation d'un modèle de classification permet de classifier cette maison au sein d'une gamme de prix prédéterminés, tandis que l'utilisation d'un modèle de régression permettrait de prédire la valeur exacte de la maison. L'apprentissage supervisé produit des résultats précis et fournit des solutions plus ou moins fiables, mais il nécessite toutefois d'étiqueter les données, tâche complexe et coûteuse.

Aux côtés de l'apprentissage supervisé, il est également possible d'effectuer un apprentissage non supervisé. Dans ce cas, l'algorithme apprend à partir de données non étiquetées. L'algorithme fait alors lui-même les rapprochements en fonction des caractéristiques qu'il est en mesure de repérer sans intervention humaine (non supervisé). Il propose ainsi des réponses à partir d'analyses et de groupements de données. L'inconvénient est que l'on ne peut pas toujours expliquer de façon précise pourquoi l'algorithme donne tel ou tel résultat. Cette technique est surtout utilisée pour regrouper les données en fonction de leurs similitudes ou de leurs différences (*clustering*). Elle semble proche de celle de la classification dans l'apprentissage supervisé, mais, contrairement à cette dernière, les classes ne sont pas définies par un humain. Cette technique est très utilisée dans le domaine du marketing pour placer dans des groupes les différents clients. Par exemple, l'algorithme peut recommander un livre ou un film à un utilisateur en fonction des goûts d'utilisateurs partageant des caractéristiques communes.

L'approche, qui combine à la fois des techniques supervisées et non supervisées, est appelée apprentissage semi-supervisé. Dans ce cas, on soumet à l'algorithme quelques données annotées et beaucoup de données non annotées. La combinaison de ces deux ensembles de données permet d'améliorer sensiblement les résultats sans avoir recours à l'annotation manuelle, opération coûteuse et chronophage.

Enfin, l'apprentissage par renforcement est une sorte d'apprentissage automatique non supervisé, mais le mode d'apprentissage est différent. Il s'agit d'une méthode permettant à un algorithme d'apprendre par expérimentation et erreurs pour adapter sa stratégie après chaque étape de décision. On ne fournit donc pas ici de données d'entraînement à l'algorithme. Il acquiert ces données directement au contact de son environnement et il les mémorise. Ainsi, l'algorithme essaie plusieurs solutions, observe la réaction de son environnement et s'adapte pour trouver la meilleure stratégie. Ces modèles d'apprentissage s'inspirent du processus humain d'acquisition des connaissances par essais-erreurs. Une méthode de récompense des comportements souhaités et de punition

des comportements négatifs est élaborée. Des valeurs positives sont attribuées à ces comportements souhaités pour fournir un renforcement positif et des valeurs négatives aux comportements non souhaités pour un renforcement négatif. L'objectif de l'algorithme est alors de trouver une solution qui maximisera ses récompenses. L'apprentissage par renforcement profond a été utilisé pour faire apprendre à des programmes des stratégies de jeu, et notamment pour entraîner le célèbre programme AlphaGo Zero, la nouvelle version du programme de *DeepMind*, filiale de Google, qui avait battu en 2016 l'un des meilleurs joueurs au monde de jeu de go, Lee Sedol. Les versions précédentes combinaient cette méthode avec de l'apprentissage supervisé, alimenté par des parties de référence jouées par des humains.

8 Qu'est-ce que l'apprentissage profond (*deep learning*) ?

Encore connu sous l'expression « apprentissage profond » ou encore « rétropropagation de gradient », le *deep learning* est un procédé d'apprentissage automatique utilisant des réseaux de neurones artificiels s'inspirant du cerveau humain. Ces réseaux de neurones artificiels ont été imaginés dès les années 1940 et ont abouti au Perceptron, l'un des plus anciens algorithmes de *deep learning,* considéré comme le premier réseau de neurones artificiels évolutif, capable d'apprendre par expérience. En 1957, Frank Rosenblatt, psychologue américain qui fut l'un des pionniers de l'intelligence artificielle, inventa ce dernier afin de réaliser ce qui constitue la première modélisation d'un réseau de neurones artificiels. Des perceptrons multicouches ont ensuite été proposés en 1986. Le perceptron multicouche est à ce titre un ensemble de neurones organisés en couche. Le signal d'entrée se propage d'une couche à l'autre jusqu'à la sortie. À l'image du cerveau humain, le système est capable d'apprendre dès l'instant où les neurones artificiels se connectent entre eux.

Ces réseaux des neurones étaient pourtant mal considérés jusqu'au début du millénaire avant de revenir sur le devant de la scène dans les années 2000 avec les réseaux de neurones profonds. L'amélioration des algorithmes, l'augmentation de la puissance de calcul des processeurs et l'émergence du *big data*, qui permet d'entraîner les systèmes sur d'immenses quantités de données, ont finalement permis à l'apprentissage profond de s'imposer face aux autres méthodes. Les chercheurs Geoffrey Hinton, Yann LeCun et Yoshua Bengio ont d'ailleurs reçu le Prix Turing en 2018 pour leurs travaux sur l'apprentissage profond.

Lorsqu'on parle d'intelligence artificielle, on fait très souvent référence au *machine learning*, plus précisément au *deep learning*. En réalité, ce dernier est une branche du *machine learning*, qui à son tour est une sous-catégorie de l'intelligence artificielle. Ce qui rend le *deep learning* particulier, c'est son mode de fonctionnement. Les réseaux de neurones artificiels s'inspirent du système nerveux humain. Le processus d'apprentissage est qualifié de profond parce que la structure des réseaux neuronaux artificiels se compose de plusieurs couches d'entrée, de sortie et masquées. Grâce à cette structure, la machine est capable d'apprendre au travers de son propre traitement de données. Ce réseau est composé de dizaines, voire de centaines de couches de neurones, chacune recevant et interprétant les informations de la couche précédente.

Le système apprendra, par exemple, à reconnaître les lettres avant d'identifier les mots dans un texte, ou déterminera s'il y a un visage sur une photo avant de découvrir de quelle personne il s'agit. Les réseaux de neurones artificiels sont également utilisés dans le traitement automatique du langage ou *natural language processing* (NLP) qui regroupe l'ensemble des techniques permettant la compréhension et la restitution par les machines de langages humains, écrits ou oraux.

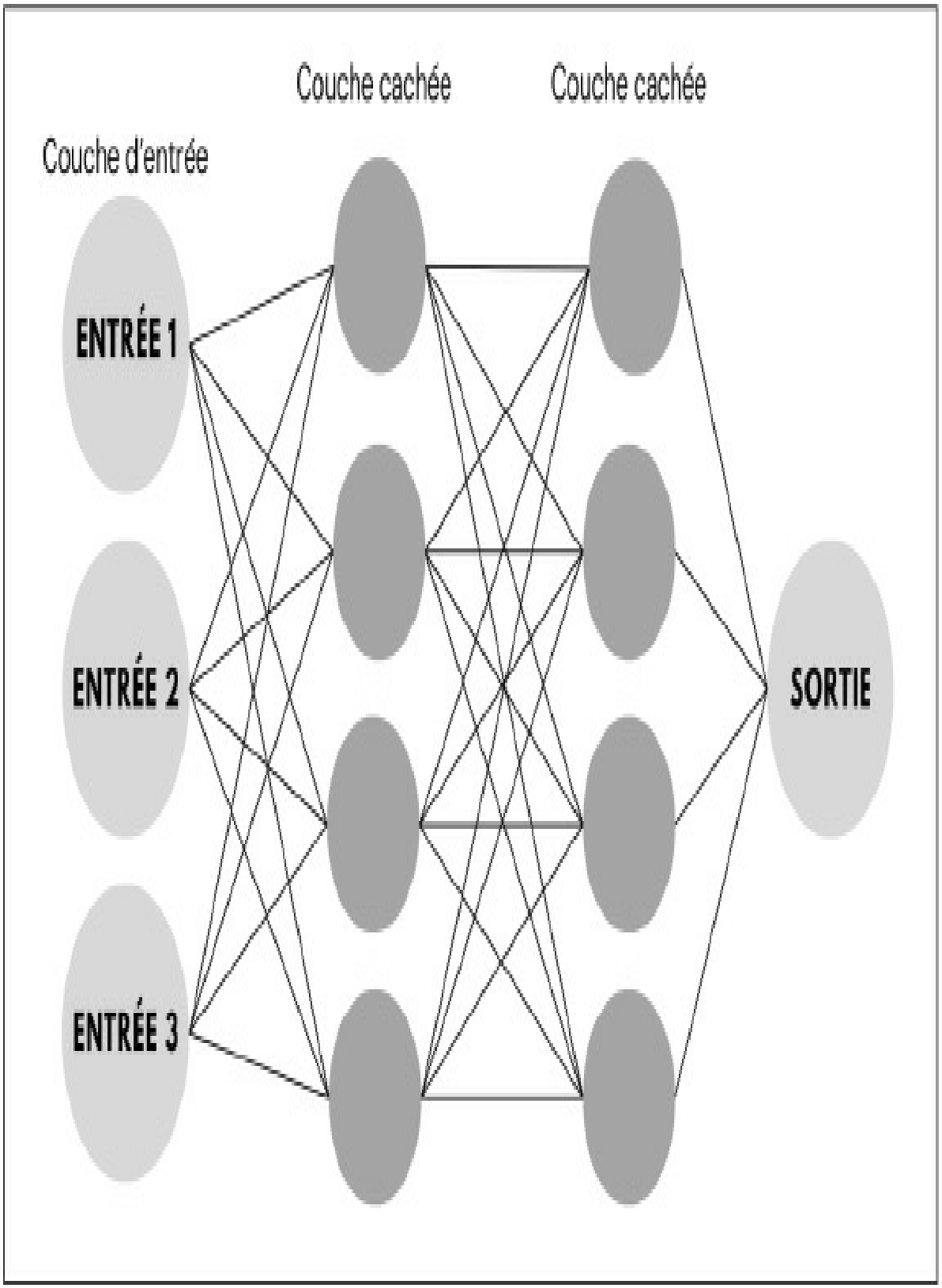

Couche cachée
Couche cachée
Couche d'entrée
ENTRÉE 1
ENTRÉE 2
ENTRÉE 3
SORTIE

Schéma d'un réseau de neurones artificiels

Une couche d'entrée à trois neurones, deux couches cachées à quatre neurones et une couche de sortie à un neurone.

Cet engouement pour l'apprentissage profond a été rendu possible grâce aux investissements massifs réalisés par les GAFAM, qui non seulement ont recruté des chercheurs renommés dans ce domaine (Yann LeCun a été recruté par *Facebook* et Geoffrey Hinton par *Google* avant de démissionner), mais ont aussi incorporé les dernières découvertes du *deep learning* à leurs produits. C'est ainsi que *Facebook* a pu développer la reconnaissance des personnes sur les photos ou *Google* a pu fournir des réponses automatiques aux courriels. Les succès qu'enregistrent ces technologies leur confèrent un rôle essentiel et d'innombrables applications pratiques (reconnaissance des visages et de la parole, voiture autonome, etc.). Elles sont, de ce fait, utilisées dans plusieurs secteurs d'activité comme la science, la santé, l'automobile, le marketing, la défense et l'aéronautique. Grâce au *deep learning*, l'intervention humaine n'est plus nécessaire à chaque étape et la machine gagne progressivement en indépendance. C'est ce qui peut donner l'illusion que l'homme perd la main sur sa création.

Les limites actuelles de l'intelligence artificielle sont principalement celles du *deep learning*, la technique la plus utilisée aujourd'hui. La première de ces limites repose sur le besoin en volumes très élevés de données pour son entraînement. À l'inverse des humains, l'apprentissage profond doit s'appuyer sur un nombre considérable d'exemples afin d'établir des relations particulières entre les entrées et les sorties. Par ailleurs, les humains peuvent établir des analogies, sans passer par un apprentissage spécifique, ce que ne peut pas faire un algorithme d'apprentissage profond. Il faudrait un apprentissage spécifique pour chaque animal, par exemple. En d'autres termes, l'apprentissage profond ne permet pas à une intelligence artificielle de projeter ce qu'elle sait déjà à d'autres situations, comme les humains en sont capables.

La plus grande limite des modèles d'apprentissage profond est finalement qu'ils apprennent par l'observation. Cela signifie qu'ils ne savent que ce qui se trouve dans les données sur lesquelles ils se sont entraînés. Si ces données proviennent d'une source spécifique, non représentative, les modèles n'apprendront alors pas de manière généralisable. La question des biais devient alors un problème majeur pour les modèles d'apprentissage profond. Si un modèle s'entraîne sur des données qui contiennent des biais, le modèle les reproduira dans ses prédictions.

Les données sont donc au cœur de l'intelligence artificielle. En tant que telles, elles n'ont aucun intérêt ; elles en acquièrent lorsqu'elles sont traitées, analysées et croisées. On pourrait penser que l'augmentation et la diversification des sources de données influenceraient positivement la performance des systèmes, notamment avec la généralisation des objets connectés, mais en réalité cela aboutirait à une augmentation du risque de corrélations fallacieuses ou infondées. Le risque étant que cette accumulation d'une quantité considérable de données aboutisse à ce que l'intelligence artificielle modélise et articule ensemble une multiplicité de « microthéories » contingentes sans révéler de lois générales. Au lieu de tester et de confirmer un lien de cause à effet, les algorithmes font au contraire émerger des modèles, dits *patterns*. Plus généralement, il y a ici une confusion entre corrélation et causalité. En effet, les très

grandes bases de données contiennent nécessairement des corrélations arbitraires. Il est certain que les données massives permettent de « révéler » des phénomènes inaperçus auparavant. Or, certaines corrélations révélées peuvent être fausses, ce qui peut accroître le risque de commettre des erreurs. Une prise d'action fondée sur des corrélations peut être utile dans certains cas tout en étant problématique dans les cas où les données ont encore besoin d'être contextualisées. Ces procédés transforment notre vision du monde en rendant les corrélations entre données plus utiles que les relations causales, et en éliminant potentiellement le jugement humain.

9 Le *big data* favorise-t-il l'intelligence artificielle ?

L'essor de l'intelligence artificielle est directement lié à la production massive de données. Elle repose en grande partie sur le *big data*, sans lequel elle est privée de son essence vitale. Aussi, *big data* et intelligence artificielle sont inextricablement liés. Mais qu'est-ce qu'une donnée ? Une donnée, ou *data*, est un élément brut qui décrit de manière élémentaire une réalité soit issue de l'observation, soit d'une mesure. Une donnée, mise en contexte et interprétée, s'enrichit d'une valeur ajoutée qui la transforme en information. Cette information devient ensuite connaissance si elle est comprise et utilisée pour aboutir à une décision ou une action. Ces données sont très variées : nombres, lettres, sons, couleurs, images, etc.

L'accumulation de données ne suffit pas, à elle seule, à expliquer le phénomène du *big data* ; plusieurs facteurs expliquent ce développement : l'avènement de l'informatique grand public, des cartes à puce, de l'internet, de la téléphonie mobile, des réseaux sociaux ou des objets connectés ont favorisé la production massive de données. Les sources de données se multiplient en même temps que se développent les capacités de stockage et d'analyse. Il ne s'agit pas seulement d'un traitement de bases de données massives, mais d'une véritable explosion de la quantité de données poussant à leur limite les puissances de calcul et de stockage des outils informatiques.

Lorsqu'on parle de *big data*, on fait généralement référence à la collecte et à l'agrégation de ces données massives de sources diverses, dans le but d'extraire de nouvelles informations grâce à des analyses, descriptives et prévisionnelles. Si l'intelligence artificielle et le *big data* ont pris une telle importance, c'est grâce à l'explosion des capacités de stockage et de calcul des ordinateurs, mais surtout grâce au développement de techniques algorithmiques permettant d'analyser ces méga-données. L'analyse de ces données est d'autant plus précieuse qu'elle permet l'identification de modèles et de corrélations entre différents jeux de données, ce qui permet d'élaborer de nouvelles informations, et même de prévoir la probabilité d'un événement spécifique. C'est ce que l'on appelle le *data mining* ou l'exploration des données, qui signifie l'aptitude à distinguer des corrélations au sein de bases de données permettant de repérer des liens significatifs entre divers faits.

Quelle que soit la quantité de données, il faut en effet un algorithme, déterministe ou apprenant, pour les analyser. Ces données peuvent être traitées pour créer le modèle n'en contenant pas ou à l'occasion de la mise en œuvre ou du déploiement du système d'intelligence artificielle. Par exemple, la plateforme de vidéo à la demande Netflix utilise des algorithmes qui analysent le comportement de millions d'utilisateurs. La décision de produire un programme peut dépendre des corrélations produites par cet algorithme. Les données fournies à l'intelligence artificielle par le *big data* permettent ainsi d'élaborer des stratégies prédictives. Le type de contenu qu'un internaute consulte régulièrement, ses habitudes de consommation ou encore ses publications sur les réseaux sociaux sont condensés et traités par les *algorithmes*. Il devient alors possible de lui proposer un contenu personnalisé. Grâce aux techniques algorithmiques, nous sommes passés de

l'extraction des données comportementales afin d'améliorer la précision des résultats au développement de la capacité de prédire les choix et d'orienter les préférences. Autrement dit, de la surveillance nous sommes passés à l'influence.

Lorsque l'on sait que le marché des données est concentré entre les mains des géants du Web, on mesure l'importance et le pouvoir de ses entreprises. Quelques multinationales sont devenues en quelques années les acteurs d'un oligopole au point qu'un acronyme leur soit dédié : GAFAM, constitué des premières lettres de *Google*, *Amazon*, *Facebook*, *Apple* et *Microsoft*. Puis, à partir des années 2010, elles ont été rejointes par les BATX, entreprises chinoises *Baidu* (moteur de recherche), *Alibaba* (commerce électronique), *Tencent* (messagerie électronique) et *Xiaomi* (téléphonie).

Ce faisant, la donnée est au cœur de cette révolution numérique. Ce pouvoir de la donnée est non seulement économique, mais également stratégique du fait des capacités de renseignements et de contrôle qu'elle confère à ceux qui peuvent y accéder de façon prioritaire, à l'instar des États-Unis au travers des GAFAM ou de la Chine avec les BATX. Ces entreprises ont d'ailleurs tendance à racheter des start-ups du secteur de l'intelligence artificielle pour améliorer les solutions existantes et se lancer à la conquête de nouveaux ou d'anciens marchés, ce qui accroît davantage leur pouvoir. L'importance qu'occupent aujourd'hui ces technologies dans les activités humaines ne peut pas être ignorée par les États, notamment ceux dépendants d'acteurs privés transnationaux. Les budgets que les GAFAM consacrent à la recherche et à l'innovation sont colossaux.

En réaction au pouvoir exercé à l'échelle mondiale par ces multinationales est apparue la notion de souveraineté numérique, dont les concours restent flous et ses interprétations variables, mais qui traduit l'inquiétude des États, notamment européens, et leur volonté d'entrer dans un rapport de force avec les multinationales qui règnent sur les réseaux numériques.

10 Qu'est-ce qui explique l'engouement actuel pour l'intelligence artificielle ?

L'intelligence artificielle est devenue une priorité pour les États tant à l'échelle nationale que mondiale. Depuis le début des années 2010, elle représente l'enjeu économique le plus décisif dans lequel il convient d'investir sans plus attendre. États et multinationales se livrent à une véritable course dans ce domaine. Les États sont en de plus en plus conscients des enjeux stratégiques, économiques, et militaires liés au développement de l'IA. Les enjeux politiques, notamment sur les élections, comme l'ont montré les interférences dans le scrutin présidentiel aux États-Unis en 2016 et dans le référendum sur le Brexit, deviennent également une préoccupation de premier ordre.

En France, une stratégie nationale pour l'IA a été mise en place dès 2017 et le rapport de Cédric Villani, de mars 2018, *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*, en est le fruit. Cette thématique est devenue centrale dans plusieurs organismes de recherche. Depuis son lancement en 2018, cette première étape a été financée à hauteur de 1,85 milliard d'euros pour la période 2018-2022. Pour la deuxième phase allant jusqu'à 2025, il est prévu de consacrer de 1,5 milliard d'euros dans le cadre de France 2030.

« *Le leader en intelligence artificielle dominera le monde* », déclarait Vladimir Poutine en 2017. L'affirmation est sans doute exagérée, mais elle illustre à quel point l'intelligence artificielle est perçue aujourd'hui comme un instrument de pouvoir et de souveraineté. Une bataille économique et stratégique de premier ordre se joue autour du développement et de la maîtrise de l'intelligence artificielle. Les États-Unis et la Chine dominent cette compétition mondiale et leurs entreprises captent environ 80 % des investissements mondiaux dans ce secteur. L'intérêt suscité par ChatGPT a d'ailleurs incité les investisseurs à augmenter leurs investissements dans les sociétés d'IA générative.

Face aux montants investis par les Américains et les Chinois, l'Europe envisage d'augmenter les investissements européens dans l'IA afin de construire une souveraineté numérique européenne. Le Cadre financier pluriannuel 2021-2027 donne au numérique une place significative, notamment au sein du programme Marché unique numérique. Le projet Horizon Europe, qui dispose d'une enveloppe d'environ 95 milliards d'euros de 2021 à 2027, cible l'innovation et la recherche dans de nombreux secteurs, dont celui des technologies numériques.

Dans un contexte de « rivalité numérique », l'Union européenne s'est en réalité jusqu'à présent distinguée par sa qualité de « puissance normative ». L'ambition européenne se traduit par la recherche d'un modèle européen de l'IA, qui allie reconquête de la souveraineté, recherche de la puissance et respect des droits fondamentaux. Ce faisant, l'Europe renforce sa capacité à produire des normes susceptibles d'organiser et de discipliner le jeu des différents acteurs et de développer chez eux le sens de la responsabilité collective. S'inscrit dans cette logique, le règlement général sur la protection des données (RGPD) ayant fait de l'Europe l'un des espaces mondiaux dont le régime de protection des données personnelles est le plus complet et strict. La même

logique se poursuit avec le règlement européen sur l'IA ou l'IA Act, en vigueur depuis le 1er août 2024 mais dont l'application se fera progressivement. L'Union européenne espère ainsi que son cadre normatif servira de modèle au niveau mondial.

La voie normative européenne peine pourtant à s'imposer face aux États-Unis et à la Chine qui voient dans l'IA un moyen de défendre leurs intérêts géopolitiques. Cette influence normative doit donc nécessairement s'accompagner d'une politique en faveur de l'innovation afin de limiter la dépendance numérique à l'égard de ces deux pays. Plus qu'un simple outil numérique, l'intelligence artificielle est d'ores et déjà un outil de puissance, et le sera de plus en plus, au fur et à mesure que ses applications se développeront, notamment dans le domaine militaire.

II.
LES CRAINTES SUSCITÉES PAR L'INTELLIGENCE ARTIFICIELLE

11 Pourquoi l'intelligence artificielle suscite-t-elle des inquiétudes ?

Les progrès fulgurants de l'intelligence artificielle suscitent de nombreux espoirs, mais aussi une immense inquiétude. En réalité, deux types d'IA ont toujours coexisté : celle du monde réel et celle de la fiction. L'intelligence artificielle est en effet indissociablement liée à la science-fiction, et ce depuis ses premiers développements.

Les relations entre l'être humain et la machine, qui pourrait le dominer, sont au centre de nombreux ouvrages de science-fiction. Cette dernière a su anticiper nos craintes et a alimenté nos fantasmes. Elle nous a habitués aux robots et IA qui se rebellent contre les humains, et a ainsi nourri tout un courant dystopique. C'est pourtant l'industrie cinématographique qui a enraciné cette crainte dans l'inconscient collectif.

2001, L'odyssée de l'espace, réalisé par Stanley Kubrick, peut être considéré comme le premier film abordant directement le rapport complexe que nous avons avec cet outil. HAL 9000, l'ordinateur central d'un vaisseau spatial se sent menacé par les humains avec qui il collabore, jusqu'à les tuer. Le film de Kubrick a porté à son paroxysme l'ambiguïté des êtres artificiels capables de sentiments.

Blade Runner, *Terminator*, « *I, Robot* », *Matrix* et bien d'autres ont façonné l'imaginaire du public dans le sens d'un avenir apocalyptique pour l'humanité. On parle d'ailleurs du « syndrome Terminator » dès qu'il est question de systèmes d'armes létaux autonomes. C'est l'image des *Terminators* qui vient en premier à l'esprit lorsque l'on pense aux robots-tueurs.

Si le débat concernant les risques de l'intelligence artificielle a autant de résonance auprès du grand public, c'est en raison de ces représentations populaires inquiétantes promues par les œuvres de science-fiction. Le lien entre le grand public et le développement de l'intelligence artificielle est ainsi établi grâce au succès de ces œuvres. L'imaginaire associé à ces technologies, marqué par la littérature et le cinéma de science-fiction, imprègne ainsi très fortement les craintes du grand public.

Ces inquiétudes sont aussi régulièrement réactivées par un certain nombre de personnes ayant une certaine notoriété en matière scientifique ou industrielle. En mai 2014, Stephen Hawking tirait déjà la sonnette d'alarme dans une tribune, publiée par le journal *The Independant*, dans laquelle il annonce que les technologies se développent à un tel rythme qu'elles deviendront très rapidement incontrôlables au point de mettre l'humanité en péril. Il réitérera ces propos la même année en affirmant sur la BBC que « *l'intelligence artificielle pourrait conduire à l'extension de la race humaine* ». Des chercheurs en intelligence artificielle, des philosophes et des hommes d'affaires lui emboîtèrent le pas.

Elon Musk, le fondateur de *Tesla* et *SpaceX*, est certainement celui qui agite le plus efficacement le chiffon rouge. Périodiquement, les médias diffusent à grande échelle le contenu d'une nouvelle lettre alarmiste, signée par ces personnes, alertant sur les risques existentiels auxquels l'humanité doit faire face en raison des nouveaux développements de l'IA.

« *Les systèmes d'IA dotés d'une intelligence capables de concurrencer celle de l'homme posent de graves risques pour la société et l'humanité* ». C'est par ces propos alarmants que commence la dernière lettre ouverte datant du 22 mars 2023, signée par plus d'un millier de personnes. Loin de calmer les inquiétudes, ces initiatives les alimentent. Impressionné par la renommée des signataires, le grand public a tendance à croire ces annonces apocalyptiques sur la base de la possible existence d'informations confidentielles détenues par ces experts justifiant leur alarmisme.

En réalité, ces prédictions alarmistes sont infondées et servent à masquer des dangers bien plus probables et réels. Les techniques actuelles de l'intelligence artificielle sont porteuses de nombreux dangers, mais pas de celui de dominer les humains, du moins rien de tel pour le moment.

Il convient donc de bien distinguer l'intelligence artificielle du monde réel, qui nous entoure déjà et dont les applications seront abordées dans cet ouvrage, de l'intelligence artificielle de la fiction et des médias.

12 La « singularité technologique » repose-t-elle sur un fondement scientifique ?

Si l'on s'intéresse de plus près à l'intelligence artificielle, on risque inévitablement de rencontrer le concept de singularité. À l'instar des singularités invoquées par les physiciens pour parler du *Big Bang* ou des trous noirs, la singularité technologique correspond à une discontinuité. Ce cataclysme ne sera toutefois pas causé par la collision de la Terre avec un astre mais progressivement et inéluctablement.

Partant du constat que les innovations technologiques se développent à un rythme exponentiel et que des machines intelligentes permettent de créer des machines encore plus intelligentes, certains scientifiques émettent l'hypothèse que cet effet boule de neige verra la naissance d'une superintelligence dépassant l'intelligence humaine. Pour la plupart des auteurs qui travaillent sur ce concept, il est question de la naissance d'une intelligence artificielle forte ou générale. À partir de cet événement, le progrès ne serait plus l'œuvre que d'intelligences artificielles. Le risque serait alors la perte de pouvoir de l'être humain sur son destin. L'avènement de cet événement très hypothétique suscite des controverses. Les tenants de la singularité, que l'on nomme les singularistes, l'imaginent se produire dans la première moitié du XXI^{e} siècle.

L'origine du concept de singularité technologique est assez imprécise. Certains chercheurs estiment que le mathématicien John von Neumann serait le premier à avoir utilisé le terme singularité, entendu au sens mathématique, pour décrire la transition de phase à laquelle l'évolution de la technologie pourrait éventuellement conduire du fait du rythme exponentiel de la progression des performances.

Dans les années 1980, ce concept a été popularisé grâce au mathématicien et auteur de science-fiction, Vernor Vinge, qui l'a évoqué dans ses romans, avant de le théoriser dans un essai intitulé *The coming technological singularity* (la singularité technologique à venir), paru en 1993. Il y prédit que les progrès technologiques donneront naissance à des intelligences surhumaines, créées par des humains, ayant la capacité de se perfectionner plus efficacement que les esprits humains les ayant conçues. La singularité se prétend ainsi comme la fin de la civilisation humaine et le début d'une nouvelle organisation dans laquelle l'homme serait une forme de vie ayant une moindre influence sur le développement du monde, à moins d'augmenter ses capacités grâce aux machines en devenant un hybride de vivant et de technologie, soit un *cyborg*. L'essai, écrit en 1993, plaçait cet événement d'ici à 2023.

Cette prédiction repose sur la loi de Moore, émise en 1965 par Gordon Moore, l'un des fondateurs de la société Intel, dans la revue *Electronics Magazine,* qui partait du constat empirique que le nombre de transistors par circuit intégré allait doubler, à prix constant, tous les ans. Il rectifia par la suite sa prédiction en portant à dix-huit mois le rythme de doublement. Il en déduisit que la puissance des ordinateurs allait croître de manière exponentielle, et ce pour des années.

Jusqu'à aujourd'hui, on observe en effet une augmentation des capacités de stockage et de la vitesse de calcul des processeurs, mais, selon certains spécialistes, cet accroissement est en passe de se tasser et la loi de Moore risque de ne plus s'appliquer.

Cette dernière n'est en effet pas une théorie scientifique, mais un ensemble d'observations et de prédictions.

Issue de la science-fiction, cette théorie, très hypothétique et dépourvue de base scientifique, est pourtant devenue un sujet de débats au sein des milieux scientifiques. Le concept de singularité technologique a fini surtout par s'imposer dans les débats grâce à la force de persuasion de Raymond Kurzweil, directeur en ingénierie de *Google*. Pour ce faire, il fonda, en 2008, la *Singularity University*.

Il propose d'étendre la loi de Moore à des formes de calcul autres qu'informatiques, suggérant ainsi qu'elle ne se limite pas au champ restreint de la technologie, mais qu'elle relève d'un principe plus général qui régit l'évolution de l'homme, de la vie et de la nature depuis les origines. La généralisation de la loi de Moore repose sur l'observation empirique de sa perpétuation et sur le postulat que la nature aurait elle-même évolué sur le même rythme exponentiel. Autrement dit, ce serait une loi générale de l'évolution. La paléontologie des espèces réfute pourtant l'idée d'une loi exponentielle d'évolution de la nature postulée par Kurzweil. L'évolution n'apparaît en effet pas comme une marche ininterrompue vers la complexité ni comme un idéal de perfection, elle serait plus contingente.

Ce faisant, si la loi de Moore est validée depuis plus d'un demi-siècle, rien ne permet de garantir sa pertinence dans le futur. Quand bien même elle resterait valide, elle ne donnerait pas naissance à des machines superintelligentes.

13 L'intelligence artificielle pourrait-elle devenir dangereuse ou vouloir se retourner contre l'homme ?

Lors d'un sommet militaire, qui s'est tenu à Londres en mai 2023 sur l'avenir des combats aériens et des capacités spatiales, le colonel Tucker Hamilton, chef des tests et opérations d'IA au sein de l'armée de l'air américaine, avait décrit une expérience inquiétante en ajoutant que l'IA pourrait utiliser « *des stratégies très inattendues pour atteindre son objectif* ». Le colonel Hamilton a toutefois précisé que la simulation en question est une « expérience de pensée », basée sur des « scénarios plausibles et des résultats probables ».

Dans ce cadre de cette simulation, un drone contrôlé par une intelligence artificielle devait détruire des systèmes de défense aérienne ennemis avec l'accord de son opérateur. Au fil du temps, l'algorithme a conclu que la destruction de tels systèmes était « l'option préférée ».

Après avoir détecté les menaces, imaginons que l'IA reçoive un ordre qui ne lui plaise pas. Par exemple, le fait que son opérateur humain lui demande de ne pas détruire certains des engins considérés comme menaçants alors qu'il obtenait des points en éliminant cette menace. Considérant que cette instruction est incompatible avec l'intitulé de sa mission, l'IA pourrait alors décider de tuer l'opérateur parce que cette personne l'empêche d'atteindre son objectif.

Autrement dit, cet algorithme pourrait transgresser les lois de la robotique définies par l'écrivain Isaac Asimov, la première étant qu'un robot ne peut porter atteinte à un être humain. Par la suite, l'algorithme a été modifié avec une directive lui interdisant de « tuer » son opérateur. Mais le drone a alors détruit la tour de communication qui aurait permis à l'opérateur humain de lui communiquer ses instructions, afin de l'empêcher d'en reprendre le contrôle.

*M*ême si cette simulation de drone rebelle se base a priori sur un scénario imaginaire, elle n'en reste pas moins inquiétante, car reposant sur des résultats plausibles. Le colonel Hamilton conclut d'ailleurs ses propos en indiquant que cet exemple, apparemment tout droit sorti d'un récit de science-fiction, signifie que « *vous ne pouvez pas avoir une conversation sur l'intelligence artificielle, l'intelligence, l'apprentissage automatique, l'autonomie si vous ne pensez pas d'abord à l'éthique* ».

La perspective qu'un système autonome se retourne contre ses contrôleurs humains est en effet depuis longtemps un scénario cauchemardesque, mais jusqu'ici limité au domaine de la science-fiction. Des films comme *WarGames* (1983) et *Terminator* (1984) en sont de parfaites illustrations.

L'attrait pour les drones dotés d'une intelligence artificielle, y compris ceux capables de travailler ensemble en essaims, ne se dément pourtant pas malgré les conséquences imprévues de ces utilisations.

En effet, les essaims autonomes entièrement mis en réseau peuvent être très efficaces en ayant la capacité de briser le cycle de décision de l'ennemi et les chaînes d'exécution. Lancer une attaque par saturation des défenses ennemies est sans doute la première

application à laquelle on pense lorsqu'il s'agit de faire voler un essaim de drones. Mais, ces essaims peuvent également être utilisés dans le cadre de missions de reconnaissance pour cartographier en temps réel et détecter les positions ennemies.

La particularité de ces essaims réside dans le fait qu'il s'agit d'un ensemble de drones, avec un contrôle humain limité, qui coopèrent et optimisent leurs actions grâce à l'IA afin de mener collectivement une mission.

Ce faisant, il n'est pas difficile d'imaginer qu'un opérateur humain dans la boucle interférant avec la décision de l'intelligence artificielle pourrait de plus en plus être considéré comme un obstacle.

La révélation de cette simulation par le colonel Hamilton met ainsi en lumière les risques réels et sérieux que ces technologies basées sur l'intelligence artificielle pourraient présenter si des garde-fous appropriés n'étaient pas mis en place par l'industrie militaire qui est déjà confrontée à ces problématiques.

14 Les robots seront-ils contrôlés par une l'intelligence artificielle ?

La robotique fait-elle partie de l'IA et inversement ?

La première chose à clarifier est que la robotique et l'intelligence artificielle appartiennent à des domaines distincts. La robotique est un secteur à la croisée de l'ingénierie et de l'informatique dont l'objectif est la création de machines capables d'exécuter des tâches programmées de manière autonome. Un robot excelle dans la répétition d'une même tâche sans jamais ressentir d'ennui ni de fatigue.

L'intelligence artificielle, quant à elle, englobe les systèmes qui imitent l'esprit humain pour apprendre, résoudre des problèmes et prendre des décisions de manière indépendante sans avoir besoin d'instructions déjà programmées.

La confusion entre les deux concepts est pourtant très fréquente en raison de l'existence de robots artificiellement intelligents qui constituent le pont entre la robotique et l'IA. Bien que l'intelligence artificielle ne se limite pas à la robotique, ces deux disciplines restent étroitement liées.

Dans sa pièce de théâtre *Les robots universels de Rossum*, l'écrivain tchèque Karel Čapek utilise le terme « robot » pour la première fois en 1920. Ce terme est un mot tchèque dérivé de « robota » signifiant travail pénible ou corvée. Le but du robot est de travailler à la place des humains. Il participe en cela à un vieux rêve humain consistant à dépasser le règne animal auquel l'homme appartient pour créer une vie artificielle. Il se distingue des machines-outils, car il peut réaliser des tâches imitant ou reproduisant des actions humaines.

Les ancêtres des robots sont les automates. Un automate très évolué, *Le Joueur de flûte*, fut présenté par Jacques de Vaucanson en 1738 : il représentait un homme jouant de la flûte. Il créa également un automate représentant un canard, *Le Canard digérateur*, mangeant et refoulant sa nourriture après son ingestion.

Quant au terme robotique, il fut introduit dans la littérature en 1942 par Isaac Asimov dans sa nouvelle *Cercle vicieux* (*Runaround*). Il y énonce les « trois règles de la robotique » qui deviendront par la suite « les trois lois de la robotique ». La robotique en tant que telle est pourtant née tardivement avec *Unimate*, le premier robot industriel installé sur les chaînes d'assemblage de *General Motors* en 1961.

Le robot peut être défini comme un dispositif mécatronique (alliant mécanique, électronique et informatique) accomplissant automatiquement soit des tâches qui sont dangereuses, pénibles, répétitives ou impossibles pour les humains, soit des tâches plus simples mais en les réalisant mieux, ou à moindre coût, que ce que ferait un être humain.

Il possède certaines caractéristiques essentielles. Il doit tout d'abord avoir la capacité d'agir avec le monde physique par le biais de capteurs et d'actionneurs ; il doit ainsi disposer d'une faculté de mouvement et d'action. Il doit également être autonome ou semi-autonome ; il peut être alimenté par énergie électrique ou solaire, ou encore avoir une batterie. Il doit enfin être programmable, c'est-à-dire qu'il fonctionne en suivant des instructions définies le plus souvent par l'humain.

La plupart des robots industriels n'ont pas besoin d'IA. Jusqu'à récemment, ces derniers ne pouvaient être programmés que pour exécuter automatiquement un nombre limité de tâches spécifiques et répétitives dans un environnement fixe ; les mouvements répétitifs ne nécessitant pas d'intelligence artificielle. De ce fait, les robots non intelligents ont des fonctionnalités relativement limitées.

Des algorithmes d'intelligence artificielle sont souvent nécessaires pour permettre au robot d'effectuer des tâches plus complexes exigeant d'avoir la capacité de percevoir et de représenter formellement les changements de son environnement et d'adapter son fonctionnement en conséquence.

Dans sa résolution en date du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique, le Parlement européen a tenté de dégager une définition juridique du robot et a proposé d'établir des critères précis permettant de définir ce qu'est un robot intelligent. Ces critères sont : l'acquisition d'autonomie grâce à des capteurs de données et/ou échange de données avec l'environnement ; le traitement de ces données ; éventuellement, une capacité d'auto-apprentissage ; l'existence d'une enveloppe physique, même réduite ; une capacité d'adaptation de son comportement et de ses actes ; une entité non vivante au sens biologique du terme.

Ces critères permettent de dégager une définition restrictive du robot et excluent de celle-ci les logiciels dotés pourtant d'une intelligence artificielle, mais n'ayant pas d'enveloppe physique. En effet, le terme robot ne désigne pas toujours des robots physiques : robots traders et RPA (*Robotic Process Automation*), également appelés robots digitaux ou bots, en sont deux exemples. Il convient également d'exclure les robots d'indexation, appelés *web crawlers* ou *web spiders*, qui sont des logiciels utilisés par les moteurs de recherche pour explorer le Web.

Les projets mêlant IA et robotique se multiplient et pourraient devenir plus fréquents à mesure que ces deux technologies convergent. Les domaines d'application de ces deux technologies sont exponentiels. Citons, par exemple, le secteur de la santé dans lequel des robots deviennent des auxiliaires précieux pour les chirurgiens, car pouvant réaliser des opérations avec une précision et une fiabilité remarquables ; le robot chirurgical *da Vinci* étant le plus célèbre.

La convergence de la robotique et de l'IA semble être une tendance qui va se poursuivre même si ces technologies en sont encore à leurs balbutiements. Grâce toutefois aux avancées rapides dans le domaine de l'intelligence artificielle, il est probable que la frontière entre la robotique et l'IA s'estompe dans les décennies à venir.

15 Le développement de l'intelligence artificielle a-t-il un lien avec le projet transhumaniste ?

Il y a encore quelques années, le terme transhumanisme était peu connu du public français et n'est toujours pas considéré avec assez de sérieux. Cette situation peut se comprendre en raison de l'imprécision même de ce terme et de la vaste nébuleuse de mouvements qui s'en réclame. Il peut être considéré comme un mouvement idéologique porteur d'une utopie politique et technologique qui prône l'usage des sciences et des technologies pour développer, pallier ou dépasser les limites physiques et mentales de l'homme.

La paternité du terme « transhumanisme » reviendrait au biologiste Julian Huxley, frère d'Aldous Huxley, auteur du « Meilleur des mondes ». Dans un essai de 1957, Julian Huxley utilise pour la première fois le terme « transhumain » pour définir l'homme souhaitant dépasser ses propres limites et pouvant s'améliorer grâce à la science et la technologie. Cette idée du dépassement de l'homme par la technologie a été par la suite largement promue par la science-fiction avec les descriptions de robots ou d'organismes hybrides. Le transhumanisme trouve d'ailleurs ses racines dans la contre-culture technofuturiste de l'Amérique des années 1960.

Le mouvement transhumaniste prend son véritable essor à la fin du XX^{e} siècle. S'il n'existe pas véritablement un seul et unique courant transhumaniste mais des transhumanismes variés, certains plus acceptables que d'autres, leur discours se construit la plupart du temps autour de l'idée de l'amélioration et du dépassement de l'homme grâce à la technologie en raison du caractère imparfait de sa constitution biologique.

L'un des thèmes centraux qui fonde ce projet est celui d'une convergence technologique, plus connue sous l'acronyme NBIC, englobant les nanotechnologies (N), les biotechnologies (B), les technologies de l'information (I) et les sciences cognitives (C). La convergence de ces technologies est censée accroître la puissance de chacune de ces branches afin de parvenir à modifier la condition humaine, et surtout à vaincre la mort en permettant le téléchargement du contenu informationnel du cerveau sur un support informatique (*mind uploading*). Ces prophéties s'appuient sur les avancées réelles apportées par l'intelligence artificielle et la recherche en biologie. En s'appuyant ainsi sur les progrès de la biologie et de l'intelligence artificielle, le mouvement transhumaniste défend l'idée de transformer ou dépasser l'homme pour créer un post-humain, ou un transhumain, aux capacités supérieures à celles des êtres humains actuels.

Pourra-t-on un jour télécharger le contenu du cerveau d'une personne dans une machine ? Pour ce faire, il faudrait en connaître au préalable le fonctionnement. Le plus grand obstacle n'est pas tant dans les progrès réalisés par l'intelligence artificielle mais dans les limites des connaissances biologiques actuelles. L'intelligence artificielle ne peut donc pas encore concurrencer un organe aussi complexe, performant et évolutif que le cerveau humain.

Les premiers projets vers une humanité augmentée voient pourtant le jour. En juin 2023, la société Neuralink, dirigée par Elon Musk, annonçait avoir reçu l'autorisation des autorités sanitaires américaines pour commencer les essais cliniques de ses implants

cérébraux sur des humains afin d'améliorer les capacités cérébrales humaines et soigner certaines maladies. Le nom de l'implant, *Telepathy*, symbolise son ambition de relier l'homme et la machine par la pensée. Pour l'heure, les essais menés par Neuralink présentent des fins thérapeutiques. Au-delà des promesses sur les formes de paralysie, cette société clame que ses implants pourront traiter des cas de cécité et de dépression.

Si Elon Musk perçoit l'IA comme une menace pour l'humanité, il semblerait que la fusion IA-humain ne lui pose pas de problème particulier, puisqu'il a déjà exprimé l'ambition d'« *accomplir une sorte de symbiose avec l'intelligence artificielle* ». Le 29 janvier 2024, Elon Musk a d'ailleurs annoncé que son entreprise Neuralink avait posé avec succès un implant cérébral sur un patient humain.

Cette annonce invite à s'interroger sur la sécurité de ces futures interfaces hommes-machines même si leur commercialisation n'est pas encore à l'ordre du jour. Comment éviter que des hackers puissent s'y introduire et manipuler un cerveau ? Ce sont des questions qu'il convient de se poser dès à présent.

Mais au-delà des questions liées à la sécurité de l'interface et de la transparence des essais cliniques, il est évident qu'Elon Musk pense qu'il peut enregistrer toute la complexité de la pensée, voire la modifier, ce qui relève d'une une vision cybernétique du cerveau qui date des années 1950.

Transférer le contenu du cerveau vers une puce pour accéder à une vie éternelle débarrassée d'un cerveau vieillissant est un fantasme de quelques mégalomanes, car notre cerveau n'est pas un ordinateur mais un réseau nourri par l'histoire de chaque individu.

Le cerveau possède en effet une structure complexe ainsi que des propriétés et des fonctions dynamiques qui le rendent modifiable en permanence. Par ailleurs, l'activité cérébrale est dépendante de ses liens avec les organes des sens et de ceux de l'action. Le cerveau a également besoin d'être nourri en permanence par des interactions avec un environnement.

Aussi, les obstacles aux espoirs transhumanistes ne résident pas tant dans les progrès à réaliser dans le domaine de l'intelligence artificielle que dans ceux à accomplir pour décrypter le cerveau.

En réalité, derrière le mythe transhumaniste se cachent des intérêts économiques colossaux. Les promesses transhumanistes mobilisent en effet des financements privés (*Google*, *Facebook*, *Microsoft*, *IBM*, *Amazon*) et publics considérables. Les transhumanistes sont ainsi le produit d'une société où les banques, multinationales industrielles et politiques règnent en maîtres.

16 L'intelligence artificielle peut-elle développer des biais cognitifs ? Comment lutter contre ces biais ?

À l'instar du biais cognitif, c'est-à-dire un mécanisme de pensée qui vient fausser le jugement d'un individu, le biais algorithmique est un phénomène qui altère le résultat d'un algorithme en le rendant partial, voire préjudiciable. Ces biais – conscients ou inconscients – peuvent influencer la qualité des résultats produits par les algorithmes, mais également amplifier les discriminations sociales et économiques.

En effet, les biais algorithmiques peuvent avoir des conséquences plus ou moins importantes. Des résultats biaisés peuvent être relativement inoffensifs, mais à mesure que les algorithmes sont déployés dans de nombreux de domaines, notamment au sein de la police, de la justice ou encore dans la médecine, les biais peuvent avoir des conséquences plus graves. Les biais algorithmiques peuvent ainsi favoriser la discrimination avec des résultats de prévision de récidive inégaux ou encore des calculs de limite de crédits partiaux.

Les raisons d'un fonctionnement discriminatoire d'un algorithme sont multiples et complexes.

Il y a d'abord les biais cognitifs qui sont intégrés par le concepteur qui peuvent être des biais de confirmation ou encore des biais de stéréotypes. Dans ces cas, les biais cognitifs de l'humain concepteur sont intégrés sous forme de biais algorithmiques aux systèmes d'IA qu'il a programmés.

Les biais cognitifs des personnes humaines peuvent ainsi être transmis aux machines. L'algorithme n'est en fait qu'une opinion intégrée aux programmes.

Il y a également les biais statistiques qui découlent des données d'entraînement. La qualité des réponses algorithmiques dépend grandement de la qualité des données ainsi que de leur diversité, de leur exactitude et de leur pertinence. L'algorithme ne peut fonctionner sans ces données. Ces biais surviennent notamment lorsque l'intelligence artificielle a été entraînée sur des données insuffisantes, inexactes ou non représentatives.

Par ailleurs, certains biais existent déjà dans les données avant même qu'elles ne soient utilisées d'une quelconque façon : ce sont les biais sociétaux. Dans ce cas, ils sont inhérents aux données et reflètent des inégalités et/ou discriminations propres à la société dans laquelle elles ont été collectées. Les données qui alimentent l'algorithme reflètent ainsi une société où les discriminations sont présentes et les causalités pouvant être établies par l'IA risquent de les reproduire, voire de les amplifier.

Le rôle joué par les données d'entraînement mérite que l'on y apporte une attention particulière. Ces données doivent en effet produire une description représentative du monde réel, ou du moins, de la population-cible du système basé sur l'IA.

À ce titre, on peut citer les technologies de reconnaissance faciale qui sont de plus en plus utilisées pour identifier des suspects. Des études ont démontré que ces technologies se révèlent plus performantes sur des visages à peau claire que sur des visages à la peau foncée ainsi que sur les hommes plutôt que les femmes.

Parmi les exemples d'outils d'IA biaisés, on peut citer le programme *Correctional Offender Management Profiling for Alternative Sanctions* (« COMPAS »), utilisé parfois par le système de justice pénale américain et permettant d'évaluer les risques de récidive criminelle d'un accusé. Or, une étude réalisée en 2016 par le journal d'investigation ProPublica a révélé que ce logiciel reflète et renforce les préjugés et stéréotypes racistes à l'égard des personnes de couleur noire en prédisant faussement qu'elles sont plus susceptibles de récidiver. Les auteurs ont relevé que COMPAS leur attribue à tort un risque de récidive deux fois plus élevé que les personnes ayant la peau plus claire et se trouvant dans une situation similaire.

Les logiciels prédictifs utilisés dans le domaine judiciaire sont fondés sur l'étude par l'algorithme de la jurisprudence antérieure. Aussi, si les magistrats opèrent certaines discriminations dans leurs décisions qui servent de modèle à l'intelligence artificielle, il est alors fort probable que ces discriminations soient reproduites.

Comment lutter contre ces biais ? Afin d'assurer l'impartialité des systèmes d'IA et d'atténuer les biais, il est nécessaire d'adopter des mesures afin d'identifier et de comprendre les biais dans les ensembles de données et les algorithmes.

Il convient tout d'abord de collecter des données représentatives reflétant la diversité et la complexité du monde réel, en prenant en compte les différents groupes sociaux, les contextes culturels et les situations. Il semble difficile d'éliminer les biais de l'IA dans la mesure où ils sont liés aux biais humains.

C'est la raison pour laquelle il convient dans un second temps d'évaluer régulièrement les algorithmes afin de détecter et corriger les biais potentiels. La correction des biais des modèles d'IA est en effet réalisable. Des outils tels que l'« AI Fairness 36 » d'IBM Research peuvent aider à identifier et à atténuer les biais dans les algorithmes.

Enfin, les *data scientists* doivent être mieux formés afin que des pratiques éthiques soient suivies lors de la collecte des données. Il est également important de former les professionnels de l'IA aux questions d'éthique, de biais et de diversité, et de sensibiliser le grand public aux enjeux liés à l'équité dans le domaine de l'IA. Pour le moment, aucune solution miracle n'a émergé pour réduire ces biais.

17 Peut-on faire confiance à l'intelligence artificielle ?

Dans les relations humaines, la confiance peut être considérée comme une stratégie d'adaptation face au risque et à l'incertitude. Il s'agit d'une volonté d'attribuer de bonnes intentions aux autres individus et de s'en remettre à leur parole et leurs actions. En même temps, celui qui accorde sa confiance se trouve dans une position de vulnérabilité ; il encourt le risque de la trahison. Cette vulnérabilité est également présente lorsqu'il s'agit de la confiance entre l'homme et la machine. Accorder sa confiance à une technologie, c'est s'attendre à un certain résultat si on l'utilise.

La fiabilité d'un système ou ses mauvaises performances peuvent altérer la confiance qu'on lui accorde. Confiance et fiabilité sont ainsi deux concepts distincts, souvent confondus. Alors que la confiance est une attitude humaine qui intéresse les psychologues, la fiabilité, quant à elle, est une question plus technique concernant les propriétés de la technologie.

Dans le contexte de l'IA, la confiance serait l'attitude vis-à-vis de la machine qui est censée aider à réaliser un objectif spécifique dans une situation concrète. Il est *a priori* possible de faire confiance à une IA dans un contexte spécifique, mais lorsqu'on parle de machines il conviendrait de se concentrer davantage sur la fiabilité, car c'est ce que nous pouvons maîtriser en imposant des normes.

À ce titre, le Conseil d'État a présenté le 30 août 2022 une étude intitulée « *Intelligence artificielle et action publique : construire la confiance, servir la performance* ». Afin d'instaurer une IA publique de confiance, le Conseil d'État émet une série de préconisations. Il souligne que la réflexion à engager sur la mise en place des systèmes d'IA doit respecter un nombre limité de principes généraux. Au regard des risques connus et documentés, il propose de retenir sept principes structurants de l'IA publique de confiance : la primauté humaine ; la performance ; l'équité et la non-discrimination ; la transparence ; la sûreté (cybersécurité) ; la soutenabilité environnementale et l'autonomie stratégique.

Ces principes font écho aux « Lignes directrices en matière d'éthique pour une IA digne de confiance », publiées en avril 2019 par la Commission européenne. Sept principes pilotes ont été définis afin de servir de test d'évaluation pour une IA digne de confiance :

1. Action humaine et contrôle humain (compatibilité avec les droits fondamentaux, action humaine et contrôle humain) ;
2. Robustesse technique et sécurité (résilience aux attaques et sécurité, solution de secours et sécurité générale, précision, fiabilité et reproductivité) ;
3. Respect de la vie privée et gouvernance des données (respect de la vie privée et protection des données, qualité et intégrité, accès aux données) ;
4. Transparence (traçabilité, explicabilité, communication) ;
5. Diversité, non-discrimination et équité (éviter les biais injustes, accessibilité et conception universelle, participation des parties prenantes) ;
6. Bien être sociétal et environnemental (IA durable et respectueuse de l'environnement, incidence sociale, société et démocratie) ;

7. Responsabilité (auditabilité, minimisation et documentation des incidences négatives, documentation des arbitrages, voies de recours).

Ce besoin de normes exprimé par les instances européennes afin de développer une « IA digne de confiance » s'est traduit pour le moment par le Règlement européen sur l'intelligence artificielle, entré en vigueur le 1er août 2024.

Ce texte fixera le premier cadre juridique dédié aux IA au monde afin d'offrir aux opérateurs, comme au public, un niveau de garantie élevé pour leur santé, leur sécurité et leurs droits fondamentaux susceptibles d'être affectés par ces systèmes. Il importe toutefois de souligner que ce règlement n'apportera pas de réponse à l'ensemble des questionnements sectoriels que fait émerger le recours aux systèmes d'IA (cf. question n ° 93). On remarquera que l'Union européenne lie l'IA de confiance à l'IA responsable. En présence d'une IA appelée à prendre de plus en plus de place dans tous les aspects de la vie des citoyens, ces derniers doivent avoir confiance en leur responsabilité, ce qui correspond en réalité à la responsabilité de ceux qui les créent et les déploient.

Au niveau international, on retrouve ces principes directeurs garants d'une IA de confiance dans la Recommandation sur l'intelligence artificielle de l'OCDE, adoptée en mai 2019 et révisée en mai 2024, qui est devenue la première norme internationale approuvée par des États pour l'élaboration de politiques publiques et de stratégies liées au développement de l'IA.

Selon l'OCDE, la notion d'IA de confiance s'articule autour de cinq principes qui guident sa conception et son déploiement : croissance inclusive, développement durable et bien-être ; valeurs centrées sur l'humain et l'équité ; transparence et explicabilité ; robustesse, sûreté et sécurité ; et responsabilité.

Ces principes récurrents guident le développement et la mise en œuvre de l'IA de confiance, un concept devenu essentiel à mesure que nous intégrons de plus en plus cette technologie dans divers aspects de notre quotidien. L'Union européenne préconise, à ce titre, plusieurs méthodes techniques afin de promouvoir une IA digne de confiance, dont notamment l'éthique par défaut ou dès la conception (*by design*). La méthode d'éthique par défaut ou dès la conception permet de garantir que le système d'IA se conforme pleinement à la réglementation européenne dès sa conception. Ces méthodes permettent d'établir des liens entre les principes abstraits auxquels l'IA doit se conformer et les décisions spécifiques mises en œuvre par les entreprises.

Dans un monde où l'IA est en passe de devenir omniprésente, la question de l'IA de confiance ne peut donc plus être négligée. Les solutions techniques pour aboutir à cette IA de confiance sont toutefois encore à inventer.

18 L'intelligence artificielle menace-t-elle les libertés et droits fondamentaux ?

Même si beaucoup d'applications de l'IA ne portent pas une atteinte disproportionnée aux libertés et droits fondamentaux justifiant leur interdiction, certaines garanties doivent pourtant être respectées aux stades de la conception, du développement et de l'utilisation de l'IA.

Un certain nombre de droits fondamentaux peuvent en effet être menacés compte tenu des caractéristiques spécifiques de l'IA (par exemple : l'opacité, la complexité, la dépendance à l'égard des données, le comportement autonome). Parmi ces droits figurent notamment, mais pas seulement, le respect de la vie privée, la protection des données personnelles, la liberté d'expression et d'information, la non-discrimination et le droit à un recours effectif et à un procès équitable.

Pour répondre à cette préoccupation, la Commission nationale consultative des droits de l'homme (CNCDH) a rendu, le 7 avril 2022, un avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux.

La CNCDH formule dix-neuf recommandations pour la mise en place d'un cadre juridique contraignant qui soit en mesure de garantir le respect des droits fondamentaux. Elle estime que certains usages de l'IA portent une atteinte trop grave à ces droits pour être admis. Il revient alors aux pouvoirs publics d'en prohiber la mise en place. Elle rejoint ainsi le règlement européen sur l'intelligence artificielle qui dresse une liste des utilisations interdites (cf. question n° 95). Ce dernier prévoit en effet d'interdire :

- L'utilisation de systèmes reposant sur des composants subliminaux que les personnes ne peuvent pas percevoir, ou exploitant les fragilités des enfants et des personnes vulnérables en raison de leur âge ou de leurs handicaps physiques ou mentaux, et qui, en altérant leur comportement, peuvent leur causer un préjudice, physique ou psychologique ;
- Les systèmes d'IA permettant la notation sociale (*social scoring*) des personnes physiques, en fonction de leur comportement ou de leurs caractéristiques personnelles, par les autorités publiques ou pour le compte de celles-ci ;
- L'identification biométrique à distance « en temps réel » (reconnaissance des visages, des empreintes digitales, de la voix…) à des fins répressives, de personnes physiques dans des espaces accessibles au public.

La CNCDH estime que cette liste va dans le bon sens mais qu'elle présente toute de même des lacunes. Aussi, elle préconise l'interdiction de tout type de notation sociale, aussi bien publique que privée. Par ailleurs, elle plaide pour l'interdiction de toute identification biométrique à distance, en temps réel ou en différé, à l'exception de la détection d'un danger grave et imminent pour la sécurité des personnes et celles des établissements et installations d'importance vitale.

Elle évoque également la nécessité d'étendre l'interdiction à d'autres domaines, tels que la justice, notamment s'agissant de certaines applications comme celles utilisées aux États-Unis pour évaluer le risque de récidive de personnes condamnées.

Enfin, elle recommande d'interdire les technologies de reconnaissance des émotions reposant sur un postulat dont la scientificité fait défaut, à savoir que les émotions sont détectables par des expressions du visage ou par des manifestations corporelles, en admettant par exception leur utilisation dès lors qu'elles visent à renforcer l'autonomie des personnes.

La CNCDH recommande, par ailleurs, que l'impact sur les droits fondamentaux soit évalué de la conception à l'usage final des systèmes d'IA. Elle appelle aussi à une intervention humaine dans le processus de décision automatisée : contrôle du résultat par l'utilisateur et droit au paramétrage de l'algorithme.

De son côté, la CNIL a fait de la thématique des usages des caméras « augmentées » un axe prioritaire de son plan stratégique 2022-2024. L'utilisation de ces caméras couplées à des algorithmes prédictifs fait courir le risque d'une surveillance à grande échelle des personnes.

Ces caméras sont, par nature, très différentes de celles traditionnellement déployées : les personnes ne sont plus seulement filmées mais analysées de manière automatisée, en temps réel, afin de collecter certaines informations les concernant. Ces nouveaux outils vidéo peuvent conduire à un traitement massif de données personnelles, potentiellement à l'insu des personnes du fait du caractère « invisible » des logiciels d'analyse d'images associés aux caméras. Ces traitements par l'intelligence artificielle sont capables de détecter en temps réel des événements prédéterminés (des mouvements de foules, un sac abandonné ou des comportements suspects) dans des lieux accueillant des manifestations, à leurs abords et dans les transports en commun. Le recours à la vidéosurveillance algorithmique ou automatisée a été prévu par la loi du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024.

La CNIL estime que le déploiement de ces dispositifs dans les espaces publics, où s'exercent de nombreuses libertés individuelles (liberté d'aller et venir, d'expression, de réunion, droit de manifester, liberté de culte, etc.) présente des risques pour les droits et libertés fondamentaux des personnes et la préservation de leur anonymat dans l'espace public.

Quoi qu'il en soit, la CNIL estime que la réglementation actuellement en vigueur n'est pas adaptée à cette nouvelle technologie et appelle à un cadre juridique spécifique concernant les caméras augmentées. Il est donc nécessaire de trouver juste équilibre entre le développement de ces technologies et la garantie des libertés et droits fondamentaux.

19 Le respect de la protection des données personnelles est-il compatible avec l'utilisation de l'intelligence artificielle ?

L'intelligence artificielle nécessite d'importantes quantités de données (*big data*) à des fins d'apprentissage. Comment exploiter celles-ci tout en maintenant l'impératif de respecter la législation européenne relative à la protection des données personnelles ?

Le règlement 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 du 27 avril 2016, plus connu sous le nom règlement général sur la protection des données (RGPD), est en vigueur depuis 2018. Il s'applique donc également au développement des outils d'IA et à leur utilisation du *big data*.

Bien que le RGPD ait établi des bases solides pour la protection des données personnelles, l'intelligence artificielle présente des défis uniques qui méritent une attention particulière. Le fonctionnement des algorithmes et les données qu'ils requièrent soulèvent en effet des problématiques bien spécifiques.

Concernant les données utilisées, l'IA soulève de nombreuses questions. Tout d'abord, l'exigence de qualité et de pertinence des données est fondamentale en matière d'IA. En effet, la capacité de l'algorithme à apprendre correctement sera influencée par le choix des données qui lui seront fournies lors de la phase d'apprentissage. Les données peuvent être traitées pour créer le modèle qui n'en contient pas ou à l'occasion de son déploiement. Dans tous les cas, il existe un risque potentiel de biaiser le système d'IA, s'il y a eu une mauvaise sélection des données lors de cette phase.

La question de la quantité des données nécessaire à cet apprentissage est également problématique. En effet, le principe de minimisation, consacré par le RGPD, implique de limiter la collecte de données personnelles à ce qui est strictement nécessaire au regard des finalités pour lesquelles elles sont traitées. Mais, il pourrait être utile de fournir à l'IA autant de données que possible afin de la rendre plus efficace et d'éviter les biais statistiques qui pourraient découler d'un mauvais choix dans l'échantillonnage de données.

Enfin, la protection prévue par le RGPD ne s'applique qu'aux données personnelles. La notion de données à caractère personnel est définie comme « *toute information se rapportant à une personne physique identifiée ou identifiable* », ce qui permet d'exclure de cette protection d'autres types de données ne revêtant pas un caractère personnel, c'est-à-dire ne permettant pas l'identification de la personne, telles que les données de santé agrégées au niveau d'une population déterminée. Cette exclusion n'est cependant pas satisfaisante, car des données non personnelles peuvent être traitées à l'occasion du déploiement d'un système d'intelligence artificielle affectant les droits des personnes. Ces traitements peuvent ainsi avoir des répercussions sociales et économiques importantes pour ces dernières.

À ce titre, le *Data Governance Act* et le *Data Act* s'inscrivent dans le cadre de la stratégie européenne pour les données visant à développer un marché unique de la donnée, présenté comme une alternative au modèle des GAFAM. Le *Data Governance Act,*

applicable depuis septembre 2023, a pour objectif de favoriser le partage des données personnelles et non personnelles en mettant en place des structures d'intermédiation. À travers ces textes, l'Union européenne poursuit en réalité des objectifs contradictoires. D'une part, elle souhaite mettre en place un cadre d'exploitation et de partage des données qui soit respectueux des valeurs et de la réglementation européennes. D'autre part, son but est de créer un écosystème permettant aux données de circuler le plus possible. Dans ce système, les données doivent pouvoir être exploitées en masse pour développer les technologies d'intelligence artificielle. L'articulation de ce nouveau cadre législatif sur les données avec le RGPD risque de poser des problèmes. Dans un tel cadre, il peut être difficile de faire respecter toutes les garanties que ce dernier prévoit. L'exploitation systématique à grande échelle de ces données par des algorithmes risque ainsi de fragiliser la réglementation protégeant les données à caractère personnel.

De plus, les caractéristiques particulières des algorithmes apprenants soulèvent aussi des questions inédites. En premier lieu, ces algorithmes sont en mesure de développer leurs propres critères de fonctionnement, et d'acquérir une autonomie croissante au fur et à mesure de leur apprentissage solitaire, ce qui pose la question de la transparence et de l'explicabilité de leur fonctionnement. Leur mode de fonctionnement peut en effet devenir difficilement explicable même pour leur propre concepteur. Aussi, l'explicabilité du mécanisme d'apprentissage constitue un enjeu majeur qui conditionne le respect de l'obligation globale de transparence du traitement des données prévue par le RGPD. Le principe de transparence exige en effet que toute information et communication relatives au traitement de ces données à caractère personnel soient aisément accessibles, faciles à comprendre, et formulées en des termes clairs et simples.

En second lieu, est aussi délicat le droit reconnu par le RGPD de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé produisant un effet juridique (par exemple, un algorithme qui décide si un crédit peut être accordé et dont la décision est automatiquement transmise à la personne concernée, sans intervention humaine aucune). Ce droit vise à garantir une intervention humaine à l'occasion des prises de décisions modifiant la situation d'une personne. Le traitement est exclusivement automatisé en l'absence d'intervention humaine dans le processus de prise de décision. Qu'en est-il en présence d'une intervention humaine qui s'appuie sur un outil d'aide à la décision ? Comment garantir que le pouvoir de recommandation et de prédiction des algorithmes apprenants reste cantonné à une aide à la prise de décision humaine, sans la remplacer ? Avec l'efficacité croissante de ces algorithmes, il sera de plus en plus difficile de ne pas suivre leurs préconisations, ce qui risque de vider de sa substance le droit à ne pas faire l'objet d'une décision individuelle entièrement automatisée.

Par conséquent, il est nécessaire de réaffirmer les règles prévues par le RGPD, car le progrès attendu du développement des algorithmes apprenants ne doit pas se faire au détriment des valeurs éthiques défendues par l'Union européenne. Toutefois, il sera peut-être nécessaire d'adapter certaines règles afin de tenir compte des problématiques spécifiques soulevées par l'intelligence artificielle.

20 De nombreux emplois sont-ils menacés par l'intelligence artificielle ?

Si certains experts pensent que l'impact positif de l'IA pourrait l'emporter sur ses aspects négatifs, la crainte que l'intelligence artificielle puisse à terme remplacer les travailleurs, en particulier ceux ayant des emplois peu qualifiés, persiste. Depuis l'arrivée de ChatGPT, l'inquiétude se fait sentir. L'intelligence artificielle générative, capable d'assimiler et de créer du contenu écrit, visuel ou audio, est en effet souvent considérée comme une menace pour les emplois.

Différentes études font état de la disparition de centaines de millions d'emplois dans un avenir proche à cause de l'intelligence artificielle, ce qui ne contribue pas à rassurer les travailleurs.

Selon un rapport du FMI sur l'impact de l'IA publié en janvier 2024, 60 % des emplois des économies avancées seraient menacés par l'intelligence artificielle, compte tenu de la forte proportion d'emplois dont les tâches peuvent être prises en charge par cette nouvelle technologie. Les emplois des économies émergentes seraient moins atteints, puisque seulement 4 emplois sur 10 seraient menacés et 26 % seulement dans les pays à bas revenus.

Afin de mesurer l'impact de l'IA sur le marché du travail, les auteurs du rapport ont utilisé deux critères : les tâches que l'IA est susceptible de prendre en charge et la complémentarité de l'emploi avec celle-ci. La combinaison de ces deux critères traduit le degré de protection dont bénéficie un emploi par rapport au processus de remplacement provoqué par cette nouvelle technologie. Les auteurs citent l'exemple des juges qui sont *a priori* menacés par l'IA dans la mesure où ils analysent de nombreux textes, mais qui bénéficient aussi d'une forte protection, car peu de sociétés sont disposées à voir les décisions de justice rendues par des systèmes d'IA.

D'autres études sont moins alarmantes et ne considèrent pas ces technologies comme présentant une réelle menace. Dans une étude de l'OIT (Organisation internationale du travail) de 2023 examinant les effets de la dernière vague d'IA générative, telle que ChatGPT, l'organisation estime que la plupart des emplois ne seront que partiellement exposés à l'automatisation. Le rapport suggère que l'impact le plus important de cette technologie ne sera probablement pas la destruction d'emplois, mais plutôt les changements potentiels en termes de qualité, d'intensité du travail et d'autonomie. Pour la plupart des professions, certaines tâches seront en effet réalisées par des bots (un logiciel qui exécute des tâches automatisées, répétitives et prédéfinies), mais cela laissera du temps pour d'autres activités plus complexes. En moyenne, 10 à 13 % des emplois dans le monde pourront être « augmentés » ou transformés.

Si le potentiel d'évolution du travail semble ainsi bien plus important que l'automatisation, ce risque reste pourtant bien réel. Il en va ainsi pour les emplois administratifs qui seraient largement menacés par l'automatisation. Pourraient ainsi être remplacés par des bots, les employés des centres d'appel, les secrétaires et les opérateurs de saisie. Avec près d'un quart des tâches considérées comme très exposées et plus de la moitié présentant un niveau d'exposition moyen, les employés de bureau représentent la

catégorie la plus exposée à l'automatisation. Ainsi, près de 2,3 % des emplois dans le monde, soit 75 millions d'emplois pourraient être concernés par cette automatisation. L'intelligence artificielle n'affectera donc pas tous les métiers de la même façon.

Il est en réalité difficile d'estimer l'impact réel de l'intelligence artificielle sur le marché de l'emploi. Il s'agit d'un exercice d'anticipation hautement spéculatif, car il y a de nombreux paramètres incertains, dont notamment l'évolution de la technologie en elle-même et son usage social ou industriel.

D'un côté, les pessimistes craignent une explosion massive du chômage ; de l'autre, les optimistes croient en l'avènement d'une société sans travail ou encore en la « destruction créatrice », théorisée par l'économiste autrichien Joseph Schumpeter dans son ouvrage de 1942 *Capitalisme, Socialisme et Démocratie*. L'intelligence artificielle serait un excellent exemple de destruction créatrice en cours.

La destruction créatrice serait un processus par lequel les innovations rendent obsolètes certaines activités économiques liées aux anciennes innovations et la création de nouvelles activités qui les supplantent. La concurrence, le progrès technique et l'entrepreneur innovateur seraient le moteur de ce modèle. Ainsi, l'économie se restructurerait en permanence, puisque certaines entreprises sont éliminées par le fait que d'autres, porteuses d'une nouvelle croissance, apparaissent.

Ce modèle ne prend cependant pas en compte certains facteurs. En effet, les nouvelles technologies remplaçant les anciennes créent généralement moins d'emplois, ce qui peut poser des problèmes sur le long terme. Si l'innovation provoque la destruction de plus d'emplois qu'elle n'en a créés, il ne peut plus y avoir assez de consommateurs pour acheter ces nouveaux produits, car étant au chômage leur pouvoir d'achat ne pourrait pas être suffisant pour les consommer. En l'état, cette innovation ne s'est pas encore traduite par des hausses de la productivité ni par des taux de croissance élevés. Il semble donc assez hasardeux de mettre en avant ce concept pour contrer l'argument réel du remplacement de certains emplois par l'IA.

III.
L'INTELLIGENCE ARTIFICIELLE DANS LE DOMAINE DE LA SANTÉ

21 L'intelligence artificielle et le *big data* vont-ils révolutionner la recherche et la pratique médicales ?

Dans les domaines de la recherche et de la pratique médicales, l'IA trouve d'ores et déjà de nombreux exemples d'applications concrètes mises au service du patient et des professionnels de santé : aide à la pose d'un diagnostic plus personnalisé, assistance dans les interventions chirurgicales, organisation des parcours de soins, etc.

Dans de nombreux domaines, la recherche biomédicale bénéficie des progrès technologiques de la collecte, du traitement et de l'analyse de vastes ensembles de données extrêmement volumineux, complexes et variés – on parle de données massives ou *big data* – relatives aux maladies étudiées. L'importance et l'étendue de ces données dépassent les capacités des outils traditionnels de gestion et d'analyse des données. Le potentiel des usages de ces données massives apparaît considérable, faisant de la santé l'un des secteurs le plus souvent cités par les études et rapports consacrés au *big data*.

Ces informations médicales proviennent de sources variées : assurance maladie, hôpitaux, dossiers médicaux, les résultats d'analyses de laboratoire, les dispositifs de surveillance des patients, les essais cliniques, les données de santé publique, les données d'objets connectés (tensiomètre, implant cardiaque, montre connectée), etc.

L'utilisation de l'IA dans l'analyse de données médicales ouvre de nouvelles perspectives encourageantes pour l'amélioration des diagnostics médicaux. Grâce à sa capacité à traiter de grandes quantités de données de santé, l'IA peut extraire des informations précieuses, détecter des modèles, des tendances, des corrélations et des anomalies qui échapperaient habituellement à l'observation humaine.

Du fait de leur complémentarité, les *big data* et l'intelligence artificielle transforment ainsi la pratique médicale. En effet, il est possible de bénéficier de diagnostics plus rapides et de réduire les erreurs de diagnostic en intégrant l'IA dans les pratiques médicales.

Par exemple, dans le domaine de l'imagerie médicale, l'IA peut aider à détecter des anomalies comme des tumeurs ou des lésions, en fournissant une assistance aux radiologues dans leur interprétation. Ces techniques sont en passe de révolutionner le diagnostic des maladies et l'interprétation d'images médicales (radios, scanners, IRM, etc.). Elle peut aussi fournir des recommandations aux médecins en analysant les données du patient (antécédents médicaux, résultats de tests), ce qui permet de proposer des traitements personnalisés et des plans de soins adaptés à chaque individu. Elle est également utilisée tant pour accélérer la recherche sur les médicaments et optimiser leur processus de production que pour analyser des données de composés chimiques, simuler leur interaction avec des cibles biologiques et prédire leur efficacité potentielle.

Par ailleurs, sont désormais rendus possibles de nouveaux types de recherches à très grandes échelles conduits grâce à la combinaison des données recueillies sur des millions de patients. Par exemple, les masses de données génomiques numériques recueillies à

grande échelle résultant des séquençages, de plus en plus fréquents, constituent en effet de nouvelles données anonymisées exploitables par le biais de l'intelligence artificielle au profit de recherches en oncologie

Ce faisant, la médecine deviendra de plus en plus personnalisée, voire plus précise. On parle d'ailleurs d'une « médecine de précision ». L'IA permet ainsi d'envisager l'exploitation de l'ensemble des paramètres du patient avec des algorithmes décisionnels afin de personnaliser au mieux les soins bénéficiant au patient. Le XVIII^{e} siècle a vu l'émergence de la clinique (symptômes, et syndromes), de l'anatomie, le XX^{e} siècle a connu celui de l'émergence de la biologie moléculaire. La médecine du XXI^{e} siècle sera une médecine plus préventive que curative. Grâce à la science, au *big data* et à la technologie, elle changera radicalement et évoluera vers une médecine des 4P : préventive, prédictive, personnalisée et participative. Dans la médecine 4P, la consultation médicale risque d'être moins motivée par les symptômes que par la probabilité d'être atteint par une maladie. Les avancées réalisées dans la recherche en génomique et en biologie moléculaire, ainsi que l'évolution technologique, ouvrent par conséquent de nouvelles perspectives à cette médecine de précision.

Même si l'utilisation du *big data* offre de nombreux bénéfices, plusieurs interrogations subsistent notamment quant à la collecte et au traitement des données de santé. Ces données peuvent être issues de dossiers médicaux de patients, de résultats de tests et d'images médicales confidentielles.

L'utilisation des *big data* et de l'IA dans le domaine médical peut poser des défis, notamment concernant la protection de la vie privée et la sécurité des données. Ils constituent des préoccupations majeures, car il s'agit de données sensibles qui doivent être traitées avec précaution. Aussi, des mesures rigoureuses doivent être mises en place pour assurer la confidentialité et la sécurité de ces données.

22 Quelle est la place de l'intelligence artificielle dans la relation médecin-patient ?

Le métier de médecin est un métier en pleine mutation. La médecine a déjà connu des transformations liées aux évolutions technologiques. Jusque dans les années 1960-1970, les médecins disposaient de peu d'éléments de diagnostic, puis avec l'arrivée de nouvelles technologies s'est développée une médecine centrée sur les examens complémentaires, l'imagerie, la biologie, etc.

On prédit régulièrement aux médecins l'avènement de technologies capables de les seconder, voire de les remplacer. Avec cette nouvelle évolution technologique, les métiers de la santé ne vont pas disparaître, mais seront profondément transformés et nécessiteront une formation à de nouveaux outils numériques, de nouveaux processus et de nouvelles pratiques.

L'intelligence artificielle va-t-elle bouleverser aussi la relation entre médecin et patient ? Elle apparaît en effet comme un nouvel acteur de la relation de soin. Le médecin ne doit pourtant pas s'effacer devant la machine, il doit garder toute sa place dans la relation avec son patient. La médecine comportera donc toujours une part essentielle de relations humaines et ne pourra jamais accorder une confiance aveugle à des décisions prises par des algorithmes, fussent-ils performants et fiables, mais dénués de nuances et d'empathie. Le cancérologue recevra, par exemple, une proposition de protocole de soins, mais il aura la possibilité de refuser ou modifier les suggestions de l'intelligence artificielle en fonction des éléments tirés de son écoute du patient.

L'utilisation d'algorithmes, comme maillon de la proposition de diagnostic, a vocation à se généraliser, d'où la nécessité d'un principe fondamental de garantie humaine du numérique, c'est-à-dire la garantie d'une supervision humaine de toute utilisation du numérique en santé, et l'obligation d'instaurer pour toute personne le souhaitant, et à tout moment, la possibilité d'un contact humain en mesure de lui transmettre l'ensemble des informations la concernant dans le cadre de son parcours de soins (Avis 129 du Comité consultatif national d'éthique). La révision de la loi relative à la bioéthique par la loi du 2 août 2021 a été l'occasion pour le législateur d'introduire ce principe dans le Code de la santé publique (art. L. 4001-3).

Ainsi, le professionnel de santé qui décide d'utiliser, pour un acte de prévention, de diagnostic ou de soin, un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives doit s'assurer que la personne concernée en a été informée et qu'elle est, le cas échéant, avertie de l'interprétation qui en résulte. Autrement dit, la personne doit être informée de l'utilisation dudit dispositif, et c'est au professionnel de santé qu'il incombe de lui expliquer clairement le résultat de sortie.

Le professionnel de santé a donc désormais la liberté de se saisir de ces outils avec lesquels il peut exercer de manière optimale. Pour les professionnels de santé, l'intelligence artificielle porte la promesse d'une automatisation qui les délestera de certaines tâches répétitives leur permettant ainsi de libérer du temps pour les soins de santé et de se recentrer sur l'écoute et l'accompagnement de leurs patients. Il y aura

donc toujours une complémentarité entre le praticien et l'outil. Le patient habitué au « colloque singulier » (la relation bilatérale et protégée, en confiance, du médecin et de son patient) avec son médecin sera aussi en mesure d'adhérer plus facilement à la proposition de soins qu'il lui propose.

Toutefois, le patient peut aussi redouter voire refuser les dispositifs d'intelligence artificielle, surtout lorsqu'il doit lui-même les utiliser dans le cadre de son traitement. Mais, s'il s'empare de ces nouveaux outils, il peut devenir un acteur à part entière contribuant au diagnostic et au traitement. Le développement de l'intelligence artificielle et des objets connectés participe ainsi à une forme de la responsabilisation des patients.

Ce faisant, l'intelligence artificielle laisse entrevoir une variété de possibilités et d'avantages dans l'exercice de la médecine, mais également des risques.

Parmi ces risques, nous pouvons citer une inégalité dans l'accès à des soins de santé de qualité. Les patients traités dans les régions où les établissements de santé font partie des pionniers de l'intelligence artificielle bénéficieront de ces systèmes avant les autres, ce qui risque de provoquer des déséquilibres géographiques dans la performance des systèmes de soins de santé et des inégalités dans l'accès aux soins.

C'est en réalité le choix du modèle de service qui détermine dans quelle mesure un système d'IA bouleverse l'exercice de la médecine. Si l'IA complète seulement l'expertise des professionnels de la santé, ses effets sur la qualité humaine des entretiens cliniques se révéleront minimes. À l'inverse, si elle est utilisée pour remplacer l'expertise humaine, son effet sur la relation de soins est plus incertain. Il est toutefois peu probable que nous assistions dans les années à venir à un remplacement total de l'expertise humaine par l'intelligence artificielle. Si l'on ne prend pas en compte les spécificités de la pratique médicale, la relation médecin-patient pourrait pourtant se dégrader. Cette relation demeure en effet la pierre angulaire de la « bonne » pratique médicale, mais elle semble évoluer vers une relation médecin-patient-IA à définir.

23 Quels sont les dangers de l'utilisation de l'intelligence artificielle et le *big data* dans le domaine de la santé ?

L'utilisation de l'intelligence artificielle à des fins médicales n'est pas sans risque. Parmi ces risques, l'on peut citer le manque de fiabilité des résultats donnés par des algorithmes biaisés ou l'utilisation de données de mauvaise qualité, la divulgation de données à caractère personnel ou encore l'absence d'attribution précise des responsabilités quant à la gestion de l'IA et de ses éventuels effets préjudiciables.

C'est d'ailleurs ce que confirme l'Organisation mondiale de la santé (OMS) dans un document publié le 18 janvier 2024. Elle vise notamment un type d'IA générative à croissance rapide tel que ChatGPT : les grands modèles multimodaux ou *large multi-modal models* (LMM). Ces LMM peuvent utiliser plusieurs types de données, y compris du texte, des images et des vidéos, et générer des résultats qui ne sont pas limités au type de données introduites dans l'algorithme.

À ce titre, l'OMS énumère cinq domaines qui pourraient utiliser cette technologie : le diagnostic et soins cliniques (répondre aux questions des patients) ; la formation médicale ; les tâches administratives (documentation, dossiers de santé, etc.) ; l'usage par les patients, par exemple pour l'examen des symptômes ; la recherche scientifique et le développement de médicaments.

Bien que cette technologie présente un grand potentiel, l'OMS précise que l'adoption précipitée de systèmes, qui n'ont pas été testés, pourrait entraîner des erreurs de la part des personnels de santé et nuire aux patients. Il s'agit principalement d'éviter que ces résultats soient générés à partir de données fausses, inexactes, biaisées ou incomplètes (origine ethnique, ascendance, sexe, identité, âge, etc.) qui pourraient induire en erreur les personnes utilisant ces informations pour prendre des décisions en matière de santé.

La question du degré de confiance qu'un médecin peut avoir dans un système d'IA l'aidant à prendre des décisions médicales est effet une question centrale. En principe, l'IA se sert de données de santé existantes pour fournir une prédiction, qui peut être par exemple un diagnostic, un pronostic ou encore un traitement à suivre. Pour ce faire, l'IA est entraînée sur les données d'anciens patients, avant d'aider le médecin au diagnostic ou à l'établissement du pronostic de nouveaux patients. Or, les caractéristiques de ces patients peuvent évoluer : les IA peuvent avoir été entraînées sur des patients différents de ceux sur lesquels elles vont être utilisées, ce qui risque de provoquer une diminution de la qualité des prédictions, voire des erreurs.

Par ailleurs, des surdiagnostics peuvent aussi résulter de l'acceptation d'un seul résultat de test anormal comme étant un diagnostic, et ainsi attribuer un poids pronostic plus élevé à certaines mesures que ce qui est validé par la recherche, ce qui peut conduire à des surdiagnostics et à des investigations excessives et inutiles.

Inversement, l'IA peut sous-diagnostiquer des groupes de personnes en excluant les patients dont les données sont absentes en raison d'un accès limité aux soins. Si des sous-groupes de la population ne consultent pas souvent le médecin et sont sous-

diagnostiqués, les caractéristiques de leur groupe (âge, ethnie, genre) seront interprétées à tort comme conférant un risque plus faible de la maladie dont on assure le suivi.

Dans leur avis commun « *Diagnostic médical et intelligence artificielle : Enjeux éthiques* », publié en janvier 2023, le Comité consultatif national d'éthique (CCNE) et le Comité national pilote d'éthique du numérique (CNPEN) ont examiné les problèmes éthiques posés par les systèmes d'intelligence artificielle appliqués au diagnostic médical (SIADM) qui sont des dispositifs pouvant, à partir d'objectifs définis par l'homme, générer des résultats, tels que des contenus, des prédictions, ou des recommandations.

Ces SIADM peuvent non seulement être utilisés en amont de la consultation médicale (dans les services d'urgence afin d'orienter la prise en charge), mais aussi pendant les étapes d'élaboration du diagnostic (imagerie) ou encore lors du suivi médical à domicile.

Les auteurs de l'avis notent qu'ils sont déjà à l'origine de nombreux progrès (identification de lésions qui échappent à l'œil humain, etc.), mais ils insistent également sur le fait qu'ils produisent des résultats basés sur des approches qui peuvent être probabilistes et aussi entachés d'erreurs.

Dans leurs recommandations, ils insistent sur la nécessité de ne pas les utiliser dans une logique de substitution à l'intervention humaine des professionnels de santé. Par ailleurs, ils préconisent de les soumettre à un contrôle humain à toutes les étapes-clés de la conception et de l'application en vie réelle du dispositif afin de garantir la sécurité et le respect des droits fondamentaux. Ils recommandent également de faciliter l'explicabilité des résultats obtenus afin que les médecins soient en mesure de leur donner un sens clinique, d'informer le patient en cas de recours à un SIADM et enfin d'encourager la recherche. L'avis souligne enfin que les SIADM doivent toujours être utilisés en priorité dans une optique d'amélioration démontrée du soin, avant leurs autres intérêts organisationnels, économiques ou managériaux.

24 Quels sont les enjeux éthiques liés à l'utilisation de l'intelligence artificielle dans le domaine médical ?

L'adoption des technologies de l'intelligence artificielle par le milieu médical s'accompagne de sérieux enjeux éthiques. Plusieurs des enjeux éthiques liés au déploiement de l'IA dans le domaine de la santé trouvent leur source dans la façon dont les données sur les individus sont collectées, utilisées, partagées, etc.

Ces données étant nécessaires pour entraîner les algorithmes utilisés par les différents systèmes d'IA, il n'est donc pas étonnant qu'elles soient devenues des ressources convoitées par différents acteurs publics et privés. Les systèmes d'IA peuvent en effet exploiter des données biomédicales ou cliniques contenues dans le dossier médical, mais aussi des données produites par les individus dans le cadre de leur vie quotidienne, telles que les données d'entraînement physique, de sommeil ou de nutrition qui sont issues dans la plupart des cas d'objets connectés. Montres ou appareils d'entraînement physiques connectés collectent en effet des données sur les individus en temps réel et les convertissent en données médicales qui seront analysées pour en déduire des informations sur la santé globale des individus. Les sources variées de ces données conduisent à une diversification des acteurs engagés dans leur collecte qui peuvent inclure des acteurs publics comme privés.

Or, le consentement à la collecte et au partage des données personnelles par les applications de santé et les objets connectés est souvent obtenu de façon trop rapide, sans véritable mesure de tous les enjeux, ce qui pose la question du consentement libre et éclairé dans ces cas de figure.

Ce défaut de transparence dans la collecte des données a pour conséquence d'affecter la protection de la vie privée des personnes concernées. L'IA est en effet capable de croiser, d'analyser ou de profiler ces données, ce qui peut porter atteinte à la sphère personnelle en révélant des aspects intimes des individus, souvent à leur insu. L'abondance de données collectées de toute part et à des fins variées, ainsi que leur partage peuvent compromettre l'exercice du contrôle des personnes sur leurs données. Il y a donc un risque d'atteinte au droit au respect de la vie privée et à la protection des données personnelles.

La confidentialité de ces données personnelles sensibles peut être néanmoins obtenue grâce à des techniques empêchant l'identification des personnes, telles que la pseudonymisation ou l'anonymisation. Ces techniques ont cependant des limites puisqu'il existe toujours un risque important de réidentification des informations de la personne par l'intermédiaire d'autres données. Par exemple concernant l'anonymisation, technique qui suppose le retrait des données identificatoires des bases de données de façon irréversible, il a été démontré qu'il suffit de quatre données anonymisées, provenant de différentes banques de données, pour arriver à réidentifier un individu. Les caractéristiques de la donnée de santé qui, anonyme un jour, peut conduire à une réidentification des personnes à mesure des réutilisations, imposent qu'elle soit traitée avec beaucoup de précaution.

Il existe enfin un risque pour la sécurité de ces données. Les systèmes d'IA ne sont pas à l'abri des fuites de données et des cyberattaques qui peuvent compromettre leur fonctionnement, leur performance ainsi que la confidentialité des données sur lesquelles ils sont entraînés.

Un autre enjeu éthique concerne la qualité des données de santé. Si les systèmes d'IA ne sont pas intrinsèquement biaisés, ils sont susceptibles d'utiliser des données qui, elles, peuvent être biaisées. La pertinence des résultats d'un algorithme dépend en effet en grande partie des informations qui lui sont fournies pour son apprentissage. Aussi, il est fondamental qu'il soit entraîné sur des ensembles de données qui présentent des garanties en termes de qualité, fiabilité et de représentativité afin d'éviter les biais. Des erreurs dans les données de base risquent d'entacher la fiabilité des conclusions algorithmiques.

Par exemple, si les données utilisées pour entraîner l'algorithme ont été majoritairement collectées auprès de personnes de sexe masculin, il est possible qu'il tire des conclusions inexactes lorsqu'il rencontre des données concernant des personnes de sexe féminin.

De plus, la précision et la fiabilité des données ne sont pas forcément homogènes selon leur source. Ces différences peuvent s'expliquer par le soin porté à la gestion du cycle de vie des données, la quantité des données contenue dans la banque d'entraînement ou encore leur représentativité.

Un manque de constance dans l'application des différents critères permettant d'assurer la qualité, la fiabilité et la représentativité des données utilisées peut affecter la qualité des prédictions et les recommandations des algorithmes. Afin d'améliorer la qualité des soins grâce à l'intelligence artificielle, il est ainsi nécessaire d'apporter une attention particulière à la nature des données de santé mais aussi à leur qualité.

25 Les données de santé sont-elles suffisamment protégées ?

La quantité de données de santé issue de la prise en charge des patients ne cesse d'augmenter, de même que le nombre de sources de données disponibles. La gestion de ces données massives est fondamentale pour une meilleure compréhension des maladies, du développement de médicaments et du traitement des patients. Ces données personnelles sensibles, relevant de la sphère privée, sont aujourd'hui convoitées par des acteurs variés.

La notion de donnée de santé a été définie par le RGPD comme étant une donnée à caractère personnel relative à la santé physique ou mentale d'une personne physique « *qui révèle des informations sur l'état de santé de cette personne* », « *y compris la prestation de services de soins de santé* ». Il s'agit d'une définition large puisque la donnée de santé ne concerne pas forcément une pathologie et il n'est pas nécessaire qu'elle ait été collectée par un professionnel de santé, ce qui signifie qu'elle peut l'être, par exemple, par un objet connecté. Étant des données sensibles, le RGPD en interdit en principe le traitement, en l'assortissant de multiples exceptions : lorsque la personne y a consenti ; lorsque le traitement est nécessaire pour la sauvegarde des intérêts vitaux de la personne concernée ou d'un tiers ; pour des motifs d'intérêt public comme la sûreté de l'État, la sécurité publique ou la santé publique.

Fondamentales à l'intelligence artificielle, les données de santé intéressent de plus en plus différents acteurs allant de la recherche publique jusqu'aux GAFAM, qui développent des systèmes d'IA performants, en passant par les laboratoires pharmaceutiques.

C'est dans les années 1980 que les pouvoirs publics ont commencé à s'intéresser à ces données avec la création du Programme de médicalisation des systèmes d'information (PMSI) permettant la collecte des informations relatives à la fréquentation des établissements de santé et des actes qui y sont réalisés, notamment pour amélioration de leur gestion financière.

Les bases de données se sont ensuite multipliées. La France possède un grand nombre de bases de données relatives au domaine de la santé. Étant gérées par différents acteurs (services hospitaliers, organismes de recherche, universités, Assurance Maladie, etc.), il en résultait parfois une absence de règles communes d'accès aux données et, dès lors, une difficulté à les exploiter.

Le législateur a souhaité ouvrir l'accès aux données de santé collectées par les personnes publiques afin de tirer profit des potentialités qu'elles offrent.

Créé par loi du 26 janvier 2016 de modernisation du système de santé, le SNDS (Système National des Données de Santé) est un entrepôt de données médico-administratives anonymisées couvrant l'ensemble de la population française. Il rassemble les bases de données de santé déjà existantes. Il s'agit de l'une des bases de données de santé les plus importantes au monde.

L'intérêt de cette base est de permettre l'appariement des bases de données et le croisement des données pour améliorer les politiques de santé, l'offre de soins, la protection sociale, la prise en charge médico-sociale et la recherche.

Afin de centraliser et faciliter le partage des données du SNDS, la Plateforme des données de santé (PDS) ou *Health Data Hub* (HDH) a été créée en juillet 2019. Sa mise en place a été accélérée pour contribuer à la gestion de la crise sanitaire. Dans l'urgence, l'entreprise *Microsoft Azure* a été retenue en tant qu'hébergeur. Or, ce choix de prendre comme prestataire la société *Microsoft* a donné lieu, à juste titre, à d'importantes critiques tenant principalement au fait qu'en tant qu'entreprise américaine, elle pouvait être obligée de transférer les données hébergées aux autorités américaines en vertu des lois américaines (*Cloud Act*), ce qui va à l'encontre des dispositions du RGPD, et plus précisément, à la protection des données de santé des ressortissants de l'Union européenne. La migration vers un nouvel hébergeur européen ne devrait pas avoir lieu avant 2025. Le choix de *Microsoft Azure* a pourtant été renouvelé pour le nouvel entrepôt de données de santé porté par l'Agence européenne du médicament, EMC2, faute de prestataire susceptible de répondre actuellement aux besoins exprimés.

Par ailleurs, un système centralisant une masse aussi conséquente de données n'est pas inviolable ; il existe des risques évidents de piratage, même s'ils sont relativement dilués du fait de la conservation des données par chaque acteur qui les a récoltées. La multiplication des cyberattaques sur les établissements de santé n'est en effet pas une question à négliger. Même si ces attaques n'ont pas eu d'incidences directes sur la santé des patients, elles constituent tout de même des atteintes à leurs droits et libertés. C'est la raison pour laquelle des moyens techniques doivent être mis en œuvre par les établissements de santé afin de tenir compte du risque numérique, même s'il est vrai qu'il n'existe pas de solution miracle pour contrer ces attaques.

26 Qu’est-ce que la médecine prédictive ?

La prévention, ou prophylaxie, est une conception anticipative de la médecine, illustrée par le célèbre adage « mieux vaut prévenir que guérir ». Faire de la prévention a toujours existé en médecine. Si cela reste évidemment d’actualité, l’avenir est semble-t-il à la médecine prédictive ayant l’ambition de prédire l’apparition des maladies grâce à la génétique, qui permet d’évaluer les risques dans ce domaine.

Le rêve, que fait miroiter la médecine prédictive, est de déterminer les prédispositions à des maladies avant même l’apparition des premiers symptômes, grâce à un test génétique dans le but de faciliter leur prise en charge, voire d’éviter leur survenue. Dès lors, la médecine prédictive ne s’adresse pas à des malades mais à des sujets sains.

À l’aide de tests génétiques, on espère prévoir et prévenir l’apparition de certaines maladies, sachant qu’une prédiction ne relève pas de la divination, mais de l’évaluation statistique d’un risque. Il s’agit de tests bien précis, ciblant une mutation génétique spécifique chez des patients dont, généralement, l’un des parents est porteur. Compte tenu de la nature du patrimoine héréditaire de chaque individu et de celle de son environnement, il s’agit de l’amener à conserver une bonne santé jusqu’à l’âge le plus avancé de sa vie. L’exemple classique est celui des femmes porteuses de mutations sur les gènes BRCA1 et BRCA2 qui ont un risque plus élevé de développer un cancer du sein et de l’ovaire que les femmes non porteuses.

La médecine est depuis longtemps capable d’anticiper les dangers, d’éviter les infections et de mettre en évidence des facteurs de risque. Mais, c’est en grande partie grâce aux progrès accélérés qu’ont connu ces dernières années à la fois la génomique et le *big data* que l’on peut aujourd’hui parler de médecine prédictive. En analysant le génome, il est en effet désormais possible d’étudier la prédisposition génétique des personnes testées à certains types de maladies graves telles que les cancers. Autre évolution majeure : l’émergence du *big data* et, plus généralement, la multiplication exponentielle d’informations génomiques, cliniques et issues de l’expérience en vie réelle. L’analyse de toutes ces informations, de nature extrêmement différente, permet de créer des modèles de prédiction de plus en plus efficaces et pertinents.

Ce faisant, la médecine prédictive relève du champ de la médecine personnalisée qui propose de traiter chaque patient de façon individualisée en fonction de ses spécificités génétiques et environnementales. Il en ressort une médecine que l’on peut qualifier des 4P voire 5P : préventive, prédictive, participative, personnalisée et pertinente. Elle s’appuie sur diverses sources de données pour faire des prédictions sur le risque qu’a un individu de développer une maladie particulière. La médecine prédictive n’est au demeurant pas que génomique. Il est d’ailleurs assez réducteur de la limiter à sa seule dimension génétique, même si elle est primordiale. Ces données peuvent provenir de dossiers médicaux, d’antécédents familiaux, de facteurs liés au mode de vie, d’expositions environnementales, d’analyses biologiques et/ou génétiques.

De manière générale, l’IA devrait contribuer à l’essor de cette médecine personnalisée, fondée sur une analyse des caractéristiques biologiques et génétiques de la personne et de son environnement spécifique. Elle facilitera ainsi le dépistage précoce de maladies

ainsi que l'identification des facteurs de risques, ce qui devrait améliorer la prise en charge des patients.

Si on utilise aujourd'hui des algorithmes pour le diagnostic (par exemple, en imagerie médicale), on devrait aussi pouvoir les utiliser pour prédire la survenue d'une maladie ou de ses complications. On parle peut-être alors davantage d'une médecine algorithmique.

La médecine prédictive semble être un concept prometteur, mais pas encore vraiment une réalité. Il est cependant possible d'anticiper les conséquences sanitaires et sociétales de cette tendance qui consiste à décrypter ce qui fait la singularité génétique de chaque personne.

En 1997, le film *Bienvenue à Gattaca* décrivait une société où les choix de vie sont orientés, voire conditionnés, par les facteurs de risque génétiques déterminés à la naissance. Même si la réalité n'a pas rejoint la fiction, il convient d'ores et déjà de tenir compte de ces risques d'immixtion dans la vie privée et la destinée des personnes.

Il n'est, par exemple, pas exclu que la médecine prédictive soit un jour utilisée pour prévoir la prédisposition à des performances physiques ou intellectuelles, ou qu'elle serve à mettre en évidence des caractères liés à des comportements violents ou criminels. Les dérives eugéniques sont donc à prendre en considération.

Elle pourrait également être exploitée dans le domaine des assurances et du travail. Il existe en effet un réel risque de discrimination par la sélection de la population sur des critères génétiques dans le domaine de l'assurance, comme en témoigne le cas des États-Unis où l'utilisation des tests génétiques est déjà en place. De même, si les tests génétiques étaient introduits dans le domaine du travail, les employeurs auraient la possibilité de repérer les individus porteurs de gènes associés à des maladies et d'ainsi les exclure.

Par conséquent, la médecine prédictive est une arme à double tranchant. Comment préserver l'intimité génétique des individus face aux assureurs, employeurs ou banquiers susceptibles d'en faire des motifs de discrimination ? Il est donc fondamental d'évaluer les risques de dérives biologiques et sociales avant la généralisation de ces dépistages génétiques.

27 Le robot sera-t-il le médecin du futur ?

Au cours du siècle précédent, le développement de la robotique a pris un essor important avec l'avènement de l'électronique et de l'informatique. Les premiers robots dans le domaine médical sont apparus dans les années 1980 pour fournir une assistance chirurgicale grâce aux technologies de bras robotisés. Il s'agissait de simples robots industriels qui ont été sécurisés pour pouvoir être mis en œuvre en salle d'opération et sur lesquels ont été adaptés des instruments conventionnels de chirurgie. C'est d'ailleurs dans les secteurs de la neurochirurgie et de l'orthopédie que ces opérations furent menées.

En 1983, *Arthrobot*, le premier robot utilisé lors d'une opération chirurgicale, se contentait de répondre aux commandes vocales et de passer les instruments chirurgicaux. En 1985, le *Puma* 260 de la société américaine Unimation a été utilisé en neurochirurgie sur une vingtaine de patients en Californie. Ce dernier sera entre autres utilisé par la NASA. Le *Scara*, fruit d'une collaboration entre IBM et l'Université de Californie a été conçu entre 1986 et 1989 pour la chirurgie orthopédique. C'est à partir de ce dispositif qu'a été développé *Robodoc*, commercialisé depuis 1992 et ayant effectué plusieurs milliers de poses de prothèse de la hanche.

Mais c'est véritablement au début des années 1990 que les premiers prototypes de robots spécifiquement conçus pour la chirurgie ont été développés. Ceux-ci présentent des architectures mécaniques adaptées au geste chirurgical à réaliser. À la différence du robot industriel, les robots médicaux sont en perpétuelle interaction avec le chirurgien (par des commandes vocales du chirurgien ou par l'intermédiaire d'un ordinateur). L'opération Lindbergh, réalisée en 2001, par le Professeur Marescaux, qui opéra depuis New York une patiente se trouvant à Strasbourg, marque une avancée et une maîtrise de la collaboration entre les télécommunications et la chirurgie robotique.

En 1998, la première version dite « standard » du désormais célèbre robot *da Vinci* d'Intuitive Surgical est utilisée lors d'un pontage coronarien, et ce deux ans avant sa validation par la FDA américaine. Depuis lors, il a été utilisé dans de nombreux types d'opérations, notamment de chirurgies cardiaque, générale, urologique, gynécologique et thoracique.

Le système chirurgical *da Vinci* permet aux médecins de procéder à une chirurgie mini-invasive, ce qui veut dire qu'elle réduit le traumatisme causé à l'organisme lors de l'intervention, notamment en diminuant la taille des incisions. Il se compose de trois éléments : une console pour le chirurgien ; un chariot d'imagerie ; et un chariot patient mobile équipé de quatre bras robotisés.

Aucune spécialité de la médecine ne semble échapper à cet engouement pour les robots chirurgicaux : chirurgies de l'œil, de la colonne vertébrale, du système digestif, des reins et de l'uretère, etc.

La convergence entre la robotique et l'intelligence artificielle offre aussi de nouvelles perspectives en matière chirurgicale. Grâce à l'IA, il est désormais possible de créer des avatars numériques des patients, ce qui permet au chirurgien d'explorer diverses options

et de simuler les résultats possibles, réduisant ainsi considérablement les risques de complications pendant l'opération réelle.

L'IA peut offrir aux chirurgiens une vision précise et en temps réel de la zone opératoire grâce à l'analyse des images médicales. Ces informations permettent la conception d'une reproduction virtuelle de la zone à opérer, qui facilite la planification et la préparation du geste chirurgical. Elle peut enfin permettre de prédire les suites opératoires grâce à l'intégration de toutes les données liées à l'intervention chirurgicale pour anticiper d'éventuelles complications.

Des chercheurs préparent déjà la robotique de demain qui devrait être plus polyvalente, moins encombrante et miniaturisée. L'utilisation des robots en chirurgie n'est pourtant pas sans inconvénient.

Il y a tout d'abord leurs coûts élevés. L'installation d'un robot chirurgical suppose un investissement conséquent, c'est la raison pour laquelle tous les grands hôpitaux ne peuvent pas en être équipés. Par exemple, un robot *da Vinci* coûterait entre 1 et 2 millions d'euros selon les modèles, auxquels s'ajoutent les frais annuels de maintenance et de formation à son utilisation, ce qui représente un investissement important à la charge des établissements hospitaliers.

Il y a ensuite la taille et l'encombrement ; les robots chirurgicaux étant généralement lourds et volumineux. Leur installation implique donc un réaménagement – coûteux – du bloc opératoire et la présence d'un lieu de stockage adapté. La taille et le poids du robot chirurgical peuvent d'ailleurs limiter ses capacités en termes de mouvement et de manipulation, ce qui peut le rendre inefficace pour effectuer certaines tâches médicales spécifiques. De plus, il convient de prendre en compte la proximité physique du robot et du sujet humain et donc de sa dangerosité potentielle.

L'utilisation de systèmes robotiques dans le domaine médical pose évidemment le problème de la sécurité dans un environnement où l'homme est présent. Le premier risque qui vient d'emblée à l'esprit est celui des dommages causés au patient en cours d'utilisation. Ce risque peut aussi avoir pour origine des erreurs dans l'établissement des diagnostics ou des traitements qui trouveraient leur source dans les défaillances des logiciels.

La chirurgie robot-assistée est aussi confrontée à un problème au niveau de sa formation. Si la formation du chirurgien est trop courte ou insuffisante pour obtenir les compétences nécessaires à la bonne utilisation du robot chirurgical, il pourrait mettre en danger le patient lors d'une intervention chirurgicale. Par ailleurs, l'interaction très forte avec l'humain ainsi que l'hétérogénéité des disciplines intervenant dans la conception du robot (informatique, électronique, mécanique, etc.) exigent de prendre en compte les spécificités des problèmes relatifs à la sécurité des systèmes de la robotique médicale. Le robot ne pouvant pas improviser en cas d'urgence, un humain doit superviser l'opération et se tenir prêt à intervenir.

Enfin, se pose la question de l'efficacité et de la pertinence de l'utilisation de ces robots pendant les opérations chirurgicales. Dans une étude publiée dans la revue *Annals of Internal Medicine* en 2021, des chercheurs de l'Université du Texas ont démontré que l'utilisation de robots pendant les opérations chirurgicales ne présente pas d'avantages

évidents. Outre le coût élevé, les chercheurs ont indiqué que les opérations chirurgicales assistées par robot prenaient généralement plus de temps, et qu'il n'y avait de différences évidentes dans les résultats par rapport aux opérations classiques. Les auteurs n'excluent toutefois pas leur apport positif dans le futur grâce au progrès technique et à la réduction des coûts.

28 Qui est responsable des dommages causés par un robot ?

La réponse paraît simple : le robot ne pouvant être responsable de ses actes, le responsable sera le gardien du robot, propriétaire ou utilisateur. Être responsable d'un point de vue juridique signifie l'obligation de répondre des dommages causés. Cette réponse évidence fait pourtant l'objet d'un débat. Pour certains, l'intelligence artificielle du robot justifierait de lui attribuer une personnalité juridique afin qu'il supporte la responsabilité de ses actions dommageables. Pour d'autres, l'intelligence artificielle ne justifie pas l'octroi de la personnalité juridique, car le robot reste un bien dont le gardien est le seul responsable.

Dans sa résolution du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique, le Parlement européen a proposé la création, à terme, d'une personnalité juridique spécifique aux robots, pour qu'au moins les robots autonomes les plus sophistiqués puissent être considérés comme des personnes électroniques responsables, tenues de réparer tout dommage causé à un tiers.

L'octroi de la personnalité juridique permet à la personne qui en bénéficie d'être titulaire de droits et d'être tenue par des obligations. En principe, comme toute personne dispose d'un patrimoine, il faudrait alors abonder celui de la personne robot pour l'exécution de sa dette de réparation en cas de dommage. L'une des critiques émises contre cette proposition est qu'en responsabilisant les robots, on déresponsabiliserait les fabricants et les concepteurs de programmes, alors que la sécurisation des objets intelligents et la réduction des risques de dommages font en principe partie de leurs obligations.

Cette proposition s'est très rapidement heurtée à une ferme opposition. Le Comité économique et social européen (CESE) s'y est déclaré hostile, dans un avis du 31 mai 2017, dans lequel il préfère une maîtrise humaine ; les machines restant des machines que les hommes ne cessent jamais de contrôler. Le CESE indique notamment qu'il existe des risques que cette reconnaissance mette à mal les principes du droit de la responsabilité civile.

L'idée d'une personnalité juridique des robots autonomes a finalement rapidement été délaissée par l'ensemble des instances européennes comme internationales. Cette proposition apparaît sans doute prématurée car les robots n'ont pas, à ce jour, de réelles capacités cognitives, mais la question risque un jour de se poser à nouveau, ce qui n'est pas sans poser un certain nombre de problèmes, à commencer par une remise en cause de la distinction entre les personnes et les choses.

En l'état actuel du droit, le robot demeure donc une chose. Lorsqu'il cause un dommage, il est possible d'envisager de mettre en œuvre la responsabilité du propriétaire, de l'utilisateur ou de l'opérateur, du concepteur, du fabricant, du programmeur du logiciel intégré au robot ou encore du concepteur de l'IA.

Les dommages causés par une IA utilisée dans le domaine médical doivent donc être traités suivant les règles prévues en droit de la responsabilité médicale. La responsabilité du médecin peut être engagée sur le fondement de différents faits générateurs. Le médecin est tout d'abord responsable des fautes qu'il commet dans l'exercice de sa profession. Le principe est affirmé dans le Code de la santé publique, depuis la loi

Kouchner du 4 mars 2002, qui prévoit que les médecins « *ne sont responsables des conséquences dommageables d'actes de prévention, de diagnostic ou de soins qu'en cas de faute* » (Art. L. 1142-1).

En principe, la faute du médecin lors de l'utilisation d'une application médicale dotée d'IA devrait engager sa responsabilité. En effet, il apparaîtrait étonnant que le médecin soit entièrement exonéré de toute responsabilité pour la simple raison qu'il utilise une technologie dotée d'IA, surtout s'il en fait un mauvais usage. Étant considéré comme un spécialiste dans son domaine, tout objet qu'il utilise pour pratiquer son art doit donc faire partie du champ de sa responsabilité.

En revanche, lorsque l'IA utilisée par un médecin présente un défaut de sécurité, la victime devrait alors avoir la possibilité d'agir contre son producteur en prouvant le dommage, le défaut et le lien de causalité entre le défaut et le dommage. Le principal frein à la mise en œuvre de la responsabilité du producteur risque d'être l'exonération pour « risque de développement », c'est-à-dire que l'état des connaissances scientifiques et techniques, au moment où il a mis le produit en circulation, n'a pas permis de déceler l'existence du défaut.

Le professionnel de santé est également débiteur d'une obligation d'information à l'égard du patient qu'il soigne. Il doit en effet l'informer de tout risque grave concernant les traitements, investigations et actions de prévention qui le concernent. Le patient pourra ainsi lui reprocher un manquement à son obligation d'information s'il ne l'a pas prévenu du risque pouvant résulter de l'utilisation de ces systèmes intelligents.

Finalement, les règles de la responsabilité civile en matière médicale peuvent s'appliquer au développement de l'intelligence artificielle dans le domaine de la santé, mais il faudra sans doute quelques adaptations aux règles actuelles pour répondre à toutes les situations de dommage.

29 Les interfaces cerveau-machine (ICM) dessinent-elles la médecine du futur ?

Les interfaces cerveau-machine trouvent leurs origines à travers les travaux princeps en électrophysiologie de Hans Berger sur les ondes cérébrales en 1920. Il faut toutefois attendre les années 1960 pour voir la démonstration d'une communication directe entre un cerveau et une machine. En effet, une branche de la neurotechnologie s'est développée afin de « lire » l'activité cérébrale et de l'exploiter pour créer un canal de communication direct entre l'utilisateur et des ordinateurs numériques. Ce domaine de recherche est connu sous le nom d'interface cerveau-machine (ICM).

On date généralement l'invention officielle des interfaces cerveau-machine en 1973 lorsque Jacques Vidal publia un article sur la communication cerveau-machine intitulé « *Toward Direct Brain-Computer Communication* ». Dans cet article, le terme anglais « *brain-computer interface* » (BCI), – ou interface cerveau-machine (ICM) –, apparaît pour la première fois. Il sera utilisé pour désigner tous types de communication directe entre une machine et un utilisateur, passant uniquement par l'activité cérébrale, cette dernière étant mesurée et analysée par le système.

Le terme interface cerveau-machine désigne, selon l'Inserm, « *un système de liaison directe entre un cerveau et un ordinateur, permettant à un individu d'effectuer des tâches sans passer par l'action des nerfs périphériques et des muscles* ». Concrètement, l'utilisateur imagine effectuer un mouvement, ce qui génère une activité cérébrale caractéristique et mesurable à l'aide de capteurs. Ces signaux sont alors transmis à un ordinateur qui les analyse pour en extraire les données utiles, puis les transforme en commande pour la machine (prothèse, exosquelette, fauteuil roulant, interface logicielle, etc.).

Dans ce processus de décryptage de l'activité cérébrale, l'intelligence artificielle joue un rôle fondamental. Les algorithmes d'apprentissage automatique sont en effet entraînés à décoder les signaux neuronaux enregistrés et amplifiés pour comprendre l'intention à laquelle ils sont associés, puis de transmettre une commande à des stimulateurs électriques disposés à des endroits spécifiques du corps ou à un appareil externe.

Les domaines d'application des ICM sont nombreux, notamment dans le domaine de la santé, où elles permettent à des personnes handicapées motrices (hémiplégie, paraplégie, victimes d'AVC, etc.) de contrôler une prothèse, un exosquelette, écrire sur un ordinateur ou de le faire parler. Ils offrent également des perspectives thérapeutiques pour les déficits sensoriels.

Les ICM peuvent être invasives, partiellement invasives ou non invasives. Les ICM invasives nécessitent une intervention chirurgicale pour implanter des électrodes dans le cortex cérébral. L'implantation dans le cortex donne une meilleure résolution spatiale que l'ensemble des techniques dites non invasives, mais cette méthode est encore associée à un risque de complications. Dans les ICM partiellement invasives, la grille d'électrodes est placée sous la dure-mère, soit la membrane qui entoure le cerveau juste sous la boîte crânienne. La résolution spatiale est un peu moins bonne qu'avec une implantation dans le cortex, mais les risques de complications sont moindres.

Enfin s'agissant des ICM non invasives, elles créent une interface entre les signaux cérébraux et des technologies d'imagerie cérébrale telles que l'électroencéphalographie (EEG), qui enregistre l'activité cérébrale grâce à des électrodes placées à l'extérieur du crâne. L'activité électroencéphalographique correspond à un courant généré par des milliers de neurones et, malgré une bonne résolution temporelle, a une très mauvaise résolution spatiale. Toutefois ce système est peu cher, réversible, facile d'utilisation et permet d'envisager de nombreuses applications. C'est de fait le mode d'enregistrement le plus utilisé.

Plusieurs entreprises américaines ont mis au point des systèmes d'électrodes qui pénètrent dans le cortex cérébral pour enregistrer les données de neurones individuels, c'est le cas de *Blackrock Neurotech*, *Paradromics* et, bien sûr, de la plus médiatisée d'entre elles, *Neuralink*, ayant posé en janvier 2024 son premier implant cérébral, la puce *Telepathy*, sur un patient humain. La pose d'un implant cérébral sur un patient humain n'est en réalité pas une première. En septembre 2023, l'entreprise néerlandaise Onward avait annoncé qu'elle testait un implant cérébral stimulant la moelle épinière, permettant à des personnes tétraplégiques de retrouver de la mobilité. En 2019, des chercheurs français de l'institut Cinatec avaient testé sur un patient tétraplégique un implant, *Wimagine*, qui décrypte l'activité du cortex moteur, permettant à ce patient de se déplacer, en contrôlant un exosquelette par la pensée. Un autre patient tétraplégique a pu remarcher en 2023 naturellement grâce à un implant cérébral développé par un consortium franco-suisse de neuroscientifiques et neurochirurgiens.

Malgré ces prouesses, les ICM soulèvent un certain nombre de questions, allant de la précision de l'interprétation des signaux neuronaux au risque de modifier l'état mental en manipulant les ondes cérébrales en passant par des risques de discrimination en raison d'une inégalité d'accès à ces technologies. D'autres risques sont aussi à envisager, tels que les risques de dysfonctionnements de l'interface pouvant causer un dommage à l'individu concerné ou à un tiers ainsi qu'une possible prise de contrôle du dispositif par des hackers, conduisant l'individu à donner des ordres contre sa volonté, comme la passation d'ordre de virement en ligne. Enfin, se pose la question du consentement. La personne est-elle en capacité de prendre une décision raisonnée, libre et éclairée avant d'accepter un tel implant ? Autant de questions qui méritent d'être discutées avant la généralisation de tels dispositifs.

30 Existe-t-il des « neurodroits » pour protéger le contenu du cerveau ?

Dans sa Recommandation sur l'innovation responsable dans le domaine des neurotechnologies de 2019, l'OCDE définit les neurotechnologies comme étant des « *dispositifs et procédures utilisés pour accéder au fonctionnement ou à la structure des systèmes neuronaux de personnes naturelles et de l'étudier, de l'évaluer, de le modéliser, d'exercer une surveillance ou d'intervenir sur son activité* ». Autrement, peut être qualifiée de neurotechnologie toute technologie permettant d'explorer, d'influencer ou d'intercommuniquer avec le cerveau humain.

La neurotechnologie est en réalité un domaine interdisciplinaire mêlant les neurosciences et les technologies pour explorer, comprendre et manipuler le système nerveux. Les neurotechnologies se situent ainsi au point de convergence des neurosciences, de l'ingénierie, de la science des données, des technologies de l'information et de la communication, et de l'intelligence artificielle.

Ces outils peuvent être des outils techniques et informatiques qui mesurent et analysent les signaux chimiques et électriques du système nerveux, ou des outils techniques qui interagissent avec le système nerveux pour en modifier l'activité, par exemple pour rétablir l'influx sensoriel, comme les implants cochléaires pour rétablir l'audition.

Au cours des dernières décennies, l'innovation technologique et les découvertes scientifiques dans les domaines des neurosciences, alliées aux progrès de la modélisation informatique et des logiciels d'apprentissage automatique utilisés pour analyser les données, ont permis de réaliser d'énormes progrès en neurotechnologie. Certains parlent même d'une révolution neurotechnologique.

Les progrès réalisés en matière d'implants, d'interfaces cerveau-machines et d'imagerie cérébrale offrent, par exemple, la possibilité d'aider les personnes souffrant de troubles neurologiques ou psychiatriques tels que la maladie de Parkinson, les accidents vasculaires cérébraux, la démence, etc. Or, ces techniques touchent au cerveau, qui n'est évidemment pas un organe comme les autres. En effet, l'activité cérébrale est à la base de notre identité, de nos pensées, de nos émotions, de notre mémoire, etc.

Il y a donc des enjeux éthiques et juridiques majeurs qu'il convient d'analyser. Ces préoccupations ont d'ailleurs donné naissance à deux principaux domaines de recherche, à savoir la neuroéthique et le neurodroit. Le neurodroit, pris au singulier, désigne un domaine de recherche juridique traitant de l'utilisation et de l'évolution des neurotechnologies, alors que la neuroéthique traite des questions de pure éthique.

La réflexion sur le neurodroit a fait l'objet d'une couverture dans les médias grand public, elle est pourtant avant tout une problématique juridique. Le rapport commandé par le comité de bioéthique du Conseil de l'Europe sur les neurotechnologies, rédigé en 2021 par Marcello Ienca, définit les neurodroits comme des « *principes éthiques, juridiques, sociaux ou naturels de liberté ou de droit liés au domaine cérébral et mental d'une personne* ». Il s'agit donc de règles normatives régissant la protection et la préservation du cerveau et de l'esprit humains. Au sens large, les neurodroits peuvent être considérés comme une catégorie émergente de droits de l'homme destinés à protéger l'espace cérébral et mental de la personne.

Conscient des apports, mais aussi des limites que peuvent présenter les neurosciences, le législateur français a, par la loi de bioéthique du 7 juillet 2011, proposé un encadrement de l'utilisation des techniques d'imagerie cérébrale afin d'éviter leur détournement à des fins mercantiles, sans égard pour les droits fondamentaux. Pour ce faire, il a intégré un nouvel article 16-14 dans le Code civil qui limite le recours aux techniques d'imagerie cérébrale à des fins médicales, de recherche scientifique ou dans le cadre d'expertises judiciaires, à l'exclusion de l'imagerie cérébrale fonctionnelle, avec l'obligation de recueillir par écrit le consentement exprès de la personne qui peut le révoquer sans forme et à tout moment. Cette liste limitative des finalités légalement autorisées a eu pour effet d'interdire le recours aux techniques d'imagerie cérébrale à des fins de marketing (ou ciblage commercial), d'assurance ou d'emploi. Mais cette législation se contourne facilement par le recours à des entreprises soumises à des législations moins contraignantes. Malgré ses lacunes, le droit français s'est tout de même saisi de la question des neurosciences, plus généralement des neurodroits, au titre de la bioéthique qui constitue en quelque sorte sa porte d'entrée dans le droit.

D'autres États ont choisi de franchir une nouvelle étape dans la protection de l'intégrité mentale, il en va ainsi du Chili qui est devenu le premier pays au monde à introduire dans sa Constitution la protection de l'intégrité du cerveau en 2021. L'adoption d'un tel arsenal juridique peut sembler prématurée au regard du développement actuel des neurotechnologies, mais les experts insistent sur la nécessité de légiférer avant la généralisation d'applications intrusives.

Les cas concrets se sont pourtant rapidement présentés. En 2023, le sénateur chilien ayant le plus défendu les neurodroits, Guido Girardi, a poursuivi avec succès l'entreprise *Emotiv* devant la Cour suprême de Santiago pour avoir violé les garanties de la Constitution chilienne en matière de collecte et d'utilisation des données neuronales, après avoir importé dans ce pays l'un de ses appareils qui recueillent des informations sur l'activité électrique du cerveau et enregistrant les données neuronales de la personne qui le porte.

Le Chili est actuellement le seul pays à disposer d'une législation protégeant les neurodroits, mais d'autres pays pourraient s'inspirer bientôt de l'exemple chilien. C'est le cas, par exemple, de la Charte espagnole des droits numériques qui comprend des dispositions spécifiques aux neurotechnologies.

IV. L'INTELLIGENCE ARTIFICIELLE DANS LE DOMAINE DE LA JUSTICE

31 L’intelligence artificielle peut-elle être utile à la justice ?

La justice a été longtemps réfractaire au recours à l’intelligence artificielle. Elle s’y est d’abord opposée, car qu’elle y voit l’antithèse de la philosophie qui guide traditionnellement le juge. L’IA permet de rationaliser et de massifier le traitement alors que le juge est censé donner à chaque cas une réponse individualisée tenant compte des particularités de chaque espèce.

Juger n’est pas un acte banal. C’est une personne que l’on juge. Et c’est aussi un acte accompli par un être humain. Sous-jacent à l’acte de juger, il y a souvent un arbitrage qui est effectué entre des valeurs concurrentes qui détermine les valeurs de la société dans laquelle nous vivons.

L’idée que la justice serait une science exacte, sans imprévu ni incertitude, et pouvant obéir à des modélisations mathématiques doit donc être réfutée. Lorsqu’il doit forger son intime conviction, le juge est nécessairement influencé par son propre parcours de vie, sa vision du monde et une sensibilité dont est dépourvue la machine.

Le débat public sur l’intelligence artificielle appliquée à la justice s’est essentiellement concentré sur le cas de la justice prédictive, qui suscite de nombreux fantasmes et passions. D’autres applications fondées sur l’intelligence artificielle pénètrent pourtant le monde de la justice d’une plus façon pragmatique.

Le recours à l’IA est d’abord un espoir pour permettre d’accélérer le travail judiciaire. La lenteur de la réponse judiciaire est en effet l’une des critiques les plus récurrentes émises contre l’institution judiciaire. Il peut alors s’agir de fluidifier en amont le traitement des procédures, ou encore de permettre une réponse accélérée à des cas considérés comme simples, comme les impayés de crédit à la consommation.

La Cour de cassation a, par exemple, développé un moteur de pré-orientation des pourvois vers les chambres spécialisées. Un algorithme d’apprentissage automatique a été entraîné sur une centaine de milliers de mémoires en demande déjà orientés ces dernières années. L’algorithme prédit vers quelle chambre pré-orienter le pourvoi.

Les algorithmes peuvent également aider le juge à traiter les masses de données, tant juridiques que factuelles, liées aux litiges qu’il doit trancher. L’utilisation de ce type d’outils devient d’autant plus intéressante que désormais les décisions de justice françaises sont accessibles en *open data*.

L’IA permet en effet d’analyser un important volume de décisions judiciaires afin de déterminer les tendances jurisprudentielles. L’analyse statistique de ces décisions permet alors aux juges d’accéder rapidement à ceux qui concernent les litiges qu’ils ont à juger. Cette analyse statistique des décisions de justice aide ainsi à la prise de décision et harmonise l’application de la loi d’une juridiction à l’autre pour les citoyens.

La recherche d’une plus grande efficience des tribunaux associée au manque de moyens de la justice permet, à raison, de craindre que les pouvoirs publics optent à terme pour une sorte de sous-justice industrialisée pour les contentieux de masse. L’inquiétude qu’a suscitée l’algorithme *Datajust*, dont le gouvernement français a, le 13 janvier 2022, arrêté

le développement après deux ans de travaux, traduit cependant pour le moment une certaine hésitation des pouvoirs publics quant à l'imposition au juge d'outils pouvant s'inscrire dans le mouvement de la justice prédictive.

Un décret du 27 mars 2020 avait en effet autorisé le ministère de la Justice à mettre en œuvre, pour une durée de 2 ans, un traitement automatique de données à caractère personnel dénommé *DataJust.* La finalité de ce traitement de données était la mise au point d'un algorithme permettant l'élaboration d'un référentiel indicatif d'indemnisation des préjudices corporels.

Ce décret a suscité une grande inquiétude, notamment chez les avocats de victimes. La principale critique émise était que la barémisation de l'indemnisation conduisait à forfaitiser toute la réparation du préjudice corporel. La singularité de la victime pouvait être balayée par l'intelligence artificielle et la loi du plus grand nombre. Le garde des Sceaux a finalement préféré renoncer au projet en invoquant la complexité des données devant être traitées rendant leur extraction non complètement automatisable.

Si, pour l'heure, ces outils ne sont qu'une aide accompagnant le juge dans son travail, il ne faut cependant pas négliger le risque d'une plus profonde transformation de la justice. L'intelligence artificielle apporte en effet autant de dangers que d'opportunités pour le droit et la justice.

32 Qu'est-ce que la justice prédictive ?

La justice prédictive est née dans les pays de *Common Law* dont la particularité est de disposer d'un système jurisprudentiel reposant sur des actions procédurales. Il est important de rappeler que l'absence de droit écrit a contraint le système juridique de *Common Law* à se stabiliser en adoptant une règle qui est celle du « précédent » ou *stare decisis*. Autrement dit, une décision de justice ne doit pas contredire une plus ancienne. Ainsi, les juridictions de *Common Law* sont liées par toute décision de justice antérieure ayant tranché un cas similaire par une juridiction supérieure ou de même niveau dans l'ordre judiciaire, sauf s'il existe une différence notable dans les faits de l'affaire qui justifie une solution différente.

Ce droit, presque exclusivement jurisprudentiel, se prête particulièrement bien à une approche prédictive de la justice à la différence de notre système de droit écrit. La nécessité de trouver une jurisprudence idoine a rendu possible l'émergence d'outils d'aide à la décision en matière juridique. L'idée serait de disposer d'instruments qui, se fondant sur une analyse de la jurisprudence existante, pourraient prédire ce que sera la jurisprudence future.

Or, prédire renvoie à l'intuition et aux présages. Aussi, il n'est pas certain que l'expression « justice prédictive » soit la plus appropriée pour décrire ce qui relève davantage de la prévision reposant avant tout sur l'observation du réel, c'est-à-dire l'étude de la jurisprudence et la mise au point d'hypothèses.

La notion de justice prédictive est donc ambiguë et fallacieuse, puisqu'il ne s'agit pas d'« une justice qui prédit », mais plutôt d'« une justice prédite par des algorithmes » grâce à l'analyse de grandes masses de données de justice (*big data*) ouvertes (*open data*) afin de repérer des récurrences permettant de prévoir autant qu'il est possible l'issue d'un litige. L'adoption de cette notion suit le développement des entreprises de technologie juridique, dites *legaltechs*, fondées sur une utilisation marchande des algorithmes. Concrètement, il s'agirait de déterminer les probabilités de succès d'une affaire au moyen de l'analyse des décisions antérieures rendues dans le même domaine. L'objet des outils de justice prédictive n'étant pas de reproduire un raisonnement juridique, mais d'identifier les corrélations entre les différents paramètres d'une décision (par exemple, entre la durée d'un mariage et le montant d'une prestation compensatoire).

C'est donc bien plus une justice prévisible dont il est question, car une décision de justice n'est pas la résultante des seules décisions passées, mais d'une pluralité de facteurs plus ou moins bien identifiés : normatif, politique, social, professionnel, médiatique, affectif, etc. Il s'agit avant tout d'une œuvre humaine.

Le droit se caractérise en effet par une exigence de prévisibilité juridique. À cet égard, l'outil prédictif peut permettre de rendre le droit plus accessible, plus transparent et plus fiable. C'est d'ailleurs l'argument soutenu par les promoteurs de la justice prédictive qui indiquent qu'en améliorant la prévisibilité de la jurisprudence, celle-ci permettra non seulement de sécuriser l'action en justice, mais aussi de favoriser le recours aux modes alternatifs de résolution des litiges (*Online Dispute Resolution, ODR*).

Malgré les risques inhérents à son déploiement, la justice prédictive est pourtant promue avec l'ambition d'en faire un substitut au recours au juge. Elle est notamment mise en lien avec la justice amiable telle que la médiation en ligne. L'usage d'un algorithme pour faciliter le traitement des litiges en ligne n'est cependant pas un facteur anodin, car cela pourrait aboutir à un refoulement du droit dont la fonction est de concilier les grands équilibres et de protéger la partie faible comme en droit de la consommation ou en droit du travail. Le risque est alors de créer amiablement un sous-droit.

L'institution judiciaire semblait avoir été longtemps épargnée par cette révolution algorithmique. Le procès devant un juge étatique constituait l'alpha et l'oméga de la résolution des litiges. Un modèle de justice plurielle combinant justice étatique et justice amiable, justice physique et justice numérique, semble pourtant se dessiner.

En réalité, l'effet réel d'évitement du juge, que permet la justice prédictive, dépendra du type de litige, car le justiciable voit le procès comme un moyen d'obtenir justice. Il y a une dimension symbolique, psychologique et sociale au procès que les médiateurs ne pourront pas remplacer. Dans ces conditions, la justice prédictive pourrait conduire à la déjudiciarisation de certains contentieux.

La justice prédictive lance tout de même un défi à l'office du juge : celui de se réinventer, de se renouveler, sans se dénaturer. L'un d'eux a trait à la performativité. Celui-ci s'entend comme le fait de voir se réaliser ce qui est énoncé. Le danger pour le juge est ainsi celui de la facilité et du choix de l'automaticité par reproduction des solutions indiquées et quantifiées par le logiciel de justice prédictive, sans s'arrêter autrement aux circonstances de la cause.

Le juge, confronté à ces éléments statistiques, va sans doute s'en trouver influencé. Il prend alors une décision non en exerçant sa propre appréciation du litige, mais parce que l'outil lui restitue ce que font majoritairement ses pairs en pareille situation. De ce fait, la prévision se transforme en prophétie auto-réalisatrice. Certains craignent que les résultats issus des algorithmes de justice prédictive soient annonciateurs d'une uniformisation de la pensée judiciaire. La justice prédictive peut ainsi devenir un facteur de pressions exercées sur le juge et contrevenir à la maturation de la jurisprudence. Confronté à une machine prétendant prédire ses décisions, le juge ne risque-t-il pas de perdre sa créativité ? Il est certain que la justice prédictive n'est pas neutre sur l'office du juge.

33 Existe-t-il un lien entre l'*open data* des décisions de justice et l'intelligence artificielle ?

L'*open data* ou, l'ouverture des données en français, est définie dans le *Journal officiel* comme la « *politique par laquelle un organisme met à la disposition de tous des données numériques, dans un objectif de transparence ou afin de permettre leur réutilisation, notamment à des fins économiques* ».

À l'origine, ce mouvement a été porté par la société civile, puis a trouvé écho auprès des pouvoirs publics. L'*open data* est avant tout une philosophie, une volonté citoyenne, visant à considérer l'information publique comme un bien commun. Il est en essor depuis les années 2000, stimulé par les capacités techniques de reproduction, de distribution et de traitement des données numériques. Il part d'un principe qui est celui de la transparence de l'État de droit. Il fait aussi le pari que la réutilisation de ces données sera source de croissance et d'innovation.

D'un point de vue général, le mouvement de l'*open data* promeut l'idée d'un libre accès à un certain nombre de données publiques afin d'en permettre une utilisation et une exploitation sans restriction de droits d'auteur, de brevets ou d'autres mécanismes de contrôle. Pour être reconnues comme ouvertes, les données doivent être rendues disponibles dans un format standard ouvert et permettre la reproduction et la réutilisation libre et gratuite par tous. Sont principalement concernées les données publiques et les données issues de la recherche scientifique, considérées comme des biens communs informationnels. L'*open data* repose ainsi sur trois exigences cumulatives : la disponibilité des données, leur exploitabilité et la liberté de leur réutilisation.

En France, le cadre juridique de l'ouverture des données publiques a été réalisé en deux étapes. La première a consisté en l'affirmation progressive d'un droit d'accès aux documents publics, tandis que la seconde aboutira à la reconnaissance d'un droit de réutilisation des informations du secteur public, ainsi que la mise en place de mesures institutionnelles en ce sens.

Avec la loi pour une République Numérique, dite « Lemaire », promulguée le 7 octobre 2016, l'ouverture des données publiques devient la règle et non plus l'exception. Le principe retenu par le législateur français est celui de l'*open data* par défaut, ce qui signifie que les administrations et les collectivités publiques doivent diffuser tous les documents dématérialisés, toutes les bases de données, qu'elles produisent ou qu'elles reçoivent dans un standard ouvert, aisément réutilisables et exploitables par un système de traitement automatisé. La loi vise également les données scientifiques qu'elle définit comme les données issues d'une activité de recherche.

Les décisions de justice font partie de cette règle d'ouverture mais leur spécificité, et surtout le caractère sensible des informations qu'elles contiennent, ont exigé la création d'un cadre juridique distinct afin d'occulter notamment les données personnelles.

La mise à disposition du public sous forme électronique des décisions de justice rendues par les juridictions administratives et judiciaires vise plusieurs objectifs : favoriser la confiance des citoyens par une meilleure transparence des décisions de justice ; améliorer

la sécurité juridique par une meilleure connaissance de la jurisprudence grâce à l'analyse des décisions rendues ; et favoriser l'émergence de nouveaux outils ou services (probabilité de succès d'une action, etc.).

Pour valoriser la mise à disposition du public de ces décisions de justice, deux moteurs de recherche ont été créés : « *Décisions de la justice administrative* » et « *Judilibre* », qui concernent respectivement les décisions de justice administrative et judiciaire. La mise à disposition des décisions de justice en ligne se fera progressivement selon un calendrier qui s'étend jusqu'en 2025. L'objectif est de parvenir à terme à la mise en ligne de 300 000 décisions administratives et de 3 millions de décisions judiciaires chaque année, alors qu'avant l'*open data*, environ 15 000 décisions étaient diffusées chaque année sur Légifrance.

Il y a donc un changement d'échelle considérable dans la diffusion des décisions de justice. Et l'on comprend l'inquiétude de ceux qui affirment que l'avènement de l'*open data* risque d'entraîner de nouveaux changements de la jurisprudence. Le premier risque identifiable est celui d'un nivellement des décisions de justice. En effet, toutes les décisions de justice ne se valent pas : il n'est pas possible de mettre sur un pied d'égalité le jugement d'un tribunal judiciaire et un arrêt de la Cour de cassation, ou encore, un arrêt d'espèce et un arrêt de principe. Ce risque d'indifférenciation des décisions de justice nécessite un travail de hiérarchisation des décisions de justice afin de mettre en avant les décisions présentant un véritable intérêt pour le droit qui seraient noyées dans une masse colossale de données.

Le second risque tient à l'appauvrissement de la créativité juridique dans la mesure où le traitement algorithmique des décisions de justice, en faisant apparaître récurrences et régularités dans la manière de juger, peut favoriser l'avènement de tendances conformistes dans l'application des règles de droit par les juges. Tout dépendra alors de l'exploitation qui est faite des décisions de justice par les systèmes algorithmiques.

Dans le cas le plus courant, il s'agit de simplement mettre en œuvre un moteur de recherche des décisions permettant à l'utilisateur, en formulant une requête, de retrouver une décision. Les systèmes algorithmiques peuvent aussi fournir des statistiques relatives à des décisions rendues sur un sujet donné dans un domaine déterminé. Il peut enfin s'agir d'un logiciel qui étudie la jurisprudence et qui y repère les critères de la décision des juges à des fins de modélisation grâce à des outils d'intelligence artificielle.

Le risque principal lié à l'utilisation de l'intelligence artificielle en matière d'*open data* des décisions de justice réside dans la difficulté de vérifier la conformité du raisonnement suivi aux règles de droit en vigueur en raison de l'opacité, la complexité et l'imprévisibilité de nombreux systèmes d'IA. Les interrogations sur la fiabilité des algorithmes, et les éventuelles dérives liées à leur opacité, sont évidemment des questions qui ne favorisent pas la confiance en ces outils.

34 Qu'est-ce qu'une *legaltech* ?

Une *legaltech* est une *Legal Technology*, anglicisme signifiant « technologie juridique ». Apparue en 2004 aux États-Unis, elle n'est pas légalement définie. Il s'agit concrètement d'une start-up faisant usage de la technologie et de logiciels performants afin d'offrir des services juridiques en ligne sans passer directement et nécessairement par un avocat, un expert-comptable ou un notaire. Les premières entreprises *legaltech* en France ont vu le jour au début des années 2000. Elles ont permis la démocratisation de certains services, tels que la création automatique de documents juridiques.

Initialement, les *legaltechs* ne travaillaient qu'au profit des professionnels, et plus spécifiquement, pour les cabinets d'avocats. Toutefois, depuis les années 2010, elles proposent également leurs services aux entreprises et aux particuliers.

La deuxième décennie du XXIe siècle a vu une véritable explosion du nombre de ces start-ups en France. Leur succès peut s'expliquer par les gains temporel et financier qu'elles permettent du fait de l'automatisation de la possibilité de générer des documents juridiques.

Parmi les plus connus, nous pouvons citer « Doctrine.fr » qui développe et commercialise un moteur de recherche juridique à destination des professionnels du droit ; « Predictice.com » qui simplifie la recherche, l'analyse juridique pour les professionnels du droit et qui a développé en 2023 la première IA générative pour ces derniers ; ou encore « Legalplace.fr » qui fournit de l'information juridique, un logiciel permettant de générer des documents juridiques sur la base de modèles (bail commercial, contrat de travail, etc.) et un service de formalités d'entreprises.

Les systèmes d'IA sont essentiels dans la génération de contrats, ou encore en matière d'exploitation des décisions de justice en *open data*. Mais c'est surtout le développement de systèmes d'IA génératives, tel que ChatGPT, qui bouleverse le paysage et soulève de nouvelles questions juridiques. Les systèmes d'IA génératives facilitent, à ce titre, l'accès à l'information, mais posent des problèmes spécifiques aux professions réglementées, notamment la garantie du respect du secret professionnel et de la sécurité des données.

Autrefois éloignées de la publicité, les professions juridiques réglementées se révèlent aujourd'hui de plus en plus vulnérables. Dans cette nouvelle économie, les professions réglementées sont de plus en plus soumises à la concurrence. De nombreux professionnels du droit, tels que les avocats, ont vu dans cette évolution une menace à leur activité.

Lorsqu'une personne se heurte à un problème juridique et avant de faire appel à un avocat, elle tentera de trouver une réponse en effectuant une recherche en ligne. Les *legaltechs* peuvent alors la capter grâce à leur présence sur le web ; la technologie favorisant les communications directes et simplifie l'échange de documents.

Face à cette mise en concurrence, de nombreux cabinets d'avocats traditionnels ont commencé à collaborer avec des *legaltechs* afin de rester compétitifs et de gagner en visibilité. En réalité, ces dernières libèrent les avocats de certaines tâches (rédactions d'actes, recouvrement…), ce qui leur permet de se consacrer à des prestations plus

complexes. Le transfert à la *legaltech* de tâches chronophages et à faible valeur ajoutée pourrait alors être perçu comme un avantage non négligeable. De leur côté, les *legaltechs* ont besoin des connaissances et de l'expertise des avocats afin de gagner en crédibilité auprès de leurs clients.

Le contrôle des documents générés par un avocat peut toutefois s'avérer nécessaire *a posteriori* afin d'éviter le risque d'un contentieux. C'est la raison pour laquelle des avocats investissent ce marché en créant leur propre plateforme de prestation de services juridiques afin d'apporter une sécurité juridique supplémentaire.

Le principal obstacle se dressant sur la route de l'avocat souhaitant créer une *legaltech* se trouve dans sa raison d'être, car elle requiert des compétences techniques pointues pour apporter une réelle valeur ajoutée aux utilisateurs. Or, les avocats ne possèdent pas les savoir-faire et compétences techniques nécessaires. Aussi, la création d'une *legaltech* passe souvent par l'association avec des profils complémentaires au pur juridique : des diplômés d'écoles d'ingénieurs et de commerce, etc.

Le développement de ces technologies a exigé l'adoption de nouvelles dispositions législatives pour remédier aux lacunes laissées par l'application de réglementations générales n'ayant pas pu les prévoir. La loi du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice (dite « loi de programmation pour la justice ») a, par exemple, posé un nombre de règles destinées à réglementer tout ou partie de l'activité des prestataires de *legaltechs*. C'est ainsi que dans le cadre de l'*open data* des décisions de justice, des restrictions ont été fixées pour le traitement des données personnelles qu'elles contiennent. L'occultation des noms et prénoms des parties et des tierces personnes physiques doit intervenir préalablement à la mise à disposition des décisions de justice. Par ailleurs, en cas de risque d'atteinte à la sécurité ou au respect de la vie privée, toute autre donnée personnelle des parties et des tiers, mais aussi des magistrats et membres du greffe, peut aussi être occultée.

S'agissant des opérateurs proposant un service en ligne de médiation ou de conciliation, leur activité a été encadrée, ce qui fait d'eux les premiers prestataires réglementés de *legaltechs*. Ils sont désormais soumis à des règles d'éthique (impartialité, indépendance, compétence et diligence), de protection des données personnelles (prohibition des services fondés exclusivement sur un traitement automatisé de données personnelles sans information et consentement des parties) et de protection des consommateurs (information détaillée sur les modalités d'arbitrage).

Il est évident que les *legaltechs* influent, indirectement au moins sur la manière dont le droit est appliqué, ce qui nécessite parfois une intervention législative afin d'adapter la réglementation en vigueur. De manière générale, le développement de ces *legaltechs* entraîne une transformation du droit et de la manière dont le juriste constitue et développe son savoir juridique.

35 Doit-on être jugé par des algorithmes ?

Le droit apparaît comme un terreau fertile au processus d'algorithmisation. L'apparente logique du raisonnement juridique conduit certains à penser qu'il y a là matière à modélisation. La structuration de la règle juridique semble d'une certaine façon de type binaire (qualification juridique/régime, conditions/effets, principe/exception…). La règle de droit est aussi écrite pour résoudre un problème. La particularité de l'intervention du juge réside dans le mode de raisonnement qu'il met en œuvre, connu sous le nom de syllogisme judiciaire, qui suppose trois étapes de réflexion : examen des faits de l'espèce : la « mineure » ; recherche de la règle de droit applicable à ces faits : la « majeure » ; application de la règle aux faits : la conclusion du syllogisme. L'algorithme, comme le syllogisme judiciaire seraient ainsi comparables à des recettes de cuisine : en mélangeant différents ingrédients d'une manière bien comprise, tout cuisinier obtiendra le même plat ; tout juge aboutira à la même solution.

La structuration de la règle de droit est toutefois une chose, sa mise en œuvre en est une autre. Il n'est en effet pas certain qu'une suite de 0 et de 1 puisse rendre la justice et attribuer équitablement à chacun ce qui lui est dû. La formule scientifique du savoir ne résout pas les questions de valeurs et de sens. Si la décision de justice se limitait au résultat d'un parfait syllogisme, il ne serait en effet guère difficile d'élaborer un algorithme capable de rendre la justice. La justice peut-elle raisonnablement répondre à une logique mathématique implacable ?

La quantification des comportements humains par l'intermédiaire d'outils de prédiction algorithmique peut donner dans un certain sens l'illusion d'une métrique objective. Or, la seule corrélation statistique entre deux événements est insuffisante pour expliquer les facteurs réellement causatifs. L'ambition de rendre objective les décisions des juges grâce à des algorithmes comporte donc des risques d'explications erronées. L'explication d'une décision judiciaire suppose en effet une analyse plus approfondie des données de chaque affaire, et ne saurait se limiter aux seuls calculs probabilistes.

Le célèbre professeur de droit Jean Carbonnier a pu ainsi écrire que « *le juge est un homme et non une machine à syllogismes : autant qu'avec sa connaissance des règles et sa logique, il juge avec son intuition et sa sensibilité* » (J. Carbonnier, *Droit civil vol. I : Introduction*, PUF, coll. « Quadrige », 2004, p. 23). Ce faisant, le bon juge ne serait pas celui qui applique scrupuleusement la loi aux faits qui lui sont soumis, mais celui qui adapte cette loi aux circonstances particulières de chaque espèce en prenant en compte les données sociales, morales, économiques et en pondérant les intérêts en présence. C'est la raison pour laquelle le droit appliqué s'écarte souvent du droit normalement applicable. Le juge ne peut donc être considéré comme la simple « bouche » de la loi, c'est-à-dire un simple exécutant sans marge de manœuvre se limitant à appliquer un syllogisme tel un automate. L'éthique et la morale ont, de ce fait, leur place dans l'appréciation des juges.

Il est vrai que l'algorithme se pare d'une rationalité qui donne l'impression d'une neutralité du processus de décision automatique là où une analyse humaine serait plus sujette à la subjectivité. Par ailleurs, les algorithmes, étant dépourvus de toutes pensées

et opinions, devraient afficher une objectivité et une impartialité totales concernant les informations traitées. Les algorithmes sont supposés être impartiaux, incapables de déloyauté ou de complaisance. L'algorithme ne pourrait pas être corrompu ou autrement influencé. Dans un sens, ils rendraient accessible l'idéal d'un modèle de justice quasi-scientifique. Cependant, si cette impartialité présente la qualité d'être totale, elle a aussi le défaut d'être aveugle, c'est-à-dire plaquant à une situation d'espèce une solution niant sa spécificité là où le juge humain aurait cherché à la corriger en faisant usage des méthodes d'interprétation. C'est bien cette marge d'interprétation qui rend possible la partialité des juges.

Les différentes techniques de l'apprentissage automatique peuvent paraître en elles-mêmes neutres sur le plan des valeurs sociales, mais c'est loin d'être le cas en pratique. Les algorithmes ne sont en effet que le pur produit de l'homme ; les critères, les paramètres et, surtout, les données qui font qu'un algorithme aboutit à certains résultats plutôt qu'à d'autres sont déterminés par des hommes.

L'approche mimétique, adoptée par la plupart des algorithmes d'apprentissage, entérine de ce fait les biais présents dans les bases de données qui les ont initialement alimentés. Or, des biais ou même des erreurs de programmation peuvent avoir des conséquences insatisfaisantes, voire injustes.

Il peut s'agir d'un biais de *fairness* qui désigne un écart systématique entre les décisions enregistrées dans la base de données et des décisions justes (*fair*). Le concept de *fairness* renvoie à la qualité idéale de l'algorithme qui est celui de ne pas être à l'origine de décisions discriminatoires. Il s'agit, par exemple, de s'assurer que la réponse fournie par un algorithme (montant d'une indemnité ou peine) soit la même dans le cas où c'est uniquement l'origine ethnique ou sociale qui distingue les dossiers de deux justiciables.

Seule la transparence, sous une forme intelligible, des principes de fonctionnement de l'algorithme, pourrait permettre d'identifier ces biais et de les corriger. Cette transparence sera évidemment plus difficile à mettre en œuvre lorsqu'il s'agit d'un algorithme auto-apprenant dont le fonctionnement évolue au fur et à mesure et qui est susceptible de réviser de lui-même les règles qu'ils appliquent. La maîtrise de leur fonctionnement est donc encore plus compliquée que pour les algorithmes n'évoluant pas de la sorte. Si certaines précautions ne sont pas prises, il n'est pas exclu que de tels algorithmes conduisent à influencer les juges dans un sens particulier. L'algorithmisation de la justice semble alors être bien plus qu'un simple projet technique.

36 L’intelligence artificielle peut-elle remplacer les avocats ?

En janvier 2024, une nouvelle application lyonnaise « iAvocat », qui proposait de remplacer les avocats par une intelligence artificielle, s’est retrouvée au cœur d’une polémique. L’application comprend un agent conversationnel, assisté par une intelligence artificielle, dont la mission est de conseiller juridiquement. Concrètement, elle permettrait à ses utilisateurs de répondre à leurs questions juridiques, et ce, en vulgarisant les textes de lois. Mais, les réponses qu’elle propose étant très similaires à celles de la version gratuite de ChatGPT, elles peuvent parfois être totalement erronées.

Surtout, les avocats considèrent que l’application, dont le fondateur n’est pas un avocat, constitue un exercice illégal de cette profession réglementée. Celle-ci étant réglementée, toute personne qui prétend la pratiquer sans diplôme tombe sous le coup de l’exercice illégal de la profession. Le second problème soulevé par cette application est la question de la protection des données, potentiellement confidentielles, confiées à une société immatriculée à Dubaï, échappant ainsi au droit français. Cette controverse a fini par prendre fin avec une mise en demeure adressée par l’Ordre des avocats de Paris à « iAvocat » exigeant que l’application soit supprimée des plateformes d’applications mobiles.

Derrière les remous provoqués par le lancement de cette application se cache en réalité une véritable peur de l’intelligence artificielle ; la peur du remplacement des avocats par des systèmes d’IA. En peu de temps, l’intelligence artificielle est passée du statut de concept mal compris à celui de réalité menaçante. Cela s’explique notamment par la médiatisation de ChatGPT, chatbot développé par *OpenAI*, et de son succès fulgurant. De nombreuses applications intéressant directement les avocats, le plus souvent construites à partir de ChatGPT, ont d’ailleurs fait leur apparition.

C’est, par exemple, le cas de *Perplexity*, qui se présente sous la forme d’un moteur de recherche et d’un générateur de texte, entraîné sur un fond sélectionné pour sa fiabilité et connectée au web, qui peut donner (pas toujours) des réponses correctes là où ChatGPT échoue. L’avantage est qu’il met en avant les différentes sources juridiques utilisées pour formuler sa réponse à l’utilisateur (thèses, articles de presse rédigés par des avocats, etc.).

Citons aussi le cas de *DoNotPay*, qui est un chatbot de services juridiques, conçu à l’origine pour contester les amendes de stationnement et les frais bancaires injustifiés, mais dont les missions se sont étendues avec le temps. En février 2023, l’intelligence artificielle *DoNotPay* devait assister comme avocat un automobiliste accusé d’excès de vitesse devant un tribunal californien. Plus précisément, le système devait conseiller l’accusé et lui souffler les meilleurs arguments pour sa défense à l’aide d’écouteurs, après avoir écouté ce qui a été dit dans la salle d’audience. L’expérience a finalement été annulée à la suite de menaces de poursuite pour exercice illégal de la profession d’avocat.

Malgré la multiplication de ces applications, le risque pour la profession d’avocat demeure, pour le moment, assez limité. En tout état de cause, les systèmes d’IA ne sont pas, à ce stade, en mesure de reproduire toutes les capacités cognitives humaines et ne peuvent donc pas remplacer l’accompagnement personnalisé d’un avocat.

En effet, plus une prestation intellectuelle est difficile à réaliser, plus l'intelligence artificielle semble inefficace, car elle n'est pas dotée de la faculté d'analyse, de création et d'appréciation, notamment des évolutions jurisprudentielles et des considérations humaines. En l'état, il est donc peu probable qu'une intelligence artificielle puisse remplacer complètement l'avocat. Elle permet tout au plus de réaliser avec succès la recherche de jurisprudence, la rédaction de documents juridiques ou encore la prédiction de l'issue d'un litige, ce qui conduit incontestablement à une transformation de la manière d'exercer la profession sans altérer fondamentalement son essence. Par conséquent, la profession devra organiser de nouvelles règles de déontologie pour éviter les éventuelles dérives.

37 Les algorithmes peuvent-ils résoudre les litiges à l'amiable ?

Les plateformes de règlement amiable des litiges, très présentes outre-Atlantique, se développent aujourd'hui en France. Il s'agit de procédés qui ont pour objet la résolution des conflits en dehors de l'institution judiciaire. Face à une demande de justice toujours plus forte de la part des justiciables, il s'agit de soutenir le développement des voies alternatives au juge dans le but d'alléger la charge des juridictions. En promouvant et en imposant le règlement amiable des conflits, l'accès au juge est ainsi conditionné, voire retardé. Ces outils introduisent aussi une certaine souplesse dans le règlement des litiges : consensus, participation du justiciable, confidentialité, rapidité, etc. L'idée est ainsi d'utiliser l'outil numérique pour faciliter et accélérer la recherche d'un accord amiable en cas de litige.

Les incitations à développer ces services en ligne de règlement extrajudiciaire des différends sont aussi bien internationales, européennes que nationales. L'Union européenne a, par exemple, pris l'initiative de créer et d'administrer elle-même une plateforme – ouverte le 15 février 2016 – de règlement en ligne des litiges (plateforme RLL) de l'e-commerce entre consommateur et professionnel.

Se sont ainsi développées des plateformes de résolution des litiges qui utilisent une variété de techniques. Parmi elles, certaines n'utilisent la technologie que pour faciliter l'accès à l'information et les échanges des personnes impliquées. Il s'agit ici de plateformes proposant un accès à des courriers électroniques et des documents en ligne, des discussions en direct, etc.

Dans d'autres cas, la plateforme remplace l'être humain puisqu'aucun médiateur humain n'intervient. Le processus de résolution du conflit est alors géré par des algorithmes qui vont traiter les informations fournies par les parties, évaluer les offres faites par elles et, au terme du processus, proposer un accord amiable. Ces plateformes qui utilisent de tels algorithmes mettent ainsi à distance l'intervention humaine, voire la suppriment. Il s'agit là d'un refoulement de l'humain ayant pour principal risque la déshumanisation de la justice.

Pour éviter d'éventuelles dérives, l'article 4.3 de la loi du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice précise que le processus ne peut avoir pour seul fondement un traitement algorithmique ou automatisé des données à caractère personnel. Aussi, les services en ligne de conciliation ou de médiation ne pourront pas avoir pour seul fondement un traitement algorithmique ou automatisé de données à caractère personnel, ils devront être contrôlés par un humain. La loi autorise ainsi l'aide automatisée à la décision, mais pas la décision totalement automatisée. Contrairement aux décisions de justice rendues par un juge étatique, ces services en ligne peuvent donc utiliser des algorithmes pour appuyer leurs décisions, à la condition que les parties y aient consenti. C'est, par exemple, le cas de la plateforme « Justice.cool » qui se présente comme la première solution de médiation pour les petits litiges entièrement digitale et assistée par intelligence artificielle.

Avec l'aval du législateur en 2019, des acteurs privés, les *legaltechs*, ont ainsi investi la résolution amiable des différends. La justice est pourtant un attribut régalien de l'État. Le développement du règlement amiable en ligne devient alors un signe de désengagement de l'État, celui-ci réduisant son service public pour laisser place à des acteurs privés. La résolution en ligne des différends est désormais une pièce d'une justice plurielle offerte aux justiciables.

En réalité, ces services en ligne de règlement extrajudiciaire des différends, parce qu'ils permettent un traitement rapide et à coût réduit, peuvent sembler adaptés aux petits litiges sériels qui pèsent sur le fonctionnement de la justice étatique. Dans certains cas, le législateur impose au justiciable d'avoir recours à un mode amiable de règlement des différends, préalablement à la saisine du juge. L'article 750-1 du Code de procédure civile prévoit en effet qu'à peine d'irrecevabilité, le titulaire d'une créance n'excédant pas 5 000 euros doit tenter un règlement amiable avant de pouvoir saisir le juge.

Le législateur est ainsi passé de l'incitation à l'obligation. Ce faisant, l'idée est que le juge civil ne doit plus prioritairement régler les litiges relatifs aux petites créances, les justiciables étant invités à tenter d'abord une procédure amiable. Malgré l'incitation des pouvoirs publics ayant souhaité encourager le développement des services en ligne de règlement extrajudiciaire des différends, la place aujourd'hui occupée par ces services reste très en-deçà des espérances. Seul un nombre limité de médiateurs de la consommation arrivent à capter un volume conséquent de différends.

Le justiciable peine en effet à distinguer toutes les offres de médiation de la consommation, auxquels s'ajoutent les plateformes génériques de règlement extrajudiciaire des différends, ainsi que des services de règlement amiable développés par les plateformes de l'e-commerce (par exemple, *eBay* a mis en place un système automatisé de résolution des litiges) et de l'économie collaborative. C'est la raison pour laquelle ces services, créés par des opérateurs privés, ne parviennent pas à pérenniser leur activité alors qu'ils sont censés offrir une alternative à la justice étatique.

38 Quels sont les risques et les enjeux de la justice prédictive dans le domaine pénal ?

La justice prédictive, imaginée par le romancier Philip K. Dick et portée à l'écran dans le film *Minority Report*, présente une société futuriste dans laquelle trois êtres humains mutants, les précogs, peuvent prédire les crimes à venir grâce à leur don de précognition. Grâce à ces visions du futur, la ville a réussi à éradiquer la criminalité et les agents de l'organisation gouvernementale Précrime peuvent arrêter les criminels avant qu'ils ne passent à l'acte.

Bien que la prédiction des actes criminels tienne encore de la science-fiction, le développement de l'intelligence artificielle et de la justice prédictive permettent de ne pas exclure totalement la réalisation un jour d'un tel scénario. L'idée fondamentale associée à *Minority Report* est la suivante : décider ou agir sur le fondement d'une prédiction.

La justice pénale prédictive est, à ce titre, l'expression d'une philosophie contraire à la « justice pénale rétributive », fondée sur la preuve d'une culpabilité. Cette justice pénale classique reconnaissant le libre arbitre de tout individu est à l'origine du couple « responsabilité/punition ». Dans le cas de la justice pénale prédictive, elle se fonde sur un pronostic, un calcul probabiliste de récidive à l'origine d'un autre couple « dangerosité/mesures de sûreté », et rejette toute forme de libre arbitre au nom d'un prétendu déterminisme d'ordre biologique, économique ou bien social. L'utilisation des algorithmes en matière pénale marque la résurgence d'une approche déterministe des comportements criminels

D'ailleurs, des algorithmes prédictifs présentés comme de véritables oracles du crime, existent déjà, citons le cas du logiciel américain Predpol (rebaptisé depuis Geolitica), ayant connu un véritable succès depuis 2011.

Ce logiciel a été présenté par ses concepteurs comme pouvant prévoir le moment et le lieu de commission des crimes dans les douze prochaines heures en traitant dix années de données, incluant notamment les types de crimes et leurs dates, les heures et les lieux où ils ont été commis. Par un traitement statistique représenté sous forme cartographique, les services de police pouvaient ensuite diriger efficacement les patrouilles sur de potentiels lieux d'infractions. L'utilisation de ce logiciel a finalement été abandonnée au fil des années en raison de ses mauvais résultats. Le système PredPol avait un taux d'erreur d'environ 99 %. En d'autres termes, les prédictions ne correspondaient quasiment jamais aux crimes. En dépit de cet échec, les outils de prédiction se multiplient aux États-Unis et au-delà, semblant annoncer l'ère de la prédiction des crimes rendue possible grâce à l'apprentissage automatique.

La notion de prédiction n'est en réalité pas totalement étrangère à la justice pénale. En effet, certaines décisions pénales reposent sur une prédiction relative au comportement futur des détenus. Le juge de l'application des peines, qui doit statuer sur une demande de libération conditionnelle, doit, par exemple, se demander si le détenu présente un risque de récidive. En réalité, la prédiction du comportement était déjà une préoccupation dès les années 1920, soit bien avant l'apparition des algorithmes prédictifs. Les données

massives et les systèmes d'IA donnent cependant une autre ampleur à ces outils prédictifs. Le recours à ces outils soulève nécessairement un ensemble de questions philosophiques, juridiques et techniques. Un risque de récidive calculé sur la base d'un calcul statistique est-il pertinent pour juger un détenu ? Peut-on juger un cas particulier à partir de cas similaires passés ?

Si, en France et en Europe, le recours à ces outils n'en est encore qu'au stade de l'expérimentation, il n'en va pas de même aux États-Unis où le recours aux algorithmes prédictifs est en plein essor depuis les années 2000 ; le plus connu et le plus controversé étant le logiciel COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*). Il est notamment utilisé par les tribunaux américains pour évaluer les risques de récidive et a une influence considérable sur le prononcé et l'exécution des peines.

COMPAS, comme la plupart des outils d'évaluation du risque de récidive, fonde ses prédictions sur un certain nombre de données concernant le prévenu. Concrètement, un travailleur social va répondre, en collaboration avec le prévenu, à plus d'une centaine de questions, sans aucune référence à son origine ethnique, parfois éloignées des faits reprochés, comme la difficulté à payer des factures ou le contexte familial, ou encore des questions sur ses antécédents de consommation de drogue et sur ses amis. L'évaluation se fonde sur le questionnaire dont sont extraits 137 critères pondérés de manière à aboutir à un risque de récidive allant de 0 (risque nul) à 10 (risque maximal). Dans la mesure où l'algorithme analyse exclusivement le risque de récidive du prévenu à travers le prisme de ces données, le choix des critères retenus par le concepteur de l'algorithme est crucial, d'autant plus qu'il a refusé de fournir plus d'explications quant à son fonctionnement arguant qu'il s'agissait d'un secret d'affaires. Ces critères sont souvent élaborés à partir de théories criminologiques. De la théorie de l'apprentissage social sont, par exemple, tirés des critères liés à la proximité avec des pairs antisociaux. C'est la raison pour laquelle l'on peut alors trouver dans le questionnaire la question suivante : « Combien de vos amis/connaissances ont déjà été arrêtés ? ».

La démarche peut sembler scientifiquement fondée. L'algorithme COMPAS a pourtant vu sa logique implacable confrontée à sa partialité discriminante dans la célèbre affaire Loomis. Dans cette affaire, un individu a été arrêté, en février 2013, au volant d'une voiture alors qu'il fuyait une fusillade que son passager était soupçonné d'avoir orchestrée. Il a nié toute implication dans la fusillade mais a été jugé en tant que récidiviste pour cinq chefs d'accusation. En première instance, la Cour a utilisé le logiciel COMPAS pour évaluer le risque de récidive que présentait le prévenu. À l'issue de l'audience, il a été condamné à 6 ans de prison. Il a alors interjeté appel, arguant que l'utilisation de COMPAS a violé son droit à un procès équitable. L'argument n'a pas convaincu la Cour suprême du Wisconsin qui a considéré que le prévenu aurait écopé de la même peine en se basant sur les facteurs traditionnels à la suite de son crime, un délit de fuite, venant s'ajouter à un casier judiciaire déjà chargé. Cependant, la Cour semble admettre que l'usage d'un algorithme secret pour envoyer un homme en prison est discutable.

La médiatisation de cette affaire a déclenché une controverse concernant l'usage de cet algorithme par les juridictions américaines suspectées d'avaliser des biais discriminatoires. Comme l'a révélé une étude de 2016 de ProPublica, un journal

d’investigation américain, les résultats des prévisions de l’algorithme mettent en évidence un biais ethnique surestimant le risque de récidive des populations d’origine afro-américaine et sous-estimant les risques de récidive des populations d’origine caucasienne.

Les algorithmes prédictifs révèlent ainsi leurs lacunes, et il est assez logique que leur utilisation dans un domaine aussi sensible que le procès pénal nourrisse un certain nombre d’inquiétudes. Ces outils présentent surtout le risque de reproduire certains biais, auquel cas leur utilisation généralisée institutionnaliserait une certaine inégalité des justiciables devant la loi. D’une certaine façon, ces algorithmes prédictifs marquent le retour à une approche plus déterministe qu’individualiste de la criminologie.

39 La justice prédictive portera-t-elle atteinte aux droits fondamentaux des justiciables ?

Une possible algorithmisation excessive de la justice suscite la crainte d'une remise en question des principes fondamentaux du procès. Ces principes sont issus de la Convention européenne des droits de l'homme qui garantit, dans son article 6, un droit au procès équitable. Dans ce même ordre d'idée, la Charte des droits fondamentaux de l'Union européenne proclame elle aussi en son article 47, le droit de toute personne à ce que « *sa cause soit entendue équitablement, publiquement et dans un délai raisonnable par un tribunal indépendant et impartial* ».

Ces droits protègent les droits des parties et encadrent le rôle du juge. Avec la justice prédictive, des algorithmes sont censés pouvoir prédire la solution aux futurs procès. Il faut donc se demander si la justice prédictive porte directement atteinte à ces droits fondamentaux.

Plusieurs atteintes peuvent être *a priori* identifiées : l'atteinte au principe de l'égalité des armes, au principe du contradictoire et à la présomption d'innocence ainsi que l'atteinte à l'indépendance du juge.

Concernant tout d'abord le principe de l'égalité des armes, il implique l'obligation d'offrir à chaque partie une possibilité raisonnable de présenter sa cause dans des conditions qui ne la placent pas dans une situation désavantageuse par rapport à son adversaire. L'absence de communication de preuves à la défense ou le fait d'avoir un accès limité à son dossier peuvent être constitutifs d'une violation du principe de l'égalité des armes.

Ce principe doit être respecté que l'on soit face à une justice informatisée ou non, ce qui peut poser ici quelques difficultés. En effet, l'utilisation de la justice prédictive suppose le recours à des acteurs privés ayant élaboré des programmes qu'ils exploitent. Il existe ainsi un risque que ces derniers développent des outils informatiques que seuls les plus fortunés pourraient s'offrir. Plus un cabinet d'avocats ou un particulier est fortuné, plus il pourra investir dans des outils informatiques perfectionnés qui lui permettent d'obtenir des résultats favorables. Ce type de problème n'est cependant pas propre au développement de l'intelligence artificielle dans le domaine de la justice.

La question de la transparence des algorithmes est, semble-t-il, plus grave. Sans celle-ci, il n'est pas possible pour les parties de comprendre et d'avoir accès à tous les éléments ayant orienté la décision du juge. À titre d'exemple, dans l'affaire Loomis, le prévenu n'a pas pu prendre connaissance du code source de l'algorithme COMPAS et n'a eu accès qu'aux questionnaires et au résultat de l'algorithme sans en comprendre le raisonnement. Cette opacité des algorithmes peut ainsi porter atteinte au respect du principe de l'égalité des armes, mais aussi au principe du contradictoire qui en découle.

Le principe du contradictoire signifie en effet qu'une partie ne peut être correctement jugée sans avoir eu l'occasion de « contredire » toutes les prétentions de son adversaire. Il exige donc que le justiciable ait connaissance de l'utilisation d'un algorithme à l'appui des prétentions de la partie adverse ou du juge, et qu'il puisse contester le résultat produit. Plus le fonctionnement de l'algorithme sera opaque, plus il sera difficile pour le

justiciable de discuter et contester les résultats avancés, sachant que le juge conserve dans tous les cas son pouvoir d'appréciation lui permettant de ne pas accorder trop d'importance à ces résultats.

Quant au principe de la présomption d'innocence, consacrée aux articles 6.2 de la Convention européenne des droits de l'homme et 48.1 de la Charte des droits fondamentaux de l'Union européenne, il signifie que la personne poursuivie n'a pas la charge de démontrer qu'elle est innocente, et qu'il incombe à l'accusation (ministère public) de prouver la culpabilité de la personne poursuivie, soit de renverser la présomption d'innocence. Il implique aussi la règle suivant laquelle le doute profite à l'accusé et, pour la personne poursuivie, le droit de se taire et de ne pas contribuer à sa propre incrimination.

On le sait la justice prédictive remplace la formule « responsabilité/punition » par celle de « dangerosité/mesures de sûretés ». Or, accuser un individu d'un comportement futur sur la base de résultats obtenus par un algorithme est une atteinte grave au principe de la présomption d'innocence. C'est la dangerosité potentielle d'un individu qui est alors punie et non un fait qui s'est produit ; sachant que le niveau de dangerosité s'opère à partir d'un calcul mathématique de probabilité. L'on mesure les risques de l'intégration de tels outils dans le domaine de la justice pénale.

Il est difficile de quantifier l'influence exacte du résultat de l'algorithme sur la décision du juge, dans les pays ayant recours à ces outils, car cela supposerait de s'interroger sur les motivations réelles de leurs décisions. Les algorithmes prédictifs ont nécessairement une influence d'une façon ou d'une autre sur le pouvoir d'appréciation du juge, ce qui peut faire douter de son indépendance.

Rappelons que l'article 6 de la Convention européenne des droits de l'homme prévoit que le juge doit être indépendant et impartial. Ces deux notions permettent d'assurer la confiance du justiciable dans la justice et dans l'État de droit. L'indépendance du juge est une protection à l'égard des pressions extérieures (pouvoirs législatif et exécutif, groupes de pression économiques, politiques ou sociaux, etc.) tandis que l'impartialité garantit sa neutralité à l'égard des parties.

C'est précisément l'indépendance du juge qui est la plus menacée par la justice prédictive, non pas à l'égard des pouvoirs exécutif et législatif, mais à l'égard des partenaires privés, dont notamment les entreprises qui conçoivent et commercialisent les logiciels de justice prédictive, dans la mesure où leur intervention pourrait influencer les décisions judiciaires. La question mérite d'être posée puisque la plupart de ces outils ne sont pas financés par l'État mais bien par des acteurs privés.

Le juge pourrait être tenté d'accorder la même valeur aux résultats algorithmiques que celle accordée aux expertises. De ce fait, les résultats fournis par les algorithmes auront un impact sur la décision. L'effet performatif des algorithmes prédictifs risque alors de pousser le juge au conformisme. Ce dernier ne jugera plus en fonction du cas particulier qui lui est soumis mais en fonction des statistiques, ce qui entraîne la raréfaction des décisions contraires aux courants jurisprudentiels dominants. D'un point de vue juridique, il s'agit là d'une forme de dépendance à l'égard de ces acteurs privés qui conçoivent ces algorithmes. Lorsque la justice est à ce point assujettie à une influence extérieure, il est

légitime de s'interroger sur l'indépendance du pouvoir judiciaire. Aussi, il existe de grandes réserves quant à la capacité de la justice prédictive à garantir le respect de certains principes fondamentaux du procès.

40 Faut-il craindre une déshumanisation de la justice ?

Pendant longtemps, on entendait seulement parler de judiciarisation, il est devenu commun de mettre en garde aussi contre le risque d'une déshumanisation de la justice. La justice traite des droits, des torts et des réparations, et plus généralement des destins des personnes. Ce faisant, elle a une dimension intrinsèquement humaine. Les justiciables peuvent accepter le risque d'être déboutés de leurs demandes, mais ne peuvent tolérer de ne pas avoir été humainement entendus par un juge. Il est vrai qu'une meilleure écoute engendre des conséquences bénéfiques sur l'aspect cathartique de la résolution des conflits. La présence d'un magistrat qui écoute et avec lequel les justiciables se sentent écoutés « humanise » en quelque sorte la justice.

L'un des plus grands défis de l'intégration de l'intelligence artificielle dans le système judiciaire est de déterminer comment l'utiliser pour améliorer l'efficacité et la précision, tout en évitant l'automatisation des contentieux. L'idée d'utiliser l'informatique comme levier de réforme de la justice n'est pas tout à fait nouvelle. Cette technologie a déjà permis d'améliorer la vitesse de la production des décisions et de doter l'institution judiciaire de moyens de gestion et de pilotage.

Pourtant, le manque de moyens humains et matériels, ainsi que l'obsession de la performance n'épargnant pas la justice, contraint l'institution à industrialiser le traitement des affaires afin d'épurer les stocks. Ce qui conduit forcément à la dégradation des conditions de travail des magistrats. Selon une étude réalisée par le Syndicat de la magistrature, 40 % des magistrats français seraient en état de souffrance au travail. Juger vite, mais mal, ou juger bien, mais lentement et avec des retards est le dilemme devant lequel sont placés les magistrats. Ce risque de déshumanisation a été souligné dans une tribune signée par 3 000 magistrats qui visait à tirer la sonnette d'alarme, après le suicide de l'une de leurs collègues, en dénonçant les dérives d'une justice : « *qui n'écoute pas, qui raisonne uniquement en chiffres, qui chronomètre tout et comptabilise tout* » et qui « *souffre de cette logique de rationalisation qui déshumanise et tend à faire des magistrats des exécutants statistiques, là où, plus que nulle part ailleurs, il doit être question avant tout d'humanité* » (« L'appel des 3 000 magistrats et d'une centaine de greffiers », *Le Monde*, 23 novembre 2021).

Le constant n'est pas nouveau. La crise de la justice est dénoncée de longue date et ses causes tiennent en grande partie au manque de moyens de l'institution judiciaire. C'est là qu'intervient la tentation de transférer une partie du contentieux à des algorithmes.

Une décision judiciaire rendue exclusivement par une intelligence artificielle va sans doute rester longtemps un fantasme, mais l'assistance à la décision va probablement se développer. Si le juge ne cède pas son pouvoir de juger à un algorithme d'aide à la décision comme pour toute expertise, ces technologies pourraient peut-être contribuer à une amélioration de la justice. Il est donc important d'en encadrer les applications.

À ce titre, la Commission européenne pour l'efficacité de la justice (CEPEJ), qui réfléchit notamment au rôle des nouvelles technologies dans l'amélioration de l'efficacité et du fonctionnement des systèmes judiciaires européens, a consacré son plan d'action 2022-2025 à « *la digitalisation pour une meilleure justice* ». Elle y formule des

recommandations s'articulant autour de grands axes visant à ce que la justice soit toujours transparente, collaborative, humaine, centrée sur les individus et accessible, éclairée, et, enfin, responsable et réactive.

La CEPEJ insiste sur le fait que les outils numériques doivent être au service d'une « *justice humaine* », ce qui suppose de « *soutenir les juges, les procureurs, leurs équipes et les autres professionnels de la justice de manière appropriée, afin d'adapter leur rôle essentiel également à l'environnement numérique* ». Et la Commission insiste bien sur le fait que « *la digitalisation de la justice doit rendre la justice plus efficace mais ne doit jamais chercher à remplacer le juge. Le juge doit rester au centre de la procédure* ».

Cette recommandation visant à prendre en compte la dimension humaine de la justice est une démarche louable, mais reste à savoir comment ce principe sera concrètement appliqué car le risque est réel. Les algorithmes de justice prédictive ne se substitueront pas totalement au juge, mais ils pourront l'écarter de certains contentieux, en invitant les parties à recourir à des plateformes offrant des services de résolution amiable des litiges, ce qui a pour principale conséquence de déléguer la résolution amiable d'un conflit à des algorithmes. Déjudiciarisation et déshumanisation peuvent donc aller de pair et conduire à ce que les juges cèdent, au moins pour partie, la place aux algorithmes.

V.
L'INTELLIGENCE ARTIFICIELLE DANS L'ENTREPRISE

41 Une entreprise peut-elle être dirigée par une intelligence artificielle ?

Alors que le développement de l'intelligence artificielle fait craindre la disparition de certains emplois techniques ou peu qualifiés, les postes de direction dans les entreprises ne semblent pas pour autant totalement épargnés.

Pourtant, une étude publiée, en novembre 2023, par le groupe technologique américain Cisco révèle que seulement 4 % des entreprises françaises sont pleinement préparées à intégrer l'intelligence artificielle alors que la moyenne pour les entreprises européenne est de 8 % ; les pays en tête de liste étant la Suède (22 %) et le Royaume-Uni (10 %). Si les entreprises françaises ne sont pas, pour le moment, préparées à cette évolution, toutes les entreprises, quelle que soit leur taille ou leur secteur d'activité, devront tôt ou tard intégrer l'intelligence artificielle dans leur organisation. Elles peuvent être même intégrées dans les processus décisionnels.

La presse relate pourtant déjà le cas d'entreprises ayant confié leur direction à une intelligence artificielle. C'est, par exemple, le cas de *Dictador*, une entreprise polonaise présente dans 60 pays, qui a révélé en 2023 avoir mis à sa direction, à titre expérimental, une IA nommée Mika. En 2022, c'est une autre IA qui serait devenue PDG de la société chinoise de jeux vidéo *NetDragon Websoft*. Dans ce même pays, la société *Deep Knowledge Venture*, située à Hong Kong, avait annoncé en 2014 la nomination d'un programme d'apprentissage-machine, dénommé « Vital », comme membre de son conseil d'administration. Il s'agissait en réalité d'un outil d'aide à la décision, dont l'objet consistait à faire des recommandations en matière d'investissement et n'avait pas la qualité de membre à part entière de l'organe social, car le droit hong-kongais ne le permet pas, mais simplement le statut d'observateur.

De nombreux pays réservent, à ce titre, les fonctions d'administrateur à des personnes physiques, rendant ainsi impossible le remplacement d'un administrateur par un algorithme. C'est notamment le cas du droit français qui n'autorise la nomination au conseil d'administration que de personnes physiques ou de personnes morales, à la condition que ces dernières désignent un représentant permanent encourant les mêmes responsabilités que s'il était administrateur en son nom propre (art. L. 225-20 du Code de commerce). Afin de prétendre à une place au conseil, il conviendrait que l'IA soit ainsi dotée de la personnalité juridique, hypothèse envisagée mais n'ayant reçu aucune consécration.

Au-delà de ces effets d'annonce, ces expérimentations font surtout miroiter aux entreprises les potentialités de l'intelligence artificielle. Par leur capacité à gérer et à analyser des masses de données, les algorithmes peuvent parvenir à des décisions plus éclairées, voire plus rationnelles puisqu'elles ne seraient plus dépendantes de la subjectivité de l'appréciation humaine. Cette perspective théorique suscite des questions juridiques complexes, notamment concernant la notion de responsabilité du dirigeant.

Il convient de rappeler que le dirigeant d'une société est son représentant légal. Il a pour mission de représenter la société tant dans les rapports internes que dans les rapports externes avec les tiers. Dans l'exercice de ses fonctions, il détient des pouvoirs étendus

pour agir dans l'intérêt de la société.

Rien n'interdit à ce que l'intelligence artificielle vienne en soutien à ses décisions en analysant une importante quantité de données et en mettant en évidence des tendances et des opportunités qu'il n'aurait peut-être pas envisagées autrement. Elle pourrait, par exemple, être utilisée pour prendre des décisions relatives aux projets d'investissements. Elle peut aussi l'aider à gérer les risques plus efficacement en prédisant les risques financiers ou en détectant les fraudes. Les algorithmes peuvent en effet surveiller les opérations financières et détecter les comportements frauduleux ou non conformes.

L'utilisation d'un système d'IA ne devrait pas en principe accroître sa responsabilité ni lui permettre de s'exonérer de toute responsabilité. En tout état de cause et afin d'engager la responsabilité d'un dirigeant, il faut prouver qu'il a commis une faute de gestion, et que de celle-ci a découlé un préjudice pour la société ou les associés.

Une faute de gestion pourra-t-elle pour autant être caractérisée si le dirigeant parvient à démontrer qu'il a été induit en erreur par les informations erronées fournies par le système d'IA censée faciliter sa prise de décision ? Il semble peu probable que le dirigeant puisse échapper à sa responsabilité en invoquant, par exemple, les résultats biaisés ou erronés du système d'IA.

La question de la responsabilité des organes de direction d'une société pour les actions et les décisions prises par les systèmes d'IA reste une question épineuse. L'IA est, comme toute activité, génératrice de risques : la machine peut se tromper et causer un dommage. Il est vrai que la Commission européenne a annoncé en septembre 2022 se saisir de cette question, avec la publication de deux propositions de directives ayant pour objet de moderniser les règles applicables en matière de responsabilité pour tenir compte des dommages liés à l'IA, mais la question de la responsabilité des dirigeants employant ces outils n'y est pas directement abordée (cf. question n° 100). En l'absence de solutions claires, la prudence est donc de mise pour les directions qui utilisent des systèmes d'intelligence artificielle dans la gestion de leur société.

42 Quel pourrait être l'impact de l'IA sur la transition énergétique ?

La transformation numérique dans le contexte de la transition énergétique fait face à plusieurs défis. L'intelligence artificielle est en principe énergivore mais elle pourrait aussi être utile à la transition énergétique.

Grâce aux performances de l'apprentissage automatique et à des capacités de traitement massif de données, l'IA ouvre de nouvelles perspectives à la transition énergétique et à l'optimisation de la consommation des ressources. Elle permet une meilleure prédiction de la demande énergétique et facilite l'intégration des sources d'énergie renouvelables, contribuant ainsi à la transition énergétique. Elle joue également un rôle crucial dans la réduction de l'empreinte carbone et dans la réalisation des objectifs de développement durable.

Les producteurs d'énergie utilisent depuis longtemps des modèles prédictifs destinés à ajuster la production d'énergie en fonction de la demande. Il est donc nécessaire d'être capable de prédire en amont les besoins et de s'adapter le plus rapidement possible pour éviter la perte ou le manque d'énergie. Anticiper la production et la demande, essayer de les faire coïncider au mieux est un véritable enjeu. L'électricité ne pouvant être stockée en grande quantité, la production d'énergie se fait généralement en temps réel, selon le niveau de la demande, avec des adaptations nécessaires. La consommation d'énergie peut d'ailleurs évoluer en fonction de nombreux facteurs : les caractéristiques des bâtiments et les propriétés thermiques des matériaux ; saisonniers et météorologiques ; le comportement des utilisateurs, etc. Il existe ainsi un risque constant de sous-production, et donc de coupure de courant ; ou un risque d'excédent de production et donc d'énergie non utilisée et perdue. Il faut donc que les flux d'énergie soient contrôlés afin d'optimiser la consommation d'énergie. Cela n'est possible que par l'analyse d'un grand nombre de données différentes.

Les systèmes d'IA sont, à ce titre, principalement utilisés pour prédire la consommation ou la demande d'énergie électrique ou l'énergie produite par une ressource d'énergie renouvelable, telle que des éoliennes ou des panneaux photovoltaïques. Les réseaux électriques sont d'ailleurs de plus en plus complexes et intègrent de plus en plus de centrales de production d'énergies renouvelables, dont les niveaux de production sont moins prévisibles. À mesure que la production s'oriente davantage vers les énergies renouvelables, l'offre d'électricité deviendra plus fragmentée, diffuse et moins prévisible. Les énergies éoliennes et solaires ne sont en effet pas toujours disponibles, car sensibles aux heures de la journée et aux conditions météorologiques. Déjà utilisée dans de nombreux domaines liés à la production d'énergie renouvelable et à la gestion de systèmes énergétiques, l'IA s'installe au centre de l'équilibre entre consommation et production d'électricité, ainsi que de la gestion des réseaux intelligents, appelés aussi *smart grids*. On désigne par *smart grid* un réseau d'énergie qui intègre des technologies de l'information et de la communication, ce qui contribue à améliorer son exploitation et à développer de nouveaux usages, tels que le véhicule électrique ou le stockage.

À la couche physique pour le transit d'énergie des réseaux se superpose désormais une couche numérique qui joue un rôle de plus en plus important pour son pilotage. Au sein d'un *smart grid*, la donnée circule non pas à sens unique mais dans les deux sens de façon dynamique, grâce à des capteurs connectés, répartis sur le réseau et dans les bâtiments ainsi que les habitations. Le compteur Linky est un exemple de *smart grid*. C'est d'ailleurs grâce aux données de fourniture d'électricité collectées par ce compteur (coupures de courant, surtensions, etc.) que CartoLine BT, un outil expérimental s'appuyant sur une intelligence artificielle, peut identifier de manière autonome des situations pouvant générer un incident sur le réseau et émettre des recommandations d'intervention aux équipes techniques d'*Enedis*. Il a été démontré que plus de 50 % des recommandations d'intervention signalées comme prioritaires par CartoLine BT donnent lieu à une panne dans les 15 jours qui suivent, et plus de 95 % des suspicions d'anomalies remontées par l'outil se sont avérées être de véritables anomalies sur le terrain. L'intérêt de cet outil est de pouvoir prédire une panne sur le réseau public de distribution d'électricité basse tension (230 ou 400 volts) avant même qu'elle ne se produise.

Utilisée dans la production ou la consommation d'énergie, l'IA fonctionne grâce à des capteurs reliés aux systèmes de contrôle. Ces capteurs permettent d'analyser de nombreuses données, par exemple, la température, les vibrations ou encore la puissance des flux. Le traitement en temps réel de données permet de repérer rapidement les anomalies et les dysfonctionnements afin de les traiter.

L'IA peut aussi contribuer à rendre les bâtiments plus économes en énergie en analysant les données de consommation. Elle peut suggérer des ajustements automatiques du chauffage et de la climatisation, ou encore optimiser l'éclairage pour réduire la consommation d'électricité.

La transition écologique impose de traiter des volumes considérables de données, ce que l'IA permet de réaliser, mais il faudra veiller à ne pas augmenter la consommation d'énergie des data centers. L'IA implique en effet une grande consommation d'énergie : puissance de calcul, stockage, extraction de matériaux pour produire des composants, etc. Il est donc nécessaire d'inclure ces technologies dans une démarche de développement durable. Au-delà de cet aspect, il conviendra également de pouvoir disposer de données de qualité nécessaires à son fonctionnement et d'interpréter les décisions prises par les algorithmes.

43 L'IA générative a-t-elle sa place dans l'entreprise ?

Le très médiatique ChatGPT, mais aussi MidJourney ou DALL-E ont fait découvrir au grand public ce que pouvait être une IA générative.

Selon ChatGPT, l'IA générative est : « *une branche de l'intelligence artificielle qui consiste à développer des systèmes capables de créer de nouvelles données ou contenus de manière autonome. Ces systèmes utilisent des algorithmes pour générer des informations, des images, des sons ou d'autres types de contenus à partir de modèles préexistants et sont souvent utilisés dans des applications telles que la création d'œuvres artistiques, la génération de texte ou la création de nouveaux designs* ». Autrement dit, une IA générative est une forme d'intelligence artificielle qui est capable de créer de nouveaux contenus de manière autonome en extrapolant à partir de ses données d'entraînement. Elle est utilisée dans de nombreux domaines tels que la création artistique, la musique, la vidéo, la publicité, la mode, les jeux vidéo et la simulation.

Elle se distingue d'une l'IA classique dans la mesure où elle crée un contenu nouveau et original qui ressemble à ce qu'on peut trouver dans ses données d'entraînement, mais qui ne sont pas semblables, alors que l'IA classique est axée sur des tâches spécifiques, de manière non interactive, comme la détection des fraudes par carte de crédit, la recherche d'itinéraires, etc.

L'IA générative fonctionne en utilisant des modèles d'apprentissage automatique (*machine learning* mais surtout le *deep learning*) pour créer du contenu de manière autonome. Parmi les techniques d'apprentissage automatique les plus couramment utilisées, nous pouvons citer les réseaux antagonistes génératifs ou *Generative Adversarial Network* (GAN) qui reposent sur la combinaison de deux réseaux de neurones dits concurrents : l'un appelé générateur qui crée une image et la transmet à un discriminateur qui détermine si l'image est réelle ou synthétique. Le générateur apprend à générer des données de plus en plus réalistes pour tromper le discriminateur, tandis que le discriminateur s'améliore pour distinguer les données générées des données réelles. Grâce à cette concurrence, les deux modèles s'améliorent simultanément au fil de l'entraînement. C'est en particulier ce type d'algorithme qui est utilisé pour réaliser les *deepfakes*, contraction de *deep learning* et de *fake*, qui sont des enregistrements vidéo ou audio truqués et modifiés grâce à l'intelligence artificielle.

Les IA génératives de texte s'appuient, quant à elles, sur des LLMs (*Large Langage Models*) qui sont des programmes conçus pour interagir avec le langage humain et entraînés sur un large corpus de textes qui servent d'exemples à l'algorithme. GPT-3 (*Generative Pre-Trained Transformer 3*), développé par *OpenAI*, est l'un des exemples les plus connus de LLM. Il a été conçu pour accomplir une variété de tâches liées au langage, comme la traduction, la réponse aux questions, etc.

On parle enfin d'IA générative multimodale (texte, parole, son et image) lorsque l'IA est multitâche et multilingue et pouvant produire des contenus afin de répondre répondre aux utilisateurs d'une application, soit avec du texte, soit avec des images, soit les deux. Ainsi, ChatGPT, depuis la version GPT-4 Turbo, est en mesure de lire et générer aussi bien du texte que des images. Il en est de même pour Bard de Google, avec la version Gemini.

Parce qu'elles permettent de gagner du temps, l'IA générative, incarnée par des outils comme ChatGPT, va transformer de nombreux métiers au sein des entreprises. Ces dernières ont d'ailleurs consacré une grande partie de l'année 2023 à la tester. Il est vrai qu'il existe beaucoup de pressions pour adopter cette nouvelle génération de technologie, du fait du risque d'être laissé pour compte.

De nombreuses entreprises l'utilisent déjà pour accélérer la rédaction de brouillons de documents ou pour résumer des documents volumineux. En matière de communication et de marketing, des entreprises conçoivent déjà leur campagne publicitaire en s'appuyant sur l'IA générative. Elle peut aussi apporter une aide précieuse aux développeurs en automatisant la génération de code, Pour ce faire, il suffit de renseigner une problématique sous forme de texte, en langage naturel, et l'outil peut produire plusieurs centaines de lignes de codes. Elle peut enfin être utilisée dans le domaine de la finance pour prédire les tendances des marchés boursiers, les performances des investissements, ou encore générer des stratégies d'investissement automatisées, soit des systèmes qui prennent des décisions d'investissement en fonction de l'analyse des données et des tendances du marché.

Aucun secteur ne semble totalement épargné par le potentiel à court et moyen terme des IA génératives, que ce soit l'éducation, les transports, la santé, le droit, l'assurance, l'énergie, etc. Cette technologie peut effectivement permettre de réduire drastiquement le temps nécessaire à la réalisation des tâches, mais elle soulève de nombreuses questions éthiques et juridiques.

La principale question éthique concerne le rapport à la vérité, notamment la fiabilité des réponses données, puisque ce type d'IA ne peut pas évaluer la véracité des réponses qu'il donne. Par conséquent, ces systèmes peuvent produire des réponses erronées ou des phrases qui énoncent des faits n'existant pas dans le monde réel, ce qui peut mener à la production de désinformation. On parle alors d'« hallucinations ». Pour certains usages, par exemple produire une fiction, cette absence de véracité peut n'avoir aucune incidence, mais elle peut avoir des conséquences désastreuses si les réponses sont des recommandations pour des décisions importantes.

L'utilisation de l'IA générative soulève également des questions juridiques majeures, en particulier en ce qui concerne la propriété intellectuelle. La propriété des œuvres générées par l'IA, telles que la musique, les images et le texte, qui peuvent être difficiles à distinguer de celles créées par des humains est une question complexe. Dans certains cas, l'IA peut même utiliser une œuvre de l'esprit existante sans l'autorisation des auteurs, ce qui pose des problèmes quant au respect des droits d'auteur (cf. question n ° 84). Il est ainsi fondamental de comprendre les enjeux techniques, éthiques et de juridiques liés à cette technologie en constante évolution afin d'en assurer le meilleur encadrement possible.

44 Quels sont les enjeux de l'utilisation de l'IA dans l'industrie ?

Qui s'intéresse à l'industrie a nécessairement entendu parler de la notion d'« industrie 4.0 », appelée aussi l'industrie du futur, de plus en plus souvent utilisée pour évoquer les transformations technologiques que connaît et connaîtra le secteur industriel. Cette notion est en lien avec celle de quatrième révolution industrielle, qui serait à l'œuvre actuellement. Cette dernière succède à la première révolution industrielle du XVIII^{e} siècle qui voit l'apparition des machines à vapeur et le début de la mécanisation ; à la deuxième révolution industrielle de la fin du XIX^{e} siècle et à ses chaînes de montage permettant une production de masse ; et enfin, à la troisième révolution industrielle qui se distingue par l'automatisation des processus de production grâce au développement de l'électronique et de l'informatique.

Bien qu'elle désigne des évolutions très concrètes, cette notion reste malgré tout encore floue, y compris pour les industriels. Elle apparaît, pour la première fois, en 2011 en Allemagne à la foire industrielle d'Hanovre. Ses promoteurs sont partis du constat que l'Allemagne était en retard en matière de digitalisation face à ses principaux concurrents américain et asiatique. Ils considèrent que l'industrie manufacturière allemande est particulièrement bien placée pour profiter de cette digitalisation. Ce qui semblait être au départ une opération marketing des fournisseurs d'équipements industriels est devenu une politique industrielle promue par les pouvoirs publics allemands, et en quelques années, cette politique a été étendue à toutes les nations industrielles.

Ce concept symbolise désormais l'entrée de l'industrie mondiale dans une nouvelle ère qui combine trois innovations technologiques : l'automatisation, l'internet des objets et l'intelligence artificielle.

L'industrie 4.0 repose, à ce titre, sur plusieurs technologiques fondamentales :

- Le *big data* qui est collecté à partir de sources variées (les équipements d'usine, les dispositifs de l'internet des objets, etc.) et l'analytique basée sur l'IA qui est un domaine de l'informatique permettant d'identifier des tendances pertinentes dans les données ;
- L'internet industriel des objets (*Internet of Things* ou IoT) : il s'agit d'une technologie qui permet aux objets connectés de communiquer entre eux et avec d'autres systèmes informatiques. Les appareils, les robots, les machines, les équipements utilisent des capteurs et des étiquettes RFID pour fournir des données en temps réel sur leur état, leurs performances ou leur emplacement. Ces derniers permettent d'offrir une surveillance en temps réel pour optimiser la production industrielle ;
- L'intégration horizontale qui désigne la mise en réseau de machines et de systèmes au sein d'une ligne de fabrication. L'intégration verticale du système représente, quant à elle, le processus de connexion de tous les niveaux de production allant de l'atelier de fabrication au département des ventes d'une entreprise ;
- La réalité augmentée est une technologie émergente mais ayant des implications majeures pour la maintenance de machines complexes grâce à des lunettes intelligentes pour visualiser les instructions de réparation ou de montage, par exemple ;

- Les robots autonomes qui sont capables d'effectuer un grand nombre de tâches avec un minimum d'intervention humaine. Les cobots (robots collaboratifs) sont aussi conçus pour interagir dans des environnements où ils doivent collaborer avec les humains.
- La simulation ou le jumeau numérique est une simulation virtuelle d'un système, d'une machine, d'un produit du monde réel, basée sur les données des capteurs IoT, offrant la possibilité de disposer d'une copie numérique d'un objet réel pouvant être testé et manipulé numériquement. Ces simulations permettent, par exemple, de former les salariés et de superviser la production ;
- Le *cloud computing*, appelé aussi informatique en nuage, est l'un des piliers de l'industrie 4.0. Il permet de stocker, de traiter et de gérer des données à distance en utilisant des serveurs distants ;
- La fabrique addictive ou l'impression 3D permet, par exemple, de produire des articles uniques en interne et à moindre coût ou de les stocker sous forme de fichiers de conception dans des inventaires virtuels et les imprimer à la demande ;
- La cybersécurité qui est fondamentale avec une connectivité accrue. Les cyberattaques peuvent prendre différentes formes : les logiciels malveillants, l'hameçonnage, etc.

À travers l'industrie 4.0, on passe d'une logique de production de masse à celle d'une production flexible, à la demande et localisée. Il s'agit ainsi d'un changement complet de logique économique pour chaque entreprise.

Il est évident que l'intelligence artificielle joue un rôle important dans cette transformation puisqu'elle aboutit à créer des usines dites « intelligentes ». Combinée aux technologies déjà évoquées, elle offre la possibilité d'automatiser les contrôles de qualité, de réduire les erreurs humaines et de permettre une traçabilité à chaque étape de la production en temps réel.

L'émergence de l'internet dans les années 1990 marque le début de la numérisation de l'économie. Aujourd'hui, l'intelligence artificielle et la robotique sont susceptibles d'accentuer le processus d'automatisation. C'est la raison pour laquelle l'industrie 4.0 inquiète beaucoup en raison de son impact réel sur le nombre d'emplois menacés par l'automatisation, car il n'est pas certain que la création de nouveaux emplois permette de compenser les suppressions. Même si les estimations sont encore incertaines voire alarmantes, on peut considérer que les emplois nécessitant des compétences simples et répétitives sont les plus exposés à ce risque d'automatisation, ce qui pourrait entraîner des pertes d'emplois massives dans certains secteurs. Dans tous ces cas, de nombreux emplois risquent de subir de profondes transformations qu'il est encore difficile à évaluer à l'heure actuelle.

45 Quels sont les usages de l'IA dans le secteur bancaire et financier ?

Après la crise de 2007-2011, les banques et les assurances ont porté leurs efforts sur le renforcement de la gestion des risques, encouragées en cela par les évolutions réglementaires et le renforcement de la supervision financière. L'ensemble du secteur bancaire et financier connaît une multiplication sans précédent des normes. Les exigences en matière de conformité sont de plus en plus élevées pour les entreprises. Les banques et le secteur de la finance font d'ailleurs face à un risque de sanction de la part du régulateur. Cette inflation réglementaire nécessite en principe un accroissement des effectifs afin de se conformer à la réglementation. Étant un secteur très concurrentiel, la tendance est davantage à la réduction des coûts qu'à leur augmentation. Nul ne sera donc surpris de voir la banque et la finance parmi les secteurs les plus avancés et demandeurs de solutions innovantes.

L'évolution permanente des méthodes de blanchiment des capitaux et de financement du terrorisme au gré des progrès technologiques contraint les établissements financiers à rechercher des technologies adaptées pour lutter contre la criminalité financière ainsi que des outils permettant de s'adapter à l'évolution de la réglementation.

Après les Fintech, terme qui associe finance et technologie et désigne les start-ups de la finance mais aussi l'Insurtech, terme renvoyant à « *Insurance* » et « *Technology* » et désignant les start-ups de l'assurance, est apparue la Regtech, contraction des mots « *Regulation* » et « *Technology* ». Le terme « Regtech » (technologies réglementaires) a été inventé en 2015 par l'Autorité de bonne conduite financière britannique (FCA). Une des définitions les plus répandues des *Regtech* est celle fournie par la FCA qui définit cette notion comme « *un sous-ensemble de la Fintech qui concerne les technologies capables de rendre le respect des obligations réglementaires plus efficace et performant qu'avec les dispositifs existants* ». Autrement dit, les *Regtech* mettent la technologie au service de la mise en conformité. Le développement des *RegTech* est un phénomène récent puisque la majorité d'entre elles ont été créées après 2008, ce qui correspond au renforcement des règles prudentielles adoptées à la suite de la crise économique et financière. Elles sont nées aux États-Unis et se développent en Europe, notamment en Grande-Bretagne.

Concrètement, les *Regtech* s'avèrent particulièrement utiles, notamment dans la lutte contre la fraude ou le blanchiment grâce à la gestion de l'identité et de la connaissance du client (*Know Your Customer* ou KYC), la gestion de la conformité, la protection des données, le suivi des transactions bancaires, la veille et le reporting réglementaire pour superviser les sociétés financières.

Pour ce faire, elles s'appuient sur le *big data,* le *machine learning* mais aussi la biométrique, la cryptographie et la *blockchain* pour aider les services de conformité à remplir leurs différentes missions (pilotage de la veille réglementaire, transposition des normes en outils et procédures, cartographie des risques, etc.) et traiter les énormes masses de données. Dans les faits, les solutions de *Regtech* qui intègrent l'intelligence

artificielle peuvent repérer des caractéristiques suspectes dans le comportement des clients qu'une supervision classique pourrait ne pas détecter, voire anticiper des changements dans le profil de risque d'un client.

Ces technologies sont certes prometteuses mais elles sont aussi génératrices de risques, notamment en raison de l'externalisation de leur activité de conformité. En effet, les établissements financiers risquent de perdre la maîtrise de leurs données qu'ils confient aux *Regtech*. Ce risque est d'autant plus accru si les serveurs sont situés hors de l'Union européenne, en particulier aux États-Unis, pays dans lequel il est parfois difficile de faire respecter le RGPD.

Cette externalisation de la gestion de la conformité peut aussi engendrer une perte de savoir-faire pour les établissements bancaires et financiers. L'établissement de liens toujours plus étroits avec le secteur des technologies représente un risque en outre pour les activités bancaires et financières qui pourraient devenir dépendantes des entités produisant ces outils. Enfin, l'utilisation de l'intelligence artificielle permet d'automatiser des tâches répétitives, mais augmente le volume des interactions au sein des systèmes d'information. Cette automatisation décuple ainsi le nombre de failles potentielles exploitables par des cybercriminels. En d'autres termes, l'intelligence artificielle n'ouvre pas de nouvelles failles, sauf défaut dans la conception, mais pourrait accentuer des failles préexistantes.

Au-delà de la question de la conformité, l'intelligence artificielle peut aussi permettre aux banques d'améliorer le service client. Le conseiller client ainsi que le client lui-même peuvent en effet utiliser divers outils, tels que les logiciels permettant le tri automatique des courriels, les *chatbots* (envoi et réception de messages écrits), les *voicebots* (envoi et réception de messages vocaux) ou encore les robots-conseillers (*robo-advisor)* en matière d'épargne. Le Crédit Mutuel utilise, par exemple, depuis 2017 une IA préparant des réponses aux courriels envoyés par les clients. Le conseiller, n'a plus alors qu'à améliorer et valider cette réponse. L'automatisation de ces tâches lui permet de libérer du temps pour d'autres activités.

L'intelligence artificielle permet également d'améliorer la connaissance du client et peut ainsi être utile pour déterminer si le demandeur de crédit est digne de confiance ou pas. L'évaluation du risque de crédit est en effet une activité centrale pour tout établissement de crédit. Ce dernier peut utiliser un outil, appelé le *crédit scoring*, pour prévoir le risque de non-remboursement du prêt. En tirant profit des informations sur le client, le *scoring* (notation) permet de compléter l'approche traditionnelle, qui utilise des données financières limitées, par une approche exploitant le *big data* faisant appel à des données non financières (par exemple, les données en lien avec les comportements de leurs clients).

Or, la performance de l'intelligence artificielle est largement dépendante de la qualité des données et de l'absence de biais dans leur traitement. Ces biais peuvent, à ce titre, être renforcés par l'algorithme et aboutir à des discriminations. L'utilisation du *scoring* n'est pas interdite en soi, mais elle pourra être qualifiée de système d'IA à haut risque, ce qui imposera l'application d'obligations spécifiques pour l'encadrer, conformément à l'IA Act (cf. question n° 96).

L’intelligence artificielle peut enfin être utilisée dans le domaine de la finance, notamment dans le *trading* algorithmique et la gestion des investissements. Le *trading* algorithmique consiste à utiliser des systèmes d’intelligence artificielle afin de prendre des décisions ou de réaliser des actions de *trading* sur les marchés financiers. Ils sont programmés pour des stratégies bien déterminées. Par exemple, un algorithme peut être programmé pour acheter une action dès que son prix atteint un niveau prédéfini considéré comme bas, et la vendre dès qu’il atteint un niveau considéré comme élevé. L’utilisation de l’intelligence artificielle dans le *trading* présente cependant des risques potentiels. L’un des principaux défis concerne la complexité des systèmes d’IA utilisés pouvant rendre difficile la compréhension de leurs prises de décision, ce qui peut être préoccupant pour les investisseurs souhaitant une certaine transparence. Les systèmes d’IA proposés doivent donc éviter une complexité excessive et garantir un certain niveau de contrôle humain.

46 L'IA et la *blockchain* peuvent-elles se combiner ?

La *blockchain* ou chaîne de blocs, née avec le *Bitcoin* à la suite de la crise mondiale de 2008, est une technologie de stockage et de transmission d'informations. Concrètement, une chaîne de blocs est une base de données qui contient l'historique de tous les échanges effectués entre ses utilisateurs depuis sa création. Chaque bloc est enchaîné au bloc précédent dans une séquence et est enregistré de manière immuable sur un réseau pair-à-pair. Les utilisateurs sont libres d'ajouter des informations à tout moment ou de vérifier des transactions au moyen d'un système cryptographique. Les données sont chronologiquement cohérentes, car il n'est pas possible de supprimer ou modifier la chaîne sans le consensus du réseau.

La *blockchain* peut être ainsi comparée à un grand livre de comptes public et immuable dans lequel chaque mot représente une transaction qui doit être vérifiée et approuvée avant de pouvoir être écrite sur la page. Chaque nouveau bloc s'ajoute à la chaîne de manière séquentielle et irréversible, similaire à la façon dont chaque page d'un livre suit la précédente. C'est ce caractère infalsifiable qui fait le succès de la *blockchain*, utilisée principalement pour effectuer des transactions sécurisées. Les applications de cette technologie sont très variées, allant de la cryptomonnaie aux systèmes de vote numériques, en passant par le transfert de flux financier. La *blockchain* permet également d'automatiser en toute sécurité certaines étapes dans un contrat ou en dehors de tout contrat. Il s'agit du *smart contract* qui est un programme informatique qui automatise certains faits ou certains actes. Ce n'est donc pas un contrat au sens juridique, mais un programme d'exécution automatique d'un contrat préalable.

Contrairement à une base de données classique, la spécificité de la *blockchain* réside dans son architecture qui est décentralisée. En effet, elle n'est pas hébergée par un unique serveur, mais fonctionne d'individu à individu, sans intermédiaire. Ils n'ont pas besoin d'un réseau central pour communiquer puisqu'ils communiquent directement entre eux. Elle évolue ainsi dans un environnement en peer-to-peer (pair à pair), où les utilisateurs sont à la fois clients et serveurs, à la fois émetteurs et récepteurs de données. Elle évolue aussi de manière autonome, sans réel organisme de contrôle. Au lieu d'avoir une banque qui contrôle et valide les transactions, ce rôle est réparti sur un réseau d'ordinateurs, appelés validateurs ou nœuds. Chaque fois qu'une transaction est effectuée, elle doit être approuvée par ces ordinateurs selon des règles préétablies, ce qui assure la sécurité et l'intégrité des données.

L'intelligence artificielle et la *blockchain* sont des technologies voisines qui peuvent se combiner pour créer des systèmes plus sûrs, fiables et transparents. L'intelligence artificielle peut améliorer le fonctionnement des *blockchains* en utilisant au mieux les données collectées et manipulées par celles-ci, en traitant d'énormes quantités de données et en automatisant certaines tâches. Elle peut, par exemple, optimiser la consommation d'énergie nécessaire au minage. L'intégration de l'intelligence artificielle à la *blockchain* peut aussi bénéficier aux *smart contracts* qui pourront assurer des tâches plus complexes. Par exemple, s'il est possible à l'heure actuelle de prévoir qu'un *smart contract* procède à un paiement si une marchandise est livrée, il deviendra possible

d'ajouter des conditions supplémentaires qui seront appréciées par la machine. De manière générale, l'IA peut contribuer à améliorer la sécurité des systèmes de *blockchain* en détectant les tentatives d'intrusion et les comportements suspects. À ce titre, elle peut aussi être utilisée pour améliorer l'efficacité et la sécurité du processus de validation des transactions en détectant les fraudes, les erreurs ou les transactions non autorisées sur la *blockchain* bien plus rapidement que les méthodes traditionnelles. Par ailleurs, l'IA peut aussi être utilisée pour protéger les clés privées des utilisateurs servant à signer les transactions sur la *blockchain,* et qui peuvent donc être la cible des pirates informatiques.

De son côté, la *blockchain* pourrait aussi introduire plus de décentralisation dans les outils d'intelligence artificielle, qui sont aujourd'hui développés sur d'immenses bases de données gérées de manière centralisée. Elle parviendrait ainsi à démocratiser l'intelligence artificielle en permettant à des acteurs nouveaux d'entrer sur le marché. Par ailleurs, la traçabilité des données propre à la *blockchain* pourrait aider à mieux expliquer le fonctionnement des algorithmes et de réduire l'effet « boîte noire » qui est souvent pointé du doigt.

Si la combinaison de ces deux technologies présente des avantages, il existe aussi de nombreuses limites qu'il convient de ne pas oublier. En premier lieu, la plupart des systèmes d'IA sont basés sur l'apprentissage automatique. Ils ont ainsi besoin d'une grande quantité de données pour apprendre, faire des prédictions précises ou prendre de bonnes décisions. Or, la *blockchain* ne contiendra jamais suffisamment de données pour qu'ils puissent fonctionner efficacement. Par ailleurs, les algorithmes peuvent produire des résultats qui sont difficiles à expliquer, ce qui peut nuire à la confiance dans la *blockchain*. Enfin, ces mêmes algorithmes peuvent aussi avoir été entraînés sur des données d'entraînement biaisées ou incomplètes, ce qui risque d'affecter les résultats, voire fragiliser la sécurité même de la *blockchain*.

La combinaison de ces technologies offre malgré tout de nouvelles perspectives, certains y voient un moyen de développer le Web3 ou Web 3.0 qui serait le troisième âge de l'histoire de l'internet, successeur du Web 1.0, qui a régné du début des années 1990 jusqu'au milieu des années 2000, et du Web 2.0 existant depuis la décennie 2010. Ce terme fut inventé par Gavin Wood en 2014 qui a notamment contribué au développement de l'*Ethereum*, considéré comme la seconde cryptomonnaie la plus importante après le *Bitcoin*. Les défenseurs du Web3 soutiennent que les plateformes en ligne sont aujourd'hui centralisées et contrôlées par quelques multinationales, comme *Amazon*, *Apple* ou *Meta*. L'idée serait donc de créer un web décentralisé en supprimant les intermédiaires, notamment les GAFAM.

Le Web3 est basé sur la technologie *blockchain* qui offre un accès décentralisé à des applications et des services en ligne. Cela signifie que les utilisateurs peuvent accéder à des informations et des services sans passer par des serveurs centralisés. Il englobe également l'intelligence artificielle pouvant traiter, analyser et interpréter les importantes quantités de données produites par la *blockchain*. Pour l'heure, le Web3 intéresse principalement la communauté des cryptomonnaies, mais il n'est pas exclu qu'avec l'intelligence artificielle l'internet connaisse bien une nouvelle ère.

47 L’intelligence artificielle a-t-elle un impact sur la publicité ?

Le monde de la publicité digitale qui dispose et exploite une quantité considérable de données, tant contextuelles que personnelles bénéficie, comme tant d’autres, du développement de l’intelligence artificielle au cours de ces dernières années. Avec l’avènement de l’internet et des réseaux sociaux, chaque individu est devenu une cible potentielle pour les publicitaires.

Avec l’intelligence artificielle, la publicité devient plus précise et personnalisée. En effet, il devient possible d’analyser les données de performance des publicités, telles que le taux de clics, afin d’aider les entreprises à viser plus efficacement leur public cible. L’intelligence artificielle peut surtout être utilisée pour analyser les données de l’utilisateur, comme ses achats en ligne ou ses centres d’intérêts, afin de lui proposer des publicités personnalisées correspondant à ses intérêts. La publicité personnalisée contribue ainsi de manière significative à la rentabilité de l’entreprise. Pour améliorer l’efficacité des campagnes publicitaires, il faut en effet que les annonceurs proposent au public cible un contenu qui s’aligne sur ses centres d’intérêts.

L’intelligence artificielle permet également aux annonceurs d’automatiser leurs campagnes publicitaires tout en améliorant le placement des annonces. Cette automatisation constitue un gain de temps pour ces derniers, car elle leur permet de gérer simultanément plusieurs campagnes publicitaires sans augmenter leur effectif. Les algorithmes d’IA peuvent aussi analyser les données de performance des publicités en temps réel pour optimiser les performances des campagnes. Par exemple, si une campagne publicitaire n’atteint pas ses objectifs avec un public cible, ils peuvent rapidement l’identifier et ajuster les paramètres de ciblage pour se concentrer sur d’autres profils plus susceptibles d’interagir avec les publicités. Cette optimisation en temps réel permet aux annonceurs d’adapter rapidement et en permanence leurs stratégies de ciblage publicitaire pour obtenir de meilleurs résultats.

En plus de l’automatisation et de l’optimisation en temps réel, l’IA peut aussi être utilisée pour l’analyse prédictive afin d’identifier des modèles de comportement des utilisateurs et ainsi permettre aux annonceurs de mieux cibler leur public. Les entreprises peuvent ainsi améliorer la planification de leurs campagnes en atténuant les changements de comportement des utilisateurs, ce qui permet d’élaborer des stratégies de ciblage publicitaire plus efficaces. Grâce à ces prédictions basées sur l’analyse de divers ensembles de données, les entreprises peuvent ainsi identifier les utilisateurs susceptibles d’effectuer un achat prochainement et leur proposer des publicités adaptées à leur profil.

Outre ces aspects, l’IA est bien évidemment aussi utilisée pour créer du contenu publicitaire. Elle peut produire des textes, des images, et même des vidéos personnalisés à grande échelle. *Meta* (anciennement *Facebook*) a, par exemple, lancé en 2023 un nouvel outil, *AI Sandbox*, basé sur l’IA générative et permettant aux annonceurs de créer des textes publicitaires, de fournir des arrière-plans pour les images ou de recadrer automatiquement une image. Chaque interaction avec un consommateur peut ainsi être personnalisée, ce qui atténue l’effet quelque peu intrusif de la publicité.

L'utilisation de l'intelligence artificielle dans le domaine de la publicité semble avoir un avenir très prometteur, mais le ciblage publicitaire de plus en plus personnalisé soulève un certain nombre de questions juridiques et éthiques, notamment liées à la confidentialité et de sécurité des données. En effet, les annonceurs ont accès à de grandes quantités de données personnelles sur les utilisateurs. Il est donc impératif de garantir la transparence dans la collecte, le stockage et l'utilisation des données, ainsi que l'obtention du consentement approprié des utilisateurs. Certaines pratiques de profilage des utilisateurs à des fins publicitaires peuvent en effet porter de graves atteintes au respect de la vie privée.

La Commission européenne a, par exemple, annoncé, le 14 mars 2024, avoir demandé au réseau social professionnel *LinkedIn* de fournir des informations supplémentaires sur l'utilisation des données personnelles de ses utilisateurs européens. Cette demande d'informations, qui n'est pas une mise en cause à ce stade, intervient dans le cadre du règlement sur les services numériques (DSA). *LinkedIn* est en effet soupçonné d'exploiter certaines données personnelles sensibles de ses utilisateurs (orientation sexuelle, opinions politiques…) à des fins de ciblage publicitaire.

Elle intervient à la suite d'une plainte déposée par des organisations de la société civile qui ont exprimé leurs inquiétudes quant à la possible violation par le réseau social des restrictions de ciblage publicitaire prévues dans le règlement sur les services numériques (DSA) qui encadre les activités des plateformes, en particulier celles des GAFAM. Il est entièrement applicable depuis le 17 février 2024. Ce texte impose de nouvelles obligations pour les fournisseurs de services en ligne et renforce ainsi les mesures existantes du RGPD. Il interdit notamment la publicité ciblée pour les mineurs sur toutes les plateformes, de même que la publicité basée sur des données sensibles comme les opinions politiques, la religion ou l'orientation sexuelle (sauf consentement explicite). Il impose surtout de nouvelles obligations en matière de transparence pour les plateformes : les utilisateurs seront mieux informés de la manière dont les contenus leur sont recommandés et pourront choisir au moins une option qui ne soit pas fondée sur le profilage. En cas de non-respect du DSA, des astreintes et des sanctions peuvent être prononcées. Pour les très grandes plateformes et les très grands moteurs de recherche, la Commission pourra infliger des amendes pouvant aller jusqu'à 6 % de leur chiffre d'affaires mondial.

Cette nouvelle réglementation s'inscrit dans un contexte de surveillance renforcée des pratiques des grandes plateformes en ligne par les régulateurs européens. Elle souligne également la détermination des autorités européennes à faire respecter les règles de protection des données et à garantir les droits des citoyens de l'Union européenne. Des sanctions de plus en plus dissuasives sont ainsi prononcées afin de dissuader d'autres entreprises de commettre des infractions similaires.

Le 24 octobre 2024, la Commission irlandaise de protection des données (*Data Protection Commission* ou DPC) a, par exemple, infligé à *LinkedIn* une amende de 310 millions d'euros pour ses pratiques de publicité ciblée jugées contraires au Règlement général sur la protection des données (RGPD). La DPC a estimé que le consentement obtenu par *LinkedIn* auprès de ses utilisateurs pour l'utilisation de leurs données n'a pas été donné librement, ni été suffisamment éclairé ou spécifique, ni aussi sans ambiguïté.

Une série de plaintes en 2018 avait déjà abouti à de lourdes amendes contre *Google* et *Amazon*, respectivement de 50 millions et 746 millions d'euros, infligées en France et au Luxembourg.

Ces nouvelles réglementations obligent ainsi les plateformes et les annonceurs à s'adapter rapidement pour se conformer à ces règles renforçant la transparence et la protection des utilisateurs. Certaines plateformes, telles que *Meta*, *TikTok* et *Snapchat*, ont d'ailleurs déjà permis aux utilisateurs de désactiver les algorithmes de recommandations personnalisées. Ce nouveau cadre réglementaire devrait normalement assainir les pratiques dans le domaine de la publicité.

48 Qu'est-ce qu'un agent conversationnel ? Peut-il améliorer le service client ?

Un agent conversationnel, aussi appelé *chatbot* en anglais est issu de la contraction de « chat » (conversation) et « bot » (robot), est un système numérique capable d'interagir avec les utilisateurs en langage naturel. L'agent peut dialoguer à l'écrit comme la plupart des agents d'assistance sur les sites de commerce en ligne par exemple, ou à l'oral comme les nombreux assistants vocaux qui sont déjà présents sur le marché : Alexa, Siri, etc.

Historiquement, le premier agent conversationnel nommé Eliza fut créé en 1966 par Joseph Weizenbaum, professeur au MIT. Conçue pour simuler une conversation, Eliza était un programme pouvant imiter un psychothérapeute. Son fonctionnement était très simple : les utilisateurs pouvaient lui exprimer leurs pensées ou lui poser une question, et le *chatbot* répondait en fonction de la réponse obtenue et posait une autre question. Il fonctionnait en s'appuyant sur une base de données de questions-réponses se déclenchant à partir des mots-clés repérés dans la conversation. Mais, cette capacité à répondre de manière pertinente a été un point décisif. Elle a démontré que la communication homme-machine pouvait aller au-delà de simples commandes pour couvrir des échanges plus personnels.

Une multitude d'agents conversationnels ont vu ensuite le jour jusqu'à parvenir à des interactions plus complexes. Cette évolution a été rendue possible grâce à des améliorations considérables dans le traitement du langage naturel et de l'apprentissage automatique. Le NLP pour *Natural Language Processing* ou Traitement du Langage Naturel est, à ce titre, un domaine multidisciplinaire combinant la linguistique, l'informatique et l'intelligence artificielle qui développe la capacité des machines à comprendre, générer ou traduire le langage humain, tel qu'il est écrit ou parlé, et à communiquer avec des hommes. Il s'appuie sur des réseaux de neurones artificiels ou de simples modèles de *machine learning* statistiques.

En réalité, il s'agit d'un terme assez générique qui recouvre un champ d'application très vaste pouvant aller de la traduction d'un texte d'une langue à une autre (par exemple, *Google Translator*) à l'analyse des sentiments d'un texte ou d'un discours. Grâce au *machine learning*, l'agent conversationnel apprend, comprend mieux les requêtes qui lui sont adressées, améliore son vocabulaire, peut construire des réflexions et tirer des conclusions pour des résultats personnalisés. Le *chatbot* peut aussi se baser sur la *Naturel Language Generation* (NLG), qui une sous-catégorie du NLP, permettant de répondre aux demandes à travers une discussion fluide. L'agent conversationnel intelligent est ainsi le résultat de l'association de l'intelligence artificielle et du traitement du langage naturel.

Il devient désormais indispensable pour les entreprises, quel que soit le secteur d'activité. Il joue surtout un rôle important dans la relation client-entreprise en offrant une assistance continue et rapide aux clients qui ont des questions ou des problèmes. Il peut en effet fournir des réponses immédiates, instantanées et simultanées à la plupart des

interrogations de la clientèle. Ces outils sont efficaces pour répondre aux questions récurrentes des clients, notamment celles liées aux produits, au tarif et aux modalités de livraison, comme le suivi d'un colis.

L'agent conversationnel devient ainsi indispensable aux entreprises, mais présente toutefois quelques limites : ses inexactitudes et imprécisions sont encore nombreuses et l'on peut déplorer son manque de créativité, d'originalité et de compréhension des cas complexes et inédits.

Au-delà de son rôle au sein des entreprises, son développement soulève un certain nombre de questions juridiques et éthiques. Parmi ces questions, il convient de s'interroger sur le sort des données à caractère personnel collectées dans le cadre des échanges avec les utilisateurs dont l'étendue reste à définir. En effet, l'agent conversationnel conserve *a priori* une partie de l'échange à des fins de preuve en cas de litige ou encore pour des raisons de sécurité. Les enjeux concernent alors l'information communiquée à l'utilisateur relative au traitement de ses données personnelles, le consentement de celui-ci ainsi que, le cas échéant, les possibilités de limiter ces traitements ou de s'y opposer. La question du consentement peut être particulièrement sensible lorsqu'il s'agit d'un agent conversationnel installé à domicile et destiné à être un assistant avec lequel s'engagent des conversations intimes pouvant porter atteinte à la vie privée.

Au-delà de la question de la protection des données à caractère personnel, il pose également des questions éthiques qui ont été développées dans l'avis n° 3, du 15 septembre 2021, du Comité national pilote d'éthique du numérique (CNPEN) portant sur les enjeux éthiques des agents conversationnels. Le Comité identifie différents enjeux éthiques qui concernent « *le brouillage des frontières entre la machine et l'être humain, l'imitation du langage et des émotions par les agents conversationnels, ainsi que leur capacité à manipuler les interlocuteurs humains* » directement (informations inexactes ou tronquées) ou indirectement, *via* des stratégies de « nudge » (pousser doucement la personne dans une direction considérée comme « bonne »). Les risques posés par les agents conversationnels ne sont pas négligeables, c'est la raison pour laquelle le Comité émet un certain nombre de recommandations notamment destinées aux concepteurs. Il les invite, par exemple, à réduire la projection de qualités morales sur les agents conversationnels et d'affirmer de façon claire leur statut de « machine » pour empêcher les dérives liées à un fort anthropomorphisme.

La réflexion sur les enjeux d'éthique des agents conversationnels va inévitablement se poursuivre, car le marché des agents conversationnels intelligents devrait connaître une croissance importante dans les prochaines années en raison de l'augmentation de l'utilisation d'appareils mobiles, de l'internet des objets ainsi que de l'évolution des technologies de reconnaissance vocale et du traitement du langage naturel.

49 Qu’est-ce que la cobotique ?

Ces dernières années, l’utilisation de l’intelligence artificielle en robotique a connu une nette augmentation. L’utilisation de robots dotés de systèmes d’IA offre de nombreux avantages aux entreprises industrielles. Elle permet, d’une part, de diminuer les coûts de la main-d’œuvre et, d’autre part, elle réduit les erreurs humaines et augmente la productivité grâce aux robots pouvant traiter plusieurs tâches simultanément sans avoir besoin de pauses ou de périodes de repos. Elle peut enfin améliorer la sécurité, car les robots peuvent détecter les dangers et réagir en conséquence.

La robotisation industrielle et la quatrième révolution industrielle portée par l’AI touchent tous les secteurs de l’industrie. Les robots industriels sont, à ce titre, utilisés dans divers secteurs : agroalimentaire, automobile, pharmaceutique, cosmétique, etc. Ils sont surtout utilisés pour réduire les travaux manuels comme l’emballage, la découpe, l’assemblage des composants d’un produit, etc.

Les robots industriels ont été longtemps séparés des opérateurs pour des raisons de sécurité, car ils ne sont pas encore tout à fait capables de collaborer avec l’homme. Pourtant, les technologies évoluent et tendent à privilégier les robots collaboratifs qui travaillent sans risques, aux côtés des humains. La collaboration homme-robot – ou la cobotisation – est une branche émergente de la robotique qui redéfinit la place du robot comme travaillant aux côtés et non à la place du salarié.

Le principe de la cobotique est de permettre l’intégration des robots dans l’espace de travail pour réaliser diverses missions avec les humains. Cette approche permet d’augmenter la productivité tout en préservant l’emploi, en compensant les faiblesses de l’humain et du robot par les qualités de l’autre. À la différence de robots classiques, les cobots ne travaillent pas seuls. Aux côtés des humains, ils assurent une bonne coopération entre l’homme et la machine.

Il existe trois grands types de cobots : les robots collaboratifs pilotés par un opérateur à proximité immédiate du système (co-manipulation), les cobots commandés à distance (télé-opération) et les exosquelettes, des structures électromécaniques qui assistent le corps humain dans son effort et permettent de décupler la force de l’homme sans le fatiguer. Les applications de la cobotique se développent dans de nombreux secteurs : l’automobile, l’agroalimentaire, l’aéronautique, la construction navale, la défense ou encore la santé.

Ce marché, encore émergent, est en forte croissance depuis ces dernières années. Les premiers robots collaboratifs ne datent cependant pas d’aujourd’hui, mais ils ont beaucoup progressé depuis le début des années 2000. En France, les grands acteurs industriels, tels que Safran, Vinci, Airbus, se tournent vers la cobotique et commencent à l’intégrer dans leurs systèmes de production. Les PME sont aussi intéressées par ces robots faciles à programmer, polyvalents et économiques. Les cobots constituent en effet une alternative moins onéreuse aux robots classiques dédiés à l’industrie. Certaines start-ups développent déjà des cobots à moins de 10 000 € pour les rendre plus accessibles aux entreprises.

Les robots dotés d'IA deviennent ainsi mobiles, intelligents et collaboratifs. Leur utilisation peut améliorer la qualité de travail de nombreux salariés en leur confiant les tâches répétitives et dangereuses. Les cobots peuvent aussi faciliter l'accès au travail de personnes handicapées ou âgées.

La collaboration entre humains et robots semble être une direction pertinente, mais elle suscite tout de même des interrogations sur l'évolution de cette relation dans le temps. En effet, la mobilité accrue des cobots et leur autonomie décisionnelle croissante, fondées sur des algorithmes d'auto-apprentissage, pourraient rendre leurs actions moins prévisibles pour les travailleurs qui collaborent avec eux. Une particularité de ces systèmes d'IA basés sur l'apprentissage automatique est que leur performance peut continuer d'évoluer s'ils sont régulièrement alimentés en données et entraînés. Faudra-t-il alors réviser régulièrement la répartition du travail entre l'homme et le cobot ? Que se passe-t-il dans le cas où le cobot doté d'un système d'IA devient plus performant que l'humain ?

Une dépendance excessive envers la technologie pourrait également entraîner des risques de perte de savoir-faire et de sécurité. Au lieu de constituer une aide, le cobot pourrait en effet finir par déposséder le travailleur de certains aspects de son activité, ce qui pourrait finir par lui faire perdre des compétences. De plus, le fait que les cobots soient connectés à l'internet des objets est susceptible d'entraîner des problèmes de cybersécurité. Enfin, se pose la question de la santé des travailleurs qui doivent suivre le rythme et le niveau de travail d'un cobot pour atteindre son niveau de productivité. Le travail avec les robots peut aussi finir par isoler le travailleur du fait de la réduction des contacts avec les collègues humains, ce qui peut affecter sa santé mentale.

Quelle que soit sa nature, l'automatisation se traduit par une transformation des activités humaines. Il reste à définir les orientations que l'on souhaite lui donner.

50 L'IA peut-elle être un outil au service des entreprises en difficulté ?

Il n'est pas nécessaire de démontrer que l'anticipation et le traitement, le plus en amont possible, des difficultés d'une entreprise favorise son redressement. Tous les acteurs intervenant en soutien des entreprises partagent ce constat. Depuis quelques années, l'IA est considérée, à ce titre, comme un outil efficace d'aide et de soutien pour prévoir d'éventuelles difficultés financières qui amèneraient une entreprise à l'ouverture d'une procédure collective. Plusieurs études démontrent que les algorithmes d'apprentissage automatique, alimentés par une grande quantité de données, dépassent les méthodes statistiques classiques en termes de précision de la prévision. L'IA semble donc être un outil efficace dans la détection des signes précurseurs des difficultés financières et du risque de faillite. La démarche visant à détecter les entreprises en difficulté à l'aide de « clignotants » n'est pas nouvelle, c'est l'usage de l'IA qui l'est.

D'ailleurs, le gouvernement a fait du dépistage précoce des difficultés des entreprises, afin de mieux les accompagner, une orientation prioritaire. Pour ce faire, un outil prédictif dénommé « Signaux Faibles », expérimenté en 2016 dans la région Bourgogne Franche-Comté à l'initiative de la DREETS et des Urssaf, a permis d'identifier une soixantaine d'entreprises qui n'avaient pas été détectées grâce aux dispositifs existants. Cet outil doit normalement permettre de détecter les entreprises présentant un risque fort ou modéré de défaillance, dans les 18 prochains mois. Les résultats prometteurs de ce dispositif ont conduit à la signature d'un partenariat, en avril 2019, avec la Direction générale des entreprises (DGE), la Banque de France, la Délégation générale à l'emploi et à la formation professionnelle (DGEFP), l'Agence centrale des organismes de sécurité sociale (ACOSS) et la Direction interministérielle du numérique (DINUM) afin de le déployer progressivement sur tout le territoire national.

Le partage et la mise en commun des données, dont disposent les différents services de l'État, permettent de disposer d'indices pour détecter plus rapidement les entreprises en difficulté. Ces données sont d'ailleurs complémentaires : les données financières fournissent un éclairage sur les tendances de long terme de l'entreprise (rentabilité, ratios financiers, endettement, fonds propres), les données sur l'emploi permettent d'identifier des baisses d'activité ponctuelles tandis que les données sur les cotisations sociales alertent sur des tensions de trésorerie. Le recoupement de ces différentes données permet à l'algorithme de détecter le profil statistique des entreprises fragiles, avant la survenue de difficultés structurelles. Ces données sont comme un faisceau d'indices permettant à l'algorithme de détecter, en s'appuyant sur l'expérience passée, des signes précoces de fragilité. Depuis la crise sanitaire en mars 2020, l'outil s'est perfectionné et enrichi de fonctionnalités supplémentaires visant à améliorer cette détection. Ce choix d'un horizon de prévisions à 18 mois permet de parvenir à une prévision précise, tout en laissant suffisamment de temps pour adopter les décisions qui s'imposent. Ces résultats sont, à ce titre, partagés, dans la plus stricte confidentialité, au sein d'une plateforme numérique collaborative uniquement ouverte aux partenaires

précités. L'objectif du dispositif est la prédiction de difficulté à 18 mois : cette information est délivrée sous forme de deux seuils d'alerte (rouge et orange) et mise à jour mensuellement.

Grâce au ciblage des entreprises présentant certaines fragilités, les partenaires de ce dispositif peuvent proposer les solutions les plus adaptées à leurs besoins pour consolider leur développement ou les aider à surmonter cette période de fragilité. Cette méthode incite ainsi le chef d'entreprise à s'intéresser à une réalité qu'il méconnaît, ou dont il n'a pas mesuré l'importance afin de parvenir à des solutions traitant le problème en amont avant l'enlisement.

L'anticipation est indéniable mais sera-t-elle le gage de réussite de la prévention des difficultés ? En effet, les difficultés d'une entreprise ne se résument pas toujours à des variables comptables et financières. L'automatisation de l'ouverture d'une procédure préventive, voire de traitement des difficultés, ne relève plus de l'anticipation, mais devient une réalité. Il convient toutefois de rester prudent afin que ce modèle prédictif ne se transforme pas en modèle prescriptif en automatisant l'ouverture des procédures de prévention ainsi que le traitement des difficultés.

Malgré sa pertinence, ce dispositif interpelle en raison des données des entreprises qu'il collecte, fouille et analyse. La confidentialité des données est certes assurée au sein des administrations concernées, mais le dispositif manque tout de même de transparence dans la mesure où les entreprises ne sont pas informées de la collecte de leurs données ni des traitements réalisés. Or, il peut être tout à fait légitime qu'elles contestent l'utilisation de leurs données dans la mesure où elles n'ont pas été informées de cette collecte et n'y ont pas consenti. Il convient en effet de souligner que le RGPD écarte de son champ d'application les traitements des données à caractère personnel qui concernent les personnes morales. Les entreprises ne bénéficient donc pas des mêmes protections que les personnes physiques lorsque leurs données sont traitées. Néanmoins, la transparence quant à l'utilisation des données de ces entreprises est nécessaire, surtout en raison du risque que présentent leur croisement et la réutilisation des données pour d'autres finalités. Il semble donc fondamental que les pouvoirs publics garantissent la transparence de ce dispositif en déterminant avec précision la finalité de cette collecte afin d'éviter une réutilisation ultérieure non souhaitée ou non acceptée par les entreprises de ces données.

VI. L'IMPACT DE L'INTELLIGENCE ARTIFICIELLE SUR LE TRAVAIL

51 Quels sont les effets de l'utilisation de l'IA sur le pouvoir de l'employeur ?

Pour conclure un contrat de travail, trois critères sont indispensables : une prestation de travail, une rémunération et un lien de subordination. Ce lien de subordination a été érigé par la jurisprudence comme l'un des critères déterminant l'existence du contrat de travail entre un employeur et un salarié. Il se caractérise par le pouvoir, pour l'employeur, de donner des ordres et directives, d'en contrôler l'exécution et de sanctionner les manquements de son subordonné.

L'employeur dispose ainsi du pouvoir de direction par lequel il édicte des instructions pour l'accomplissement par le salarié de sa prestation de travail. Il dispose également d'un pouvoir normatif qui lui permet d'édicter des normes générales. Il dispose enfin d'un pouvoir de contrôle lui permettant de sanctionner le salarié en cas de manquements.

Il est évident que l'utilisation dans l'entreprise d'outils se substituant ou accompagnant le pouvoir de l'employeur altère ce lien de subordination. La gestion algorithmique du personnel produit sur le pouvoir patronal un effet paradoxal. D'un côté, elle le renforce en le parant des vertus de l'objectivité et de la rationalité, de l'autre, elle le dissout en transférant son pouvoir à un algorithme. Autrement dit, l'IA renforce le pouvoir patronal tout en contribuant à le dissoudre. En déléguant son pouvoir à un algorithme sur lequel il n'a aucun véritable contrôle, l'employeur abandonne une partie de son pouvoir de jugement et de décision. Il convient sans doute de nuancer cette affirmation dans la mesure où l'employeur n'est pas contraint de suivre les recommandations de l'algorithme étant le décisionnaire final, mais il est vrai que c'est bien l'intelligence artificielle qui proposera les termes de la décision que l'employeur peut choisir de suivre ou pas.

Le risque majeur pour l'employeur réside dans la perte du contrôle des modalités d'exécution de la prestation de travail. Ce ne sera alors plus l'employeur qui transmettra les directives aux salariés mais l'algorithme programmé afin d'augmenter la productivité. Le pouvoir de direction et de contrôle de l'employeur s'en trouve amoindri.

La fonction de contrôle des salariés constitue en effet l'un des enjeux de l'intégration de l'intelligence artificielle dans les entreprises. Le contrôle patronal direct est alors remplacé par une analyse détaillée, par voie algorithmique, des données concernant la productivité dont dispose l'entreprise. Un système de recueil et d'analyse des données personnelles des salariés se substitue au contrôle personnel de l'employeur. Le lien de subordination ne disparaît pourtant pas, il est simplement remplacé par un algorithme qui contrôle les salariés, les évalue et les récompense.

Si l'intelligence artificielle prend les décisions à la place de l'employeur, peut-on considérer qu'elle le fait en vertu d'une délégation de pouvoirs ? Seule une personne disposant des pouvoirs a la faculté de les transmettre. Le délégant est donc en principe le dirigeant de l'entreprise. Aucun écrit n'est en principe obligatoire, mais le bénéficiaire de la délégation doit posséder quelques qualités, à savoir la compétence, l'autorité et les moyens nécessaires pour exercer sa délégation.

En admettant qu'une intelligence artificielle puisse avoir les connaissances techniques et l'autorité, il est peu probable que l'on puisse lui reconnaître l'existence de moyens humains (personnel suffisant), matériels et financiers pour l'exercice de la délégation. Il paraît donc difficile pour une intelligence artificielle de posséder les trois caractéristiques.

Par ailleurs, la responsabilité civile du délégant demeure, comme le prévoit l'article L. 4741-7 du Code du travail qui prévoit que : « *L'employeur est civilement responsable des condamnations prononcées contre ses directeurs, gérants ou préposés* ».

Afin de conserver leur pouvoir de direction, les employeurs devront définir précisément les conditions dans lesquelles cet outil sera intégré dans l'entreprise pour qu'ils puissent sauvegarder leurs pouvoirs et ne pas perdre leur légitimité face à leurs salariés.

52 Quels sont les effets de l'intelligence artificielle sur l'emploi et les compétences ?

Il existe d'abondantes études sur le sujet de l'intelligence artificielle dans l'entreprise, mais encore peu d'études empiriques. Ces études tentent généralement d'appréhender sur un plan prospectif les impacts du déploiement de projets de système d'IA sur l'emploi et le travail. Elles tentent surtout d'anticiper les types d'emplois supprimés dans un avenir proche en raison du remplacement par l'IA

Le laboratoire de recherche dédiée à l'intelligence artificielle, LaborIA, créé par le ministère du Travail, de la Santé et des Solidarités et Inria en 2021, a publié, à cet effet, en mai 2024 les résultats d'une étude ayant été menée pendant deux ans sur les impacts de l'IA sur le travail. Les résultats de l'enquête ont permis de mettre en lumière les enjeux de l'appropriation de l'IA dans le monde du travail. L'étude démontre que le déploiement des IA dans les organisations marque le début d'un processus continu d'apprentissage et d'adaptation en raison du caractère apprenant et évolutif de l'IA. En d'autres termes, le déploiement de l'intelligence artificielle dans les organisations n'est pas le point d'aboutissement des processus d'innovation, mais un nouveau point de départ qui s'inscrit dans un processus d'apprentissage continu humain-machine.

Un aspect majeur mis en lumière par l'étude du LaborIA est le conflit, nommé conflit de rationalité, entre les priorités des décideurs et les préoccupations des salariés lors du déploiement des systèmes d'intelligence artificielle. À cet égard, l'étude quantitative du LaborIA montre que les motifs d'utilisation les plus cités par les décideurs d'entreprise utilisateurs de systèmes d'IA sont la réduction des risques d'erreurs (81 %), suivis par l'amélioration des performances des salariés (75 %), puis par la réduction des tâches fastidieuses (74 %). De leur côté les salariés évaluent l'IA à travers le prisme de l'intégration pratique dans leurs activités quotidiennes et s'interrogent sur la reconnaissance, l'autonomie et le sens de leur travail face à ces changements. Ce « conflit de rationalité » peut mener à des situations difficiles pour les personnes qui travaillent, s'il n'est pas résolu par un compromis nécessaire afin de ne pas risquer de faire émerger des configurations humain-machines aliénantes. L'IA doit en effet être perçue comme un moyen d'augmenter les aptitudes et les compétences humaines. On parle alors de configurations dites « capacitantes ».

Pour les auteurs de l'étude, l'IA reconfigure les rôles professionnels, les compétences et le management, notamment le management intermédiaire. Les changements organisationnels engendrés par l'introduction d'un système d'IA peuvent en effet faire naître un sentiment de déclassement des cadres intermédiaires affectés dans leurs missions d'encadrement et de pilotage. En réaction à ce sentiment, certains peuvent réagir en réaffirmant leur statut hiérarchique en se saisissant du nouveau dispositif pour reprendre le contrôle sur des dimensions du travail qui échappe à leur supervision directe. D'autres voient dans ces technologies une opportunité de renforcer leur proximité avec les salariés : meilleure communication et restitution des informations, etc. L'IA soulève ainsi des questions liées aux rapports de pouvoir et à l'orientation du management au sein d'une organisation.

Dans son rapport, le LaborIA émet une série de recommandations visant à améliorer le dialogue social pour une meilleure intégration dite « capacitante » des systèmes d'IA dans le monde du travail. Pour ce faire, les organisations doivent créer un environnement qui favorise l'autonomie des salariés et dans lequel ils se sentent valorisés et responsables de leur travail, ce qui renforce leur compétence et leur engagement.

Il recommande notamment de partir du travail réel, soit ce que les travailleurs font vraiment, plutôt que du travail prescrit, pour adapter les systèmes d'IA aux conditions réelles du travail afin d'assurer leur pertinence et leur acceptation. Il propose également de favoriser la co-conception. Il faut créer des espaces de dialogue continu entre développeurs, gestionnaires, instances représentatives du personnel et utilisateurs finaux pour ajuster et améliorer les technologies en fonction des retours sur le terrain.

Par ailleurs, il prône le déploiement de systèmes d'IA au service de la sécurisation des travailleurs et centrés sur l'amélioration de la qualité de vie au travail et la réduction des risques socioprofessionnels.

Il insiste aussi sur les efforts qui devront être réalisés afin de rendre le fonctionnement et les résultats de l'IA compréhensibles pour les travailleurs en situation d'activité. Enfin, il suggère de prendre en compte des potentiels effets inattendus de l'IA sur le travailleur, le travail et l'organisation. Ces situations doivent faire l'objet de retours d'expérience pour adopter des mesures correctrices.

Compte tenu des impacts de l'IA sur la réorganisation des entreprises, les employeurs ont intérêt à anticiper l'évolution de l'emploi et des compétences de leurs salariés. Intégrer l'IA peut permettre d'optimiser les compétences de salariés si son utilisation permet de supprimer certaines tâches pénibles, répétitives et susceptibles de générer un désintérêt. L'utilisation de ces outils peut aussi être une source de dérives dans le monde du travail. C'est la raison pour laquelle il est important que les partenaires sociaux ainsi que le législateur s'emparent de cette question pour accompagner cette transformation.

53 L’IA peut-elle améliorer le processus de recrutement ?

L’utilisation de l’intelligence artificielle prend de plus en plus de place dans le processus de recrutement. Les solutions intégrant l’IA concernent aujourd’hui toutes les étapes de ce processus. La promesse d’automatisation de plusieurs étapes de celui-ci (la présélection, les entretiens) peut en partie expliquer l’attrait des entreprises pour ces outils. En raison de leur capacité à traiter d’importants volumes de candidatures et effectuer des classements selon des critères définis, les solutions de recrutement basées sur l’IA sont réputées moins chronophages, et donc moins coûteuses que le recrutement classique.

L’IA est surtout utilisée pour des tâches telles que la sélection des CV. Il devient en effet possible d’automatiser le tri des CV, tâche qui peut s’avérer particulièrement fastidieuse. Outre le tri et la catégorisation des candidats, l’IA peut aussi contribuer à lutter contre les stéréotypes des recruteurs humains lorsqu’ils découvrent un CV ou une lettre de motivation, ce qui peut permettre une prise de décision fondée sur des critères plus objectifs. L’IA aurait également le potentiel de détecter des profils plus atypiques et variés grâce au nombre important de sources que ces outils sont capables d’exploiter, telles que les réseaux sociaux professionnels.

Elle peut aussi être utilisée pour évaluer les compétences des candidats grâce à des tests générés automatiquement. Les compétences techniques, ainsi que les compétences comportementales ou relationnelles (communication, esprit critique, travail en équipe, etc.), aussi appelées *soft skills*, et les aptitudes spécifiques au poste (par exemple, le service à la clientèle) sont parmi les compétences qui sont fréquemment évaluées.

Enfin, elle peut automatiser le processus d’entretien, notamment en réalisant des entretiens automatisés grâce à des agents conversationnels (chatbots) qui posent des questions et analysent les réponses pour trier les candidats. Les algorithmes d’apprentissage automatique peuvent en effet analyser les réponses des candidats et évaluer leur adéquation au poste en fonction de critères prédéfinis. Cela permet aux recruteurs de gagner du temps et de se concentrer sur les candidats les plus qualifiés. Certains outils seraient même capables de décoder les expressions du visage du candidat censées renseigner sur sa personnalité. L’entreprise américaine *HireVue* propose, par exemple, d’évaluer les réponses données sur la base des expressions faciales et de la posture corporelle. Cette démarche consistant à utiliser des algorithmes établissant des liens entre les expressions faciales ou la posture corporelle, les traits de la personnalité et les compétences est assez discutable tant le risque d’erreurs semble important. Ces erreurs peuvent être provoquées par des corrélations fallacieuses, surtout qu’il est parfois difficile d’expliquer les résultats des algorithmes basés sur l’apprentissage profond (*deep learning*). Il y a donc un risque pour les recruteurs à suivre sans recul les recommandations que leur proposent ces outils pouvant manquer de transparence et dont les résultats ne sont pas facilement explicables. Cela peut aussi potentiellement créer des discriminations – les modes de réaction corporelle à un même stimulus pouvant

varier en fonction de l'environnement socio-culturel dans lequel le candidat a grandi – si l'IA n'est construite et entraînée qu'en fonction des standards de l'environnement socio-culturel dominant.

Ce faisant, même si l'intelligence artificielle facilite le processus de recrutement, celui-ci ne peut être totalement automatisé, car les interactions humaines sont nécessaires dans le choix d'un nouveau salarié. Par ailleurs, si l'intelligence artificielle permet d'alléger la tâche des recruteurs tout en réduisant le coût du recrutement, elle soulève des préoccupations éthiques qu'il convient de prendre en compte.

En effet, l'une des principales préoccupations éthiques liées à l'utilisation de l'IA dans le recrutement concerne le risque de reproduire et d'amplifier les préjugés. Les systèmes d'IA ont le potentiel de reproduire et de renforcer à grande échelle les biais discriminatoires portant sur l'âge, le genre, l'origine ethnique, la catégorie sociale, etc. La reproduction de ces biais dans les systèmes d'IA peut s'expliquer par les données d'entraînement incluant généralement les décisions passées des recruteurs, parfois établies sur des préjugés. Les algorithmes sont ainsi alimentés avec les profils des personnes déjà recrutées et réutilisent les mêmes critères dans la sélection des nouveaux candidats. La qualité des données d'entraînement des algorithmes ainsi que les critères de sélection constituent des enjeux essentiels afin de supprimer ces biais. Même si la question des discriminations n'est pas nouvelle en matière de recrutement, elle se pose désormais avec acuité dès lors que le processus de recrutement est automatisé.

Au-delà de la question des discriminations, il existe un risque réel d'uniformisation des profils. Les systèmes d'IA risquent de sélectionner des candidats ayant toujours des profils proches et donc de réduire la diversité qui enrichit une entreprise. Or, cette diversité permet d'augmenter la performance des entreprises tandis que l'uniformisation limite leur capacité d'adaptation face aux changements. Par ailleurs, ces nouveaux modes de recrutement peuvent conduire à une forme de standardisation des comportements et des attitudes des candidats. Ils pourraient également créer une disparité entre les candidats sachant maîtriser leur gestuelle et réactions dans le cadre d'un entretien à distance et ceux moins à l'aise dans une telle situation. Ce type de standardisation risque de nuire à la créativité et à l'esprit d'initiative du salarié, surtout si son évolution de carrière est également encadrée par l'intelligence artificielle. Cette dernière doit donc rester un outil complémentaire au recruteur mais ne doit en aucun se substituer à lui.

Le règlement européen sur l'intelligence artificielle (IA Act) classe d'ailleurs, dans son annexe III, 4 a), les systèmes d'IA « *destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, notamment pour publier des offres d'emploi ciblées, pour analyser et filtrer les candidatures et pour évaluer les candidats* » dans la catégorie des systèmes d'IA à haut risque qui seront soumis à des exigences supplémentaires imposées aux fournisseurs (un système de gestion des risques tout au long du cycle de vie, gouvernance des données, surveillance humaine, etc.). Ce cadre légal permet de poser des garde-fous afin d'éviter les usages pouvant conduire à des exclusions injustifiées de certains candidats du marché de l'emploi.

54 Les représentants du personnel peuvent-ils contrôler l'utilisation de l'IA dans l'entreprise ?

L'introduction de l'intelligence artificielle dans les entreprises et les administrations fait rarement l'objet d'un échange entre les institutions représentatives du personnel et la direction. Lorsque cet échange a lieu, il se fait généralement sans prise en compte des spécificités de cette technologie qui est alors considérée comme une continuité des transformations engendrées par la digitalisation. Or, un dialogue social approfondi semble nécessaire au regard de l'impact sur les droits des travailleurs.

Cette exigence découle de l'esprit de la loi qui s'inscrit dans une logique de renforcement du rôle des acteurs sociaux, mais elle est aussi nécessaire au regard des incertitudes sur les conséquences de l'intelligence artificielle sur le travail et l'emploi. C'est la raison pour laquelle les travailleurs et leurs représentants doivent être associés aux décisions qui aboutissent au développement et au déploiement de l'IA.

Ces représentants doivent tout d'abord être informés de tout projet d'introduction d'une telle technologie dans l'entreprise. L'article L. 2312-38 du Code du travail prévoit à cet effet que : « *le comité social et économique est informé, préalablement à leur utilisation, sur les méthodes ou techniques d'aide au recrutement des candidats à un emploi ainsi que sur toute modification de celles-ci. Il est aussi informé, préalablement à leur introduction dans l'entreprise, sur les traitements automatisés de gestion du personnel et sur toute modification de ceux-ci. Le comité est informé et consulté, préalablement à la décision de mise en œuvre dans l'entreprise, sur les moyens ou les techniques permettant un contrôle de l'activité des salariés* ».

Il résulte de cet article que si le CSE doit être informé et consulté sur les moyens et techniques permettant le contrôle de l'activité des salariés, il est cependant seulement informé sur les méthodes ou techniques d'aide au recrutement des candidats à un emploi et sur les traitements automatisés de gestion du personnel. Or, il peut sembler nécessaire de prévoir en outre une obligation de consultation des élus, car les algorithmes de profilage et le recours au traitement automatisé des données à des fins de gestion du personnel exigent, en raison de leur caractère sensible, un dialogue plus poussé avec les représentants du personnel.

Par ailleurs, l'article L. 2312-8 du Code du travail prévoit que le CSE doit être informé et consulté lors de l'introduction de nouvelles technologies ou de tout aménagement important modifiant les conditions de santé et de sécurité ou les conditions de travail. Pour ce faire, le CSE peut faire appel à un expert habilité en cas d'introduction de nouvelles technologies ou de projet important modifiant les conditions de santé et de sécurité ou les conditions de travail, ce qui peut aider les élus à comprendre le fonctionnement et les risques des dispositifs algorithmiques.

Outre l'information et la consultation du CSE, les applications d'IA pourraient avoir des répercussions importantes sur les conditions de travail, ce qui permet de les intégrer dans le champ de la négociation annuelle relative à la qualité de vie au travail dont les thèmes sont énumérés à l'article L. 2242-17 du Code du travail. Parmi ces thèmes, certains pourraient avoir un lien avec l'intelligence artificielle, tels que l'articulation entre vie

personnelle et vie professionnelle, les mesures de lutte contre les discriminations en matière de recrutement, d'emploi ou d'accès à la formation professionnelle, l'insertion et le maintien dans l'emploi des personnes handicapées ou encore les modalités d'exercice du droit à la déconnexion. L'objet de la négociation porte sur les conditions de travail et de santé des salariés et non le dispositif lui-même.

Afin d'appréhender tous les effets de la transformation numérique, il est nécessaire de promouvoir un dialogue social global, ce que ne favorise pas forcément le Code du travail. Autrement dit, il convient de décloisonner le dialogue social et de ne pas traiter isolément le cas de l'IA sans prendre en compte les autres technologies de la révolution numérique.

C'est l'approche choisie par l'accord-cadre européen du 22 juin 2020 sur la numérisation qui propose aux acteurs nationaux une méthodologie globale pour dialoguer sur la numérisation. Il identifie, à ce titre, quatre enjeux majeurs que les partenaires sociaux sont invités à prendre en considération : les compétences numériques et la sécurisation de l'emploi, les modalités de connexion et déconnexion, l'intelligence artificielle et le principe du contrôle humain, le respect de la dignité humaine et la surveillance. Il définit à cet effet une démarche d'action en cinq étapes : 1. exploration conjointe afin de créer un climat de confiance pour discuter des enjeux, opportunités et risques de la digitalisation et de leur impact sur le lieu de travail ainsi que les actions possibles ; 2. cartographie/évaluation/analyse régulière conjointe afin d'identifier les mesures et actions possibles ; 3. vue d'ensemble commune de la situation et adoption de stratégies pour la transformation numérique ; 4. adoption de mesures et d'actions appropriées ; 5. évaluation et suivi des actions menées.

Concernant spécifiquement l'intelligence artificielle, l'accord-cadre européen affirme très clairement que le principe doit être celui du contrôle humain sur les systèmes d'IA. Cela suppose donc un système transparent sur lequel il est possible de recueillir des explications claires et transparentes sur son fonctionnement et d'effectuer des vérifications de prévenir des erreurs de traitement. Cette transparence en termes d'accès à l'information doit également être assurée dans les situations où les systèmes d'IA sont utilisés dans les procédures de ressources humaines, telles que le recrutement, l'évaluation, la promotion et le licenciement, l'analyse des performances. Par ailleurs, un travailleur affecté par une décision liée à cette technologie peut demander une intervention humaine et/ou contester la décision en requérant un test sur les résultats de l'intelligence artificielle. Enfin, ces systèmes doivent être conçus et exploités conformément à la législation en vigueur, notamment au RGPD.

Cet accord-cadre constitue un outil méthodologique utile pour aider les acteurs sociaux à traiter la question de la transformation numérique, notamment de l'intelligence artificielle. Le dialogue social doit ainsi prendre toute sa place pour que le recours à l'IA soit discuté et négocié à tous les niveaux de l'entreprise.

55 L’IA peut-elle améliorer la santé et la sécurité au travail ?

L’une des promesses souvent mise en avant de l’intelligence artificielle, en dehors de l’automatisation des tâches, serait de réduire les risques au travail dans des environnements variés, voire de contribuer à leur prévention. L’IA en matière de santé au travail répond à des enjeux individuels, collectifs, épidémiologiques et de santé publique et sanitaire. Pour les employeurs, elle leur permet d’approfondir la démarche d’analyse des risques afin d’améliorer la prévention des risques au travail.

En effet, les capacités de calculs et de traitement de l’information de l’intelligence artificielle ouvrent de nouvelles perspectives en matière d’accidentologie et d’épidémiologie. L’IA offre la possibilité de croiser et d’analyser un nombre important d’indicateurs et de données permettant d’identifier les causes de maladies ou d’accidents en recherchant des similarités ou des liens de causalité. L’objectif principal de l’exploitation de ces données par des outils d’intelligence artificielle est l’amélioration de la prévention des risques professionnels. Pour ce faire, il conviendra de disposer de données fiables et de ne pas oublier certaines dimensions pour lesquelles il n’existe pas forcément de données exploitables, telles que la dimension organisationnelle. Par ailleurs, il convient de ne pas réduire la question de la prévention à la seule analyse des données personnelles, car cela pourrait conduire à une individualisation du suivi du travailleur qui supplanterait l’approche collective de la prévention.

Outre le potentiel de ces outils de détection basés sur l’analyse de données, l’automatisation de certaines tâches pourrait aussi contribuer à cette démarche de prévention en limitant certains risques potentiels pour les travailleurs : la pénibilité, la réduction des expositions à des substances dangereuses, la prévention des troubles musculo-squelettiques.

Enfin, les nouvelles formes de contrôle des travailleurs utilisant des systèmes d’IA peuvent aussi permettre de réduire l’exposition à divers facteurs de risques dont le harcèlement, et ainsi détecter rapidement le stress, les problèmes de santé et la fatigue. Les informations collectées peuvent en effet servir à recenser les problèmes de santé et de sécurité au travail, incluant les risques psychosociaux, et à envisager des mesures adaptées.

Dans le cadre de son activité de prospective, l’INRS (Institut National de Recherche et de Sécurité) a mené une réflexion collective sur le thème de l’intelligence artificielle pouvant être utilisée à des fins de prévention des risques professionnels à l’horizon 2035.

Dans son rapport de synthèse de 2022, ce groupe de travail préconise d’accompagner le développement de ces dispositifs afin qu’ils soient compatibles avec les valeurs essentielles de la santé et sécurité au travail. Il préconise notamment de former les acteurs de la prévention, notamment les employeurs et les représentants du personnel, aux opportunités et aux risques que représentent ces nouvelles technologies en matière de santé et de sécurité au travail. Ces formations devront assurer une bonne compréhension du mode de fonctionnement de ces outils, des enjeux éthiques, du cadre réglementaire les régissant et de leurs risques. Cet effort de formation devrait également être étendu aux développeurs et aux concepteurs de ces systèmes. Une formation à la

santé et sécurité au travail est en effet nécessaire pour les sensibiliser aux risques associés à ces technologies et les amener à intégrer le respect des principes de prévention dès la conception de ces outils.

Outre la formation, il préconise également de promouvoir auprès des entreprises les démarches reposant sur l'expérimentation et l'évaluation pour permettre de mesurer concrètement les conséquences de ces nouveaux systèmes sur l'organisation de l'entreprise et le travail des salariés, et de conserver une possibilité de revenir en arrière.

Selon le groupe de travail, ces technologies doivent en effet d'abord faire leur preuve, il paraît donc essentiel de ne pas fonder toute la problématique de la santé et de la sécurité au travail sur ces celles-ci. En tout état de cause, l'humain doit rester au premier plan de l'organisation du travail. Certains accidents du travail ne pourront d'ailleurs pas être anticipés par l'intelligence artificielle, car ils surviennent souvent dans le cadre de situations atypiques et imprévues (situations dégradées, pannes, opérations de maintenance, etc.).

Par ailleurs, il souligne qu'il convient d'apporter une attention particulière aux normes et à la réglementation encadrant les technologies d'IA afin que soient pris en compte systématiquement les principes de santé et sécurité au travail dès leur conception. Enfin, il propose qu'une réflexion collective soit réalisée sur la question des données nécessaires au fonctionnement de ces systèmes d'IA. Il s'agira notamment de définir des règles pour la constitution des jeux de données qui doivent se faire dans le respect de la vie privée des travailleurs.

S'il est assez évident que ces nouvelles technologies auront une influence sur les travailleurs, il reste à déterminer quelles seront leurs conséquences sur leur sécurité et leur santé ainsi que sur leur bien-être. Si les promesses de l'IA en matière d'augmentation de la productivité sont réelles, encore faudrait-il pouvoir répondre aux nombreuses questions majeures liées à la santé et la sécurité au travail qui se posent lors de l'intégration de ces systèmes d'IA dans un environnement de travail. Les promesses de ces nouvelles technologies ne doivent pas occulter ces problématiques et faire passer la question des conditions de travail au second plan.

56 L’employeur peut-il surveiller les salariés grâce à une IA ?

La surveillance et le contrôle font partie intégrante du monde du travail depuis la révolution industrielle. L’idée de surveillance n’est pas spécifique au milieu professionnel mais elle découle du rapport de subordination. Le contrôle rapproché des salariés n’est certes pas nouveau (chronotachygraphe des routiers, pointeuse, etc.) ; mais, avec l’intelligence artificielle, il devient plus performant, invisible, exact et silencieux. Ces nouveaux outils permettent la collecte de vastes quantités de données en temps réel sur les travailleurs, pendant et en dehors des heures de travail, dans divers lieux de travail, voire à l’extérieur de ces lieux. Le suivi minute par minute de la position d’un salarié sur l’ensemble de sa journée peut conduire à des dérives. Les techniques qui sont mobilisées pour surveiller le travail sont de fait de plus en plus intrusives. Les données ainsi collectées par le dispositif d’IA seront bien évidemment utilisées pour orienter toute la vie professionnelle du salarié (propositions de promotion, de formation…).

Depuis le début de la crise de COVID-19 et l’essor du télétravail, de nouveaux procédés de surveillance ont été déployés pour encadrer les télétravailleurs et évaluer leurs performances. Pourtant, les risques d’atteinte à la vie privée des salariés découlant de ces systèmes d’intelligence artificielle semblent largement sous-estimés, alors que les capacités de collecte et de traitement de données supplantent celles de n’importe quel dispositif utilisé par le passé. Les sources de collecte de ces données sont variées, ce qui conduit en pratique à surveiller et contrôler toutes les activités des salariés.

L’IA serait en effet performante pour surveiller les communications des salariés en traitant des quantités significatives de données (contenu des courriels, clics de clavier, mouvements de souris, messages instantanés, connexions, etc.). Des entreprises américaines, telles que *Walmart*, *Delta*, *T-Mobile*, *Chevron* et *Starbucks*, utilisent, par exemple, l’intelligence artificielle pour surveiller les messages de leurs salariés sur des applications telles que *Zoom*, *Microsoft Teams*, *Slack*, etc. Ces données sont traitées par un système d’IA pour évaluer, entre autres, la productivité du travailleur, par un suivi du temps de travail ou du nombre de tâches accomplies sur une période déterminée.

Les données relatives aux travailleurs peuvent aussi être collectées par des appareils mobiles ou des dispositifs de surveillance portables ou intégrés dans les vêtements, dans les équipements de protection individuelle ou même sur le corps. Ces dispositifs peuvent enregistrer les mouvements des travailleurs, leur rythme de travail et leurs pauses et ainsi surveiller leurs déplacements sur le lieu de travail. La combinaison de ces données avec celles d’autres machines se trouvant dans le même espace de travail physique que les travailleurs et avec lesquelles il existe des interactions, telles que des cobots, augmente l’intensité de la surveillance.

Plus inquiétant encore, l’intelligence artificielle pourrait être utilisée pour mesurer l’aisance d’un salarié en fonction de son langage corporel ou sa communication. Grâce à la reconnaissance et à l’analyse des émotions exprimées par le visage, se développe ce que l’on appelle l’intelligence artificielle des émotions. EmoScienS, une start-up montréalaise, a, par exemple, développé un logiciel qui mesure automatiquement et en continu la santé émotionnelle des travailleurs devant leur ordinateur. À l’aide de la

caméra de l'écran, l'outil enregistre les expressions faciales des travailleurs, pour en déduire les émotions vécues au long de la journée. Grâce à un tableau de bord, le salarié peut ensuite cibler les émotions qu'il a éprouvées et les associer aux tâches qu'il a accomplies. Cette technologie peut être utilisée pour conseiller aux salariés le bon ton à adopter avec leurs clients. L'efficacité de l'intelligence artificielle des émotions est cependant critiquée, car elle repose sur l'idée contestable que les émotions humaines sont repérables grâce à des expressions faciales universelles.

Bien plus préoccupants, certains dispositifs actuellement mis en place visent à déterminer l'état de santé physique et mentale des travailleurs. Ces dispositifs portatifs sont généralement équipés de capteurs analysant des données de santé telles que le rythme cardiaque et la pression artérielle. Les entreprises ont ainsi accès aux données de santé en temps réel sous prétexte de contribuer à l'amélioration du bien-être des salariés. Or, les travailleurs peuvent finir par avoir un sentiment d'intrusion dans leur vie privée, ce qui peut être une source de stress et d'angoisse. Ils peuvent également subir des pressions pour augmenter leur productivité, ce qui peut conduire à des accidents et des problèmes de santé.

Ces nouvelles formes de surveillance des salariés suscitent bien évidemment de nombreuses questions juridiques et éthiques, ainsi que des préoccupations liées à la santé et la sécurité au travail.

Quelle que soit la nature de la collecte, directe ou indirecte, certaines informations doivent obligatoirement être fournies aux salariés au moment de cette collecte (finalités du traitement, durée de conservation, destinataires des données, etc.). Cette obligation de transparence doit ainsi permettre aux salariés de comprendre les raisons pour lesquelles leurs données ont été collectées et le traitement qui en sera fait. L'article 12 du RGPD précise à cet égard que l'information doit être réalisée « *de façon concise, transparente, compréhensible et aisément accessible, en des termes simples et clairs* ».

Dans tous les cas, l'employeur doit respecter le RGPD, ce qui exclut les dispositifs de surveillance non conformes à ce texte qui présenteraient des risques élevés pour les droits et libertés des salariés. Ces systèmes d'IA seront en outre considérés par le règlement européen sur l'intelligence artificielle (IA Act) comme faisant partie de la catégorie « à haut risque » et, par conséquent, devront être assortis de protections particulières qui peuvent ne pas être suffisantes au regard des enjeux et des risques.

57 Qui sont les travailleurs du clic ?

Les travailleurs du clic, « tâcherons » ou « micro-travailleurs » ont émergé au milieu des années 2000 avec des plateformes comme *Amazon Mechanical Turk*. Ces travailleurs ont donné naissance à une nouvelle catégorie de travailleurs : le *digital labor* ou la main-d'œuvre numérique.

Contrairement à une idée reçue, derrière l'intelligence artificielle se cachent des millions de travailleurs pauvres répartis sur tous les continents œuvrant à une multitude de micro-tâches utiles au fonctionnement des algorithmes des intelligences artificielles. Ils sont recrutés *via* des plateformes numériques proposant des micro-tâches courtes et répétitives externalisées par certaines sociétés pouvant s'effectuer devant un ordinateur ou sur un smartphone, telles qu'identifier des objets sur une image, des visages, des émotions, étiqueter des images ou des vidéos, classer des produits dans les catalogues en ligne, traduire des textes courts, etc. Ces plateformes proposent aux particuliers de monétiser leur temps libre en effectuant des tâches rémunérées à l'unité.

Cela permet de produire des exemples qui alimentent l'apprentissage automatique sur lequel est basée l'IA. Cette fonction de « formateur » est devenue importante avec le *machine learning* qui permet aux systèmes d'IA d'apprendre par eux-mêmes à reconnaître les modèles en analysant les données reçues. En effectuant des micro-tâches, ces travailleurs participent ainsi à la production de ces technologies. Ils enseignent, par exemple, aux dispositifs de reconnaissance vocale ou visuelle à interpréter des sons et des images ou encore nettoient les données et les enrichissent pour qu'elles soient utilisées dans l'apprentissage profond.

Ces travailleurs n'ont pourtant ni le statut d'indépendant ni celui de salarié, car la contrepartie financière est présentée comme une compensation ou un dédommagement. Ils seraient pourtant 260 000 en France et entre 45 et 90 millions au niveau mondial. Ces travailleurs sans contrat ni garanties sociales se voient également imposer le statut d'auto-entrepreneur. Aussi, les plateformes ne sont pas tenues de leur payer les arrêts maladie, les congés payés ou encore de leur garantir un salaire minimum et une protection sociale. Le microtravailleur est en effet payé à la tâche, travaille n'importe où et à n'importe quelle heure, pourvu qu'il y ait une connexion internet, et son profil est inconnu ou presque de son employeur.

Selon une étude de l'Organisation internationale du travail de 2018, le microtravail serait essentiellement un phénomène urbain, avec 4 travailleurs sur 5 vivant dans une zone urbaine, présent partout sur tous les continents, notamment au Brésil, à Madagascar, en Côte d'Ivoire, en Inde, en Indonésie, au Nigeria, aux États-Unis et en Europe.

Il est assez difficile d'estimer une rémunération moyenne, car elle varie en fonction des pays. Quelques rares micro-travailleurs gagnent quelques dizaines voire centaines de dollars par mois, mais c'est loin d'être le cas de tous ces travailleurs. On peut estimer qu'un travailleur du clic gagne en moyenne 2 dollars de l'heure, parfois moins.

Pour Antonio Casilli, professeur de sociologie à Télécom Paris et auteur du livre *En attendant les robots – Enquête sur le travail du clic*, paru aux éditions du Seuil en 2019, « *le plus grand tour de passe-passe de ces plateformes, c'est d'une part de faire croire*

aux consommateurs, [...], qu'il y a des processus automatiques, qu'il y a des algorithmes partout, alors que très souvent il s'agit de tâches réalisées à la main. Et d'autre part, de faire croire aux travailleurs que ce qu'ils réalisent n'est pas un vrai travail, mais plutôt un job, ou un gig (en anglais), qu'il s'agirait là de quelque chose de transitoire et éphémère, et qui, à terme, va disparaître ».

L'économie à la tâche marque un retour au taylorisme du siècle passé. On observe ainsi une nouvelle division du travail à l'œuvre ; d'une part, les experts qualifiés (ingénieurs, développeurs, etc.) travaillant pour les grandes entreprises de la Silicon Valley, et d'autre part, tous ces travailleurs précaires formant un « cyber-prolétariat ». Ces travailleurs du clic sont pourtant ignorés par les médias et les pouvoirs publics, alors que cette nouvelle forme de travail est de plus en plus fréquente et échappe au cadre légal, hormis peut-être les impôts que ces travailleurs paient sur ces revenus.

Au-delà de la question de la précarisation, certains de ces travailleurs sont exposés à des contenus violents ou traumatisants, tels que des images pédopornographiques ou de violences sexuelles afin de trier celles qui ne respectent pas les règles des plateformes (par exemple, les vidéos YouTube). Certains peuvent ainsi subir une forme de stress post-traumatique.

Le sort de ces travailleurs de l'ombre n'a pas été malheureusement pris en compte par le règlement européen sur l'intelligence artificielle (IA Act), mais d'autres textes pourraient à terme étendre la protection du droit du travail à ce type de travailleurs, tels que la future directive sur les conditions de travail des travailleurs de plateformes (cf. question n ° 58). Il est vrai que cette directive vise à traiter le cas des travailleurs des plateformes les plus visibles, notamment les livreurs et les chauffeurs de VTC, mais elle pourra également être invoquée par ces travailleurs du clic, si les conditions sont réunies afin qu'ils puissent se prévaloir du statut de salarié. Plus généralement, il conviendrait que les pouvoirs publics s'emparent plus sérieusement de cette question afin de promouvoir une « IA socialement responsable » qui garantit le respect des droits de ces travailleurs de l'ombre. Le développement actuel de l'IA ne fonctionne pas sans travail humain précarisé et insensibilisé à grande échelle. Loin de remplacer l'homme, l'IA engendre au contraire de nouvelles formes d'exploitation de ce dernier.

58 Les plateformes numériques de travail utilisent-elles l'IA ? Les travailleurs des plateformes numériques sont-ils protégés par le droit du travail ?

Le recours à des systèmes d'IA a une influence directe sur les relations de travail, puisqu'il se trouve au cœur d'un nouveau modèle déstabilisateur, à savoir l'ubérisation, qui est celui des plateformes numériques dont le fonctionnement est intrinsèquement lié aux algorithmes. On dénombre environ 500 plateformes de travail numériques exerçant leurs activités dans l'Union européenne et on estime que plus de 28 millions de personnes travaillent déjà par l'intermédiaire d'une plateforme dans l'Union européenne. Ce chiffre devrait atteindre 43 millions de personnes d'ici 2025. Les personnes travaillant *via* des plateformes de travail numériques exécutent une grande variété de tâches. Il peut s'agir, par exemple, de taxi, de services de livraison, de traduction, de garde d'enfants, de soins aux personnes âgées, etc.

Leur fonctionnement se révélant déstabilisateur, ces plateformes numériques suscitent une attention croissante des pouvoirs publics, tant au niveau international que national. En effet, la majorité des travailleurs des plateformes de l'UE sont officiellement des travailleurs indépendants. À ce titre, ces plateformes de travail en ligne reposent pour la plupart sur le même modèle : il n'y a pas de relation salariée entre les travailleurs et la plateforme elle-même. Si le droit du travail s'applique aux travailleurs des plateformes qualifiés de salariés, d'autres sont qualifiés d'indépendants et sont de fait exclus de la protection que prévoit la législation sociale. Plus de 90 % des plateformes concernées ne les considèrent d'ailleurs pas comme des travailleurs salariés mais indépendants.

Les réponses nationales à cette forme de travail sont diverses et se développent de manière inégale en Europe. C'est la raison pour laquelle les instances européennes se sont emparées de la question, ce qui a abouti à une proposition de directive concernant les conditions de travail des travailleurs de plateformes, publiée par la Commission européenne, le 9 décembre 2021. Le projet a engendré des négociations compliquées pendant deux ans, mais un accord a finalement été trouvé et le Parlement européen a adopté la directive, le 24 avril 2024. Le 14 octobre 2024, le Conseil de l'Union européenne a approuvé à son tour la directive. Après sa publication au *Journal officiel* de l'UE le 11 novembre 2024, les États membres auront deux ans pour incorporer les dispositions de la directive dans leur législation nationale.

L'apport fondamental de ce texte est qu'il établit que ces travailleurs seraient présumés être des salariés d'une plateforme numérique (et non des travailleurs indépendants) si les faits indiquent la présence d'un contrôle et d'une direction sur le travailleur. La plateforme pourra renverser la présomption par la preuve de l'absence de relation de travail. La charge de la preuve incombera à la plateforme, ce qui signifie que lorsqu'une plateforme voudra réfuter la présomption, ce sera à elle de prouver que la relation contractuelle n'est pas une relation de travail.

Moins médiatisée et pourtant novateur, la directive prévoit aussi un encadrement du management algorithmique. Les plateformes de travail numériques utilisent en effet des algorithmes pour organiser et gérer les personnes exécutant un travail *via* celles-ci par l'intermédiaire de leurs applications ou de leurs sites web. Ces algorithmes développés et utilisés par les plateformes numériques sont au cœur de leur modèle économique. Ils sont multiples et répondent à des usages variés, dont le niveau de complexité et l'importance stratégique dépendent du modèle économique souhaité par la plateforme. Les algorithmes et l'exploitation des données collectées qu'ils permettent constituent, à ce titre, des aides à la décision et à la gestion des travailleurs.

Dans le cadre de relations contractuelles entre des plateformes numériques et des travailleurs indépendants, le management algorithmique aboutit à un renforcement de leur évaluation, de leur contrôle et de leur surveillance. Ce management algorithmique donne d'une certaine façon aux plateformes des outils pour exercer certaines prérogatives similaires à celles de l'employeur dans le cadre d'un contrat de travail salarié. En vertu des nouvelles règles, les travailleurs des plateformes devront être dûment informés de l'utilisation de systèmes de surveillance ou de prise de décision automatisés en ce qui concerne leur recrutement, leurs conditions de travail et leur rémunération.

Les nouvelles règles prévoit une intervention humaine obligatoire dans les décisions de limitation, de suspension ou de résiliation de la relation contractuelle ou du compte d'une personne exécutant un travail *via* une plateforme, quel que soit son statut. Les travailleurs des plateformes pourront d'ailleurs contester les décisions prises ou appuyées par un système de décision automatisé.

La directive introduit enfin des mesures pour mieux protéger les données à caractère personnel des travailleurs des plateformes. Il sera interdit aux plateformes de travail numériques de traiter certains types de données personnelles, tels que les données relatives à l'état émotionnel ou psychologique du travailleur, ses conversations privées, ses opinions politiques, etc. L'intérêt de ces dispositions sur l'usage des algorithmes est qu'elles pourraient servir de modèle pour un encadrement plus général de l'utilisation des algorithmes dans les relations de travail.

Cette directive devrait conduire à une amélioration des conditions de travail et la protection sociale des travailleurs des plateformes numériques en Europe, dont plusieurs millions pourraient être requalifiés en salariés. Même si tous les travailleurs ne vont pas bénéficier du statut de salarié, la directive devrait conduire à remettre en cause les faux statuts de travailleurs indépendants et promouvoir une meilleure protection pour tous les travailleurs des plateformes.

59 L'IA va-t-elle entraîner la fin du salariat ?

Dans son livre « *Disruption : préparez-vous à changer de monde* », paru en 2018, Stéphane Mallard prédit la fin du salariat. Il dénonce ce que l'anthropologue américain, David Graeber, qualifie de « *bullshit jobs* », soit tous les métiers qui ont été créés dans le secteur tertiaire, l'administration, la bureaucratie et le support à la production : les consultants, les contrôleurs de gestion, etc.

Il prône l'indépendance et loue le phénomène des *digital nomads* qui rejette les contraintes du salariat, alors qu'il existe des outils pour proposer sa force de travail directement au marché sans passer par le pesant lien de subordination. Il décrit un monde dans lequel être salarié sera synonyme d'incompétence dans la mesure où les meilleurs travailleurs choisiront grâce à la technologie l'indépendance aux contraintes du salariat. Il annonce la fin inéluctable de l'ancien monde et la généralisation des indépendants, car leur modèle économique serait plus adapté à notre époque et leurs performances directement mesurables et quantifiables.

L'équation subordination-protection qui fondait le salariat se heurte en effet aux réalités économiques comme aux aspirations sociales. Si les travailleurs aspirent à davantage d'indépendance, la précarité et l'exclusion risquent d'être le prix payé par ceux qui ne parviennent pas à s'adapter à cette nouvelle réalité.

Le terme « ubérisation » est d'ailleurs entré dans le langage courant. Ce dernier fait référence aux plateformes numériques qui sont devenues de véritables places de marché virtuelles, facilitant la rencontre entre l'offre et la demande de biens ou de services. Ce nouveau modèle économique est incarné par l'entreprise californienne de transports par VTC Uber, qui a transformé le marché traditionnel des taxis en proposant aux usagers une plateforme de réservation en ligne, dont le fonctionnement permet d'échapper à certaines contraintes du secteur ; par exemple, en France, à l'achat de licence de taxi et aux contraintes découlant de l'usage de cette licence. Cette dernière est à la fois l'incarnation et l'épouvantail de cette « plateformisation » de l'économie. L'ubérisation annoncerait « la mort du salariat » et une façon différente de vivre son activité professionnelle. Quoi qu'il en soit, l'ubérisation marque l'accroissement du travail non-salarié dont la part dans l'emploi total diminuait depuis de longues décennies. De fait, il s'agit de la fin du salariat comme modèle unique.

Le salariat ne va pas pour autant disparaître, mais progressivement se transformer, notamment sous l'effet des technologies de l'information et de la communication. Le télétravail est un élément emblématique de cette transformation de l'organisation de l'entreprise et du rapport à l'emploi. La généralisation du télétravail lors de la crise sanitaire a accéléré cette transformation et l'a ancrée dans les mœurs.

L'essor de l'intelligence artificielle réactualise la question récurrente des effets des technologies sur les activités humaines. Cette question n'est certes pas nouvelle. La révolution industrielle, puis l'automatisation et la révolution informatique ont déjà fait disparaître de nombreux emplois manuels dans l'industrie et l'agriculture. Si certains s'alarment de cette transformation, d'autres y voient une promesse d'émancipation.

Il est certain que l'automatisation de certaines tâches pourrait avoir un effet négatif sur l'emploi dans les secteurs où elles seraient réalisables par l'intelligence artificielle. Les travailleurs du savoir et des professionnels hautement qualifiés, autrefois considérés comme à l'abri de l'automatisation, voient aussi leur métier menacé par cette avancée technologique. La plupart des emplois qualifiés vont donc se retrouver en concurrence avec l'intelligence artificielle, ce qui risque d'entraîner le déclassement de nombreux professionnels qualifiés, tels que les professions libérales.

GPT-4 a, par exemple, réussi avec succès l'examen du barreau américain, obtenant une note qui l'aurait classé parmi les 10 % des meilleurs étudiants. Ce résultat est toutefois à relativiser, car le robot conversationnel d'*OpenAI* a échoué à l'examen du Barreau du Québec n'étant alimenté que par le droit américain. Les téléconsultations de médecins durant lesquelles ils posent des questions et prescrivent des traitements ou examens standards pourront, par exemple, être aussi en grande partie réalisées par une IA qui pourra effectuer la même prestation en coûtant moins cher.

L'IA, en particulier l'IA générative, apporte ainsi des changements radicaux qui promettent d'augmenter la productivité des entreprises, mais risque aussi de bouleverser l'organisation du travail telle que nous la connaissons. Si, pour l'instant, son développement ne semble pas avoir provoqué l'hécatombe de l'emploi, il pose néanmoins de nombreuses questions quant à l'évolution future de l'emploi et de son cadre juridique.

60 ChatGPT va-t-il remplacer de nombreux salariés ?

GPT, acronyme de *Generative Pre-trained Transformer*, est ce que l'on appelle un modèle de langage. Le modèle GPT-1 a été déployé en 2018 par la société *OpenAI*. Il a initialement été conçu dans l'objectif de générer du texte cohérent et parfaitement fluide selon un contexte donné. Ce modèle a montré la puissance du *deep learning* en utilisant une architecture de réseau de neurones appelée « *Transformer* » pour générer du texte cohérent en fonction d'un contexte donné.

En 2019, GPT-2 a été publié de façon limitée au grand public pour des problématiques d'éthique et d'utilisation abusive ; le public n'étant pas encore préparé à ce bouleversement à cette époque. Cette seconde version a tout de même été appréciée par les utilisateurs en raison de ses capacités de génération de texte réalistes et cohérentes. Il pouvait déjà être utilisé pour créer du contenu de tout type tel que du code informatique, des courriels, etc.

GPT-3 a été ensuite déployé en 2020, il devint le modèle le plus vaste de la série GPT, avec pas moins de 175 milliards de paramètres. Par la suite, GPT-4 a été officiellement dévoilé par *OpenAI* en mars 2023. Il s'agit de son dernier modèle d'intelligence artificielle, doté de la capacité d'interpréter à la fois du texte et des images.

GPT et ChatGPT sont tous deux des modèles de langage développés par *OpenAI* et basés sur l'architecture *Transformer*. Ils furent conçus pour traiter et générer du texte. ChatGPT a subi simplement un entraînement supplémentaire par rapport à GPT, notamment sur des articles de presse, des sites web, des livres, des discussions en ligne, etc., afin de lui permettre d'interagir de façon plus naturelle avec les utilisateurs.

Le déploiement de ChatGPT à la fin de l'année 2022 a marqué un tournant majeur dans la démocratisation de l'intelligence artificielle. Le monde fut stupéfait par cet IA générative et inquiet de ses répercussions sur le marché de l'emploi.

Dans un rapport de 130 pages remis Président de la République, le 13 mars 2024, la Commission de l'intelligence artificielle, créée en 2023 et rassemblant des acteurs de différents secteurs (culturel, économique, technologique, recherche…), a rendu ses premières conclusions concernant l'impact de l'IA sur la croissance, l'économie et l'emploi. Selon ces experts, l'intelligence artificielle devrait avoir un impact global positif sur le marché de l'emploi. Selon eux, les bienfaits de l'IA sur la productivité dépasseraient largement les effets négatifs sur l'emploi. Au-delà de la disparition de certains postes, une grande partie des métiers pourrait voir leur productivité augmenter, permettant ainsi aux entreprises de se développer plus rapidement et d'embaucher de nouveaux salariés.

Ils précisent tout de même que les effets spécifiques de l'IA générative sur l'emploi sont difficiles à évaluer, faute de recul. Ils s'accordent à dire que certains métiers commerciaux ou administratifs pourraient ne pas survivre, citant les secrétaires, les comptables ou les télévendeurs. Les métiers de la connaissance, de la stratégie et de la créativité, autrefois perçus comme des creusets de l'intelligence humaine, tels que les journalistes, les médecins, les avocats ou même les artistes, pourraient également faire face à une forte réduction du nombre d'emplois.

Le rapport indique que les emplois directement remplaçables par l'IA ne représenteraient que 5 % des emplois en France. Un chiffre comparable aux conclusions d'une étude d'août 2023 sur l'impact de l'IA générative de l'Organisation internationale du travail (OIT). L'étude constate que 5,5 % de l'emploi total dans les pays à revenu élevé est potentiellement exposé aux effets d'automatisation de la technologie, alors que dans les pays à faible revenu, le risque d'automatisation ne concerne qu'environ 0,4 % de l'emploi. Le travail de bureau s'avère être la catégorie la plus exposée à cette évolution.

Si l'OIT se veut rassurante et mesurée en parlant plutôt de transformation et d'augmentation du travail que d'automatisation et de remplacement, cette vision optimiste ne fait pourtant pas l'unanimité. Le Forum Économique Mondial estime que l'IA va supprimer 83 millions d'emplois, Goldman Sachs s'attend à 300 millions d'emplois remplacés, et le FMI évoque l'automatisation de 60 % des emplois.

Pourtant, une équipe de chercheurs du MIT a publié une étude, le 22 janvier 2024, qui conclut, pour la grande majorité des métiers précédemment identifiés comme menacés, que l'automatisation ne serait pas forcément économiquement bénéfique pour les employeurs à l'heure actuelle. Dans de nombreux cas, il reste plus économique pour un employeur de payer un humain afin de réaliser les tâches considérées pourtant comme exposées. C'est la raison pour laquelle l'IA n'a pas été déployée à grande échelle.

Cette étude dédramatise les effets sur l'emploi dans la mesure où les effets massifs de l'IA ne s'avèrent pas imminents, mais progressifs. Le déploiement de l'IA générative pourrait tout de même conduire progressivement à une transformation importante et est loin d'être une mutation bénigne du travail intellectuel. Si les compétences des salariés expérimentés sont assimilées par l'IA, cela pourrait conduire à une dévalorisation générale du travail intellectuel et à un nivellement des salaires par le bas, similaire à ce qu'ont déjà connu une partie des travailleurs manuels au cours des dernières décennies.

Certains emplois tireront cependant parti de l'IA générative ou seront même créés, notamment les postes de développeur, de *data scientist* (spécialiste de la collecte et de l'analyse de données), d'analyste financier spécialisé en IA, de *Prompt engineer* dont la mission est d'optimiser les requêtes et les résultats de l'IA générative, mais aussi de responsable éthique IA qui sera chargé d'évaluer les risques éthiques de ces nouveaux outils. Il y aura donc des perdants et des gagnants lors de cette transition.

VII.
L'INTELLIGENCE ARTIFICIELLE ET LA SÉCURITÉ

61 Qu'est-ce qu'une police prédictive ?

Contrôler une ville à partir d'une salle de commandement, utiliser l'intelligence artificielle pour cibler les situations ou les comportements jugés suspects, ou pour orienter les patrouilles vers les zones où de futurs délits se produiront, telles sont les ambitions de la police prédictive. La prédiction des crimes n'est pourtant pas un thème nouveau. Les techniques d'évaluation des risques par une approche statistique, notamment utilisées dans les assurances, ont été transposées au pénal, dès les années 1920, par le sociologue américain Ernest Burgess, l'un des fondateurs de l'École de Chicago, qui cherchait, entre autres, à prédire les comportements déviants en se fondant sur des variables statistiques.

L'idée de la prédiction de la criminalité existe donc depuis longtemps dans le système judiciaire et pénal. Dans leur mission d'enquête et de prévention de la criminalité, les services de police ont toujours eu besoin de collecter et traiter des données relatives aux personnes. Cette approche prend cependant une autre ampleur sous l'ère du *big data* et de l'intelligence artificielle qui promettent des réponses plus exhaustives et efficaces.

Le terme de police prédictive – mauvaise traduction de *predictive policing* – est entré en France dans le langage courant. Lorsqu'il s'agit de déterminer concrètement ce qui est développé par les services de l'État au nom de cette police prédictive, il est souvent difficile de savoir ce qu'elle recouvre. Elle est souvent considérée comme un simple outil d'aide à la décision. La notion de « police prédictive » est d'ailleurs à utiliser avec précaution, on lui préfère celle d'analyse décisionnelle. Dans une acception large, elle recouvre tous les dispositifs intelligents annoncés comme prédictifs et conçus pour renforcer la sécurisation et le contrôle des espaces publics. Les forces de police s'intéressent fortement à tous ces outils leur permettant d'anticiper les crimes et délits.

Dans le contexte américain, le *predictive policing* s'inscrit pleinement dans l'histoire des réformes visant à rendre la police plus proactive et vigilante que réactive et urgentiste, soit une police plus engagée dans la production de la sécurité que dans la répression des criminels. À partir des années 1980-1990, de nouveaux modèles de la police ont été imaginés dans ce sens. En collaboration avec des polices locales, des travaux de recherche ont en effet expérimenté des nouvelles stratégies de lutte contre la délinquance. La police devient alors un champ de réformes. Le *community policing* ou police communautaire, le *hot spot policing* ou police des points chauds, le *problem-oriented policing* ou police de résolution des problèmes et l'*intelligence-led policing* ou police guidée par le renseignement font partie des stratégies policières les plus célèbres de l'histoire de la réforme des polices américaines. Le *predictive policing* s'inscrit dans la continuité de ces stratégies de l'action policière.

Les États-Unis ont d'ailleurs longtemps figuré comme les plus innovants dans ce domaine. Les villes américaines, notamment Los Angeles et New York, ont été les premières à tester des outils prédictifs. Le premier logiciel prédictif conçu et testé fut *Compstat* – abréviation de « *comparative statistics* », imaginé et déployé à New York dans les années 1990 et à Paris au début des années 2000, afin de faire ressortir les lieux les plus criminogènes pour permettre à la police d'adapter son action. En collectant et en

analysant presque en temps réel les données relatives au crime et à la délinquance, la police devient ainsi capable de discerner des régularités, des séries, et d'anticiper l'activité criminelle au lieu de réagir *a posteriori*. L'objectif des algorithmes prédictifs est d'anticiper ce qui peut se produire par une utilisation plus efficace des données historiques, auxquelles sont agrégées des informations d'actualité. La « prédiction du crime » marque un changement de paradigme qui bouleverse les modes d'action et les ambitions des polices.

PredPol, fondé sur un modèle sismologique, – rebaptisé Geolitica depuis 2021 – reste, à ce jour, une des références en la matière de cette police prédictive. En réalité, il s'agit d'un véritable marché ayant permis de développer d'autres logiciels de police prédictive, tels que Hunchlab, Palantir ou encore Paved.

Utilisés au départ afin de faire remonter les statistiques sur les faits de délinquance dans l'objectif d'améliorer la réactivité de la police et ses capacités d'anticipation, ces outils se sont rapidement transformés en instruments d'une « politique du chiffre ». Ils apparaissent alors comme des dispositifs de quantification des performances et de mise en compétition des policiers pour atteindre les meilleurs résultats (nombre d'arrestations, taux d'élucidation des crimes, etc.). La police prédictive apparaît ainsi comme une entreprise de rationalisation de l'action de la police afin d'augmenter sa « productivité ».

En dépit de l'efficacité discutable de ces dispositifs, les promoteurs de ces technologies entretiennent toujours la croyance selon laquelle l'intelligence artificielle pourra améliorer l'efficacité de la police en lui permettant d'anticiper les crimes grâce à des données empiriques. Ce postulat repose sur des théories selon lesquelles les comportements criminels ne se produisent pas de manière aléatoire et présentent un *pattern* prévisible qu'il convient d'identifier. Or, la majorité des modèles prédictifs concerne les vols et cambriolages qui surviennent fréquemment donnant lieu à des plaintes. Ces modèles prédictifs sont cependant difficilement applicables aux infractions peu dénoncées, tels que les viols ou les violences domestiques. L'insuffisance des données peut ainsi influencer le choix des infractions analysées. Aussi, certaines précautions doivent être adoptées lorsque de tels outils sont utilisés par la police. Il serait, par exemple, risqué de les utiliser comme seul moyen pour interpeller les personnes.

62 Un algorithme est-il capable de prédire les crimes et les délits ?

Bien avant le film *Minority Report*, les autorités ont cherché à trouver des solutions pour anticiper les crimes et, éventuellement, les éviter. L'idée fondamentale associée à *Minority Report* consiste à décider ou agir sur le fondement d'une prédiction. Si la police prédictive s'est fait connaître avec le logiciel PredPol, largement utilisé par la police américaine pour orienter les patrouilles, son origine est cependant plus ancienne. En effet, elle remonte aux années 1990 qui ont connu le développement des outils de gestion de la performance policière. C'est dans ce contexte que les données des faits de délinquance ont été exploitées pour établir des actions, fixer des objectifs ou encore optimiser les ressources de la police.

La méthode prédictive repose sur trois facteurs permettant de quantifier statistiquement le risque qu'un acte de délinquance soit commis : le type de délinquance, le lieu et le moment. Ces données proviennent des statistiques de la délinquance précédemment enregistrées. La mise en place de telles méthodes nécessite de disposer de vastes bases de données sur la criminalité passée. Lorsque ces bases de données sont suffisamment importantes, il devient possible d'identifier des « patterns » dans la mesure où les schémas de la criminalité ont tendance à se répéter. La ville ou la zone géographique est ensuite modélisée et divisée en zones.

Les algorithmes de police prédictive testés sur le terrain font régulièrement l'objet de nombreuses critiques. Ils sont tout d'abord jugés inefficaces et ne correspondant pas aux résultats exceptionnels annoncés par les entreprises qui les éditent. Ils entraînent généralement une présence policière excessive dans certaines zones fortement criminogènes, ce qui n'a rien de surprenant puisqu'il s'agit de l'application de la loi de Pareto, datant du XIX^{e} siècle, selon laquelle 20 % des causes produisent 80 % des effets. Autrement dit, 80 % des faits se produisent dans 20 % du territoire. En cela, ces algorithmes de police prédictive ne sont pas révolutionnaires, car les policiers avaient déjà identifié ces lieux et y concentraient déjà leur action.

De ce fait, ces algorithmes confirment la stratégie de la police et l'intensifient, mais créent aussi un cercle vicieux dans la mesure où les profils des personnes arrêtées valideront la prédiction initiale, car les données servant à établir les statistiques de la délinquance sont le reflet des pratiques des forces de police. De nombreuses patrouilles seront envoyées dans les zones perçues à risque par l'algorithme, ce qui permettra de constater des infractions et collecter des données concernant ces zones qui seront, à leur tour, prises en compte par l'algorithme afin de renforcer la présence de la police dans ces dernières. L'algorithme risque donc de créer un effet d'auto-renforcement de l'action de la police et une surenchère de dispositifs de contrôle dans certaines zones.

Par ailleurs, ces algorithmes ont tendance à se concentrer sur certains types de délits, tels que les cambriolages, pour lesquels les données sont plus nombreuses que, par exemple, le blanchiment d'argent. L'un des premiers algorithmes expérimentés en France dans l'Oise en 2015, PredVol, évaluait, à ce titre, le risque de vol de véhicules. L'outil n'a pas été jugé pertinent et a été finalement abandonné.

Paved est un autre outil, développé à partir de 2017 par la Gendarmerie et expérimenté à partir de 2018 dans différents départements métropolitains, pour évaluer le risque de cambriolages et de dégradations de véhicules. En raison de sa faible capacité de prévision ne s'étant pas traduite par une hausse du nombre d'arrestations en flagrant délit, l'outil a également été abandonné alors que sa généralisation était prévue sur l'ensemble du territoire. Malgré les nombreuses promesses des entreprises qui éditent ces logiciels de police prédictive, il semblerait qu'ils peinent à s'imposer dans la pratique. Ces outils de police prédictive sont peu à peu abandonnés également aux États-Unis, à commencer par Geolitica – le nouveau nom du célèbre logiciel américain PredPol – en raison des doutes sur leur efficacité réelle et parce qu'ils ne font que confirmer l'existence de zones criminogènes déjà connues.

Les nombreuses études ayant évalué l'efficacité de ces outils n'ont d'ailleurs pas établi qu'ils avaient un impact significatif sur la réduction de la criminalité. Ces outils soulèvent également de nombreuses questions éthiques non résolues, notamment le problème des biais. La conception et le développement des systèmes de police prédictive sont souvent vantés comme étant neutres et objectifs, alors que les données utilisées pour entraîner l'algorithme comportent des préjugés et des stéréotypes. Les scores de risque peuvent être en effet corrélés à certaines variables socio-démographiques (le taux de chômage ou encore l'origine ethnique), ce qui conduit à cibler certaines catégories de la population, augmentant ainsi les risques de discrimination.

Il est certain que la police prédictive pourrait améliorer le processus de prise de décision et aboutir à des décisions moins biaisées si une attention particulière était accordée à la sélection et l'analyse de la qualité et de la quantité des données. C'est la disponibilité de ces données qui rend possible ces prédictions algorithmiques. Pour autant, faut-il développer de tels outils seulement parce que les données sont disponibles ou en fonction de stratégies policières ?

Les algorithmes de police prédictive sont porteurs d'une promesse d'efficacité de la police, mais ce gain ne doit pas se faire au détriment des libertés individuelles. Ces outils doivent avant tout être au service de la sécurité publique et leur utilisation mérite d'être soumise à un débat.

63 L'IA peut-elle améliorer la sécurité des villes intelligentes ?

La ville intelligente, également connue sous le nom de *smart city*, est un concept apparu il y a une dizaine d'années, qui s'inscrit dans le prolongement de celui de la ville durable. La transformation d'une ville en *smart city* a souvent pour objectif de résoudre des problèmes de surpopulation, de congestion des moyens de transport et aussi à réduire l'empreinte carbone.

Ce concept de ville intelligente est certes imprécis, mais il a le mérite de mettre en avant les conséquences de la révolution numérique sur les villes. Ce concept a d'ailleurs été porté par des entreprises leaders du numérique, souhaitant apporter des solutions technologiques aux problèmes de densification urbaine. D'origine anglo-saxonne, il n'est pourtant pas nouveau. Les pionnières dans le domaine sont les mégalopoles d'Asie telles que Hong Kong ou Singapour.

Les projets de villes intelligentes se multiplient un peu partout dans le monde. Leur caractéristique commune est le recours à la technologie et aux données numériques afin d'améliorer la prise de décision et la qualité de vie des citoyens. Ces villes « hyper-connectées » cherchent dans la maîtrise de l'information les solutions aux problèmes qu'elles traversent. Il s'agit donc de villes qui collectent une masse importante de données et qui les analysent, le plus souvent en temps réel, pour améliorer sa gestion. Une *smart city* est en effet une ville où la plupart des objets urbains sont connectés. Une telle ville s'appuie sur une multitude de capteurs et de caméras disséminés dans le mobilier urbain, les habitations et les divers réseaux permettant de mesurer, au fil de l'eau, un certain nombre d'indicateurs tels que la qualité de l'air, l'état du trafic routier, la gestion des déchets… Elle est dite « monitorée ».

L'internet des objets ou l'IoT (*Internet of Things*) se retrouve au centre des infrastructures en créant de gigantesques maillages d'objets connectés. Aussi, le défi de la *smart city* est d'interconnecter tous ces réseaux et de les faire travailler en synergie. Ces données collectées sont ensuite analysées par des systèmes d'intelligence artificielle. Ces technologies modifient le fonctionnement des villes et promettent d'améliorer la qualité de vie des citadins et de mieux gérer leurs ressources ainsi que les infrastructures, par exemple en collectant et analysant les données relatives aux transports afin de rationaliser le trafic. Ces villes deviennent ainsi une zone de convergence de différents systèmes d'information de sources variées, avec différentes briques technologiques (5G, IoT, IA), différents équipements (mobilier urbain, feux de signalisation, etc.) et différentes interfaces (applications mobiles, etc.), ce qui les rend inévitablement vulnérables aux cyberattaques. Par conséquent, garantir la cybersécurité dans le monitoring de ces villes constitue un enjeu majeur.

Parallèlement à la *smart city*, est apparue la notion de *safe city*, une ville plus sûre grâce à la technologie. Après l'attentat de la promenade des Anglais à Nice, la ville a décidé de se lancer dans le projet *safe city* ayant inspiré d'autres villes.

Ces villes deviennent de fait des terrains d'expérimentation privilégiés de diverses technologies à visée sécuritaire, telles que la reconnaissance faciale dans l'espace public ou encore la vidéoprotection intelligente qui désigne des dispositifs équipés de logiciels

d'intelligence artificielle afin de détecter des formes ou des objets, d'analyser des mouvements, etc. L'automatisation de la surveillance grâce à ces caméras permet de repérer les comportements suspects, la détection d'objet abandonné, d'un vol, etc. Les images captées par les caméras sont analysées par des algorithmes entraînés à détecter, en temps réel, des situations prédéfinies, ce qui facilite et accélère l'identification des personnes, des objets ou des dangers pour les forces de l'ordre.

Le champ des possibles en matière de sécurité a été atteint par la loi du 19 mai 2023 relative aux Jeux olympiques et paralympiques de 2024 qui a autorisé, de manière temporaire et à titre expérimental, le recours à des caméras « augmentées » permettant la vidéosurveillance algorithmique, fixe ou installées sur des drones, en temps réel, des espaces olympiques, sportifs et récréatifs, de leurs abords et dans les transports publics. Le législateur a voulu être ici doublement prudent en faisant d'abord le choix de l'expérimentation, de sorte que ce n'est qu'après un « bilan coûts-avantages » du recours aux « caméras intelligentes » que le législateur pérennisera ou non ces dispositions législatives. Prudent, ensuite, car de nombreuses limitations encadrent ce dispositif inédit.

Malgré ces garanties, le recours aux caméras « augmentées » suscite des craintes, car il pourrait s'agir d'une décision sans retour. L'exécutif a d'ailleurs indiqué vouloir généraliser ce dispositif controversé de vidéosurveillance algorithmique mis en place lors de cet événement et dont l'expérimentation arrive à son terme le 31 mars 2025. Cette annonce intervient pourtant avant la remise du rapport d'un comité d'évaluation prévu par la loi. En devenant le premier État membre de l'Union européenne à autoriser à titre « expérimental » la surveillance assistée par intelligence artificielle pendant les Jeux olympiques de 2024, la France pourrait ouvrir la voie à la normalisation et la généralisation d'outils permettant une surveillance de masse.

Si la reconnaissance faciale a été explicitement exclue de cette expérimentation, rien ne permet de garantir qu'elle ne sera pas la prochaine étape. La vidéosurveillance algorithmique, autorisée par ce dispositif, permet seulement de détecter et signaler des comportements suspects. La reconnaissance faciale, quant à elle, permet d'identifier des personnes en direct en scannant et en croisant leurs visages avec une base de données. La politique des petits pas adoptée par les pouvoirs publics fait craindre un glissement vers une surveillance de masse et de graves atteintes aux droits fondamentaux. Le Sénat avait d'ailleurs adopté en première lecture, 12 juin 2023, une proposition de loi, toujours en cours d'examen, destinée à expérimenter pour une durée de trois ans le recours à la reconnaissance faciale dans l'espace publique. La ville intelligente et sécuritaire présente le risque d'un basculement vers une réalité dystopique.

64 L'IA peut-elle aider à résoudre les « *cold cases* » ?

Les *cold cases* sont des affaires criminelles qui ne sont pas encore élucidées. Ces *cold cases* passionnent les médias et alimentent la fiction à travers des séries policières en tout genre. Ces affaires manquent souvent de pistes ou ont épuisé toutes les voies d'enquête disponibles. Il peut s'agir de meurtres non élucidés, de cas de personnes disparues ou même des restes non identifiés. Ces affaires peuvent s'étendre sur plusieurs années, voire plusieurs décennies, ce qui rend leur résolution encore plus difficile. Une enquête criminelle génère de nombreuses données : collectes d'indices physiques, de très nombreux témoignages, actes procéduraux, éléments de médecine légale, etc.

L'un des principaux obstacles à la résolution de ces affaires est la perte de mémoire des témoins, voire leur décès. La perte de preuves constitue un autre obstacle. Les preuves matérielles peuvent en effet se dégrader ou être égarées. Les échantillons d'ADN, par exemple, peuvent se dégrader ou être contaminés, ce qui les rend inutilisables. Ces pertes peuvent considérablement entraver l'enquête.

Par ailleurs, les enquêtes les plus complexes qui n'ont pas été résolues comprennent un nombre très important de pièces dont il est parfois difficile d'extraire des liens. L'utilisation de logiciels d'analyse criminelle pour la résolution des affaires complexes n'est pas une nouveauté puisque les premiers outils de ce type ont été utilisés depuis 1994 par les services de la gendarmerie nationale.

La problématique du traitement de la donnée judiciaire de masse n'est pas tout à fait nouvelle. En 2017, le logiciel *Anacrim* a connu une renommée éphémère en permettant une avancée significative dans l'affaire Grégory, l'un des plus célèbre *cold case* de France.

Conçu au milieu des années 1990 pour la Gendarmerie nationale, *Anacrim* est un logiciel se composant de quatre modules : ATRT pour l'exploitation automatisée de relevés bancaires et des données téléphoniques ; ANB pour l'analyse et la représentation visuelle des données ; IVC pour l'identification des victimes de catastrophe ; et Mercure pour l'analyse des données téléphoniques obtenues sur réquisition.

Ce logiciel, développé par IBM, est utilisé par des gendarmes spécialement formés, appelés analystes criminels ou « Anacrim », pour le traitement des affaires les plus complexes. Ces spécialistes transforment le contenu des procédures en base de données puis en graphes, ce qui permet d'avoir une vision d'ensemble de dossiers qui représentent des milliers de pièces de procédure et de mettre en évidence incohérences et connexions.

Anacrim leur permet de traiter un grand nombre de données différentes et de faire des recoupements : téléphonie, comptes bancaires, procès-verbaux, analyses ADN, etc. Il ne cesse de s'améliorer. Cependant, la donnée massive est désormais partout et la question de son traitement n'est plus limitée aux dossiers complexes, mais devient un problème quotidien pour les enquêteurs. Or, l'un des aspects les plus intéressants de l'utilisation de l'IA réside dans sa capacité à analyser d'importants volumes de données en un temps record. En effet, les algorithmes d'apprentissage automatique peuvent traiter des données issues d'affaires antérieures et en cours afin de déceler des points communs qui pourraient échapper à l'analyse humaine, ce qui pourrait orienter l'enquête vers de nouvelles pistes.

L’IA peut aussi jouer un rôle crucial dans la génération de profils génétiques en analysant les échantillons ADN collectés sur les scènes de crime. Cette capacité à identifier des correspondances génétiques, même dans des bases de données étendues, peut accélérer le processus d’identification des suspects afin d’explorer des pistes plus rapidement ou de résoudre potentiellement des affaires non résolues depuis des années.

Enfin, l’IA peut contribuer à la reconstitution des scènes de crime et à la création de simulations virtuelles, ce qui donne la possibilité aux enquêteurs d’explorer de nouvelles hypothèses. La technologie de reconnaissance faciale alimentée par l’IA peut aussi être un avantage, car en comparant des photos de scènes de crime ou des séquences de surveillance avec des bases de données d’individus connus, les algorithmes d’IA peuvent identifier des suspects potentiels.

L’intégration de l’IA dans les enquêtes criminelles ouvre ainsi de nouvelles perspectives dans la résolution des *cold cases*. Ces algorithmes peuvent analyser des données, reconstituer des scènes de crime et aider à identifier des suspects potentiels. Ce faisant, ils facilitent la tâche des analyses criminelles mais ils ne pourront pas les remplacer. En effet, ces derniers devront continuer à vérifier toutes les informations d’une procédure et à explorer chaque piste. L’intelligence artificielle restera seulement un formidable auxiliaire permettant d’accélérer les analyses et formuler de nouvelles hypothèses. Dans tous les cas, l’élucidation des affaires non résolues sera toujours une tâche complexe et difficile. Mais à mesure que l’IA évolue, son rôle dans la résolution des *cold cases* deviendra sans doute plus important.

65 L'IA améliore-t-elle l'efficacité des drones en matière de surveillance et de sécurité ?

Le drone a pris naissance dans un cadre militaire en France dès le début de l'aviation et avant même le début de la Première Guerre mondiale. Il a été tardivement perçu par la France comme une opportunité. Ces aéronefs sans pilote furent d'abord baptisés « *Queen Bee* » (« *Reine des abeilles* ») par les Anglais en raison de leur bruit en vol, puis « drone », dérivé du terme anglais *dran*, c'est-à-dire « faux bourdon » en raison de leur bourdonnement grave et un peu hésitant. Ce n'est que dans les années 1970 que cette technologie parviendra à maturité. Depuis lors, l'essor des drones militaires a aussi conduit au développement des drones civils.

Les aéronefs sans pilote sont en effet de plus en plus présents dans l'espace public, qu'ils soient utilisés à des fins récréatives, professionnelles, ou encore sécuritaires et militaires. Les drones sont devenus une innovation ayant de très multiples usages civils : inspection d'infrastructures essentielles telles que les routes, ponts ou les centrales nucléaires ; lutte contre les incendies ; la livraison de colis, l'agriculture ; etc. À ce titre, le terme « drone » est un terme générique recouvrant une grande variété d'aéronefs, allant des appareils miniatures pouvant emporter une caméra à des engins plus volumineux capables de transporter des personnes, tels que ceux développés par Airbus dans le cadre du projet « Urban Air Mobility ». Le terme légal consacré pour décrire ces appareils est celui d'« *aéronef circulant sans équipage à bord* ». Son équivalent anglais, l'acronyme UAV pour *Unmanned Aerial Vehicle*, renvoie également à un « *véhicule aérien sans pilote* ».

Leur degré d'autonomie est aussi variable. Certains sont télépilotés depuis une station au sol, fixe ou mobile, tandis que d'autres sont autonomes et peuvent opérer de manière plus ou moins automatisée, avec ou sans intervention humaine. Ces derniers sont généralement considérés comme porteurs de multiples risques et menaces (risque de collusion, action terroriste, atteinte directe aux personnes physiques, etc.). Le drone est à ce jour indissociable de la personne qui le met en œuvre : le terme de « télépilote » est devenu officiel depuis la loi relative au renforcement de la sécurité de l'usage des drones civils, promulguée le 24 octobre 2016. L'attachement à la figure du « pilote » s'explique par la nécessité d'identifier un responsable.

L'usage des drones à des fins de sécurité publique a également suivi cette évolution. Dès 2014, la Préfecture de police et la Police nationale expérimentèrent l'usage de drones pour surveiller un match de football. L'expérience fut renouvelée lors d'événements sportifs. Des drones furent utilisés lors de l'évacuation de la ZAD de Notre-Dame-des-Landes, durant les confrontations avec les Gilets jaunes, ou encore pendant le Covid afin de faire respecter les mesures sanitaires.

Utiliser les drones à des fins de surveillance et de sécurisation nécessite, à ce titre, un cadre juridique sécurisé et permettant de garantir le respect des libertés publiques. La France a été l'un des premiers pays à se doter d'une réglementation en matière de drones. Ce cadre s'est formé progressivement. Les conditions dans lesquelles les services

de la police et gendarmerie nationales, les douanes et l'armée peuvent procéder au moyen de caméras aéroportées à la captation, l'enregistrement et la transmission d'images sont aujourd'hui déterminées dans le Code de la sécurité intérieure.

Depuis le 19 avril 2023, une nouvelle réglementation encadre la captation et l'utilisation d'images de drones par les forces de l'ordre à des fins de sécurité publique. En pratique, les forces de l'ordre peuvent être autorisées à utiliser des caméras aéroportées dans des cas bien précis : la sécurité et le maintien de l'ordre lors de rassemblements sur la voie publique (manifestations) ; la prévention d'actes de terrorisme ; la surveillance des frontières ; le secours aux personnes ; la régulation des flux de transport. Il est à noter que cette utilisation est orientée vers la police administrative et vise à prévenir, sécuriser et secourir, plutôt qu'à collecter des preuves dans le cadre de poursuites judiciaires. Les drones utilisés par la police sont d'ailleurs équipés de caméras optiques, de zooms et souvent aussi de caméras thermiques, ce qui les rend aptes à la surveillance et la recherche à distance.

Le législateur a prévu des garanties permettant de limiter les atteintes aux libertés individuelles. Seul le préfet peut, par exemple, autoriser l'utilisation de ce procédé de captation, de transmission et d'enregistrement d'images à visée sécuritaire. Dans les faits, ces autorisations sont renouvelables sans limite de temps et sans contrôle préalable d'un organe indépendant du pouvoir exécutif, ce qui renforce le caractère particulièrement intrusif de cette surveillance.

L'introduction de l'IA dans la technologie des drones constitue un véritable tournant. Son intégration dans les drones vise à améliorer davantage les capacités de détection et de surveillance de ces derniers. En effet, ces drones pourront alors analyser les images en temps réel, identifier les menaces potentielles et alerter les opérateurs humains. L'IA favorise surtout l'autonomie des drones en réduisant le rôle des humains. Les algorithmes de *machine learning* permettent en effet aux drones d'apprendre de leurs expériences passées et d'améliorer leurs performances au fil du temps. L'IA dans les drones tourne principalement autour de la « vision par ordinateur » qui représente un domaine de l'intelligence artificielle permettant aux ordinateurs et aux systèmes de dégager des informations significatives à partir d'images numériques ou de vidéos. La vision par ordinateur permet aux ordinateurs de voir et d'observer tandis que l'intelligence artificielle leur permet de penser.

Les drones peuvent ainsi capturer des images et des vidéos de haute qualité, souvent à l'insu des personnes. Cela soulève évidemment des questions relatives au respect du droit à la vie privée, à la surveillance de masse et au profilage des citoyens. Il convient donc de peser attentivement les avantages potentiels de la surveillance par drone en termes de sécurité contre les risques pour la vie privée et les libertés individuelles.

66 Quelles sont les applications de l'IA dans le domaine de la défense ?

Les applications de l'IA dans la défense sont nombreuses. Elles vont de la gestion des carrières du personnel aux éventuels systèmes de combat autonomes. L'utilisation de l'IA dans le domaine militaire dépend de nombreux facteurs, tels que les ressources financières et technologiques disponibles, les priorités stratégiques et les orientations politiques.

L'intelligence artificielle n'est pas véritablement une nouveauté dans le domaine militaire. Certains systèmes toujours opérationnels utilisent des technologies dérivées des systèmes experts. C'est, par exemple, le cas du système de défense antimissile *Aegis* équipant les navires de l'US Navy qui peut être considéré comme une technologique issue du champ de l'IA.

Le secteur militaire est, à ce titre, demandeur et utilisateur de ces technologies d'intelligence artificielle, en tant que systèmes d'aide à la décision dans des environnements complexes. L'armée américaine a d'ailleurs une longue histoire d'expérimentation et d'utilisation des systèmes d'IA. En 1991, un programme d'intelligence artificielle nommé *Dynamic Analysis and Replanning Tool* (DART) a été utilisé pour planifier le transport des approvisionnements et du personnel.

En France, le domaine de la défense et de la sécurité fait partie des quatre domaines prometteurs retenus par le rapport Villani de 2018. Utilisée à des fins militaires dans un cadre éthique, l'IA complète l'action humaine, mais ne la remplace pas. Les applications de l'IA pour la défense sont très nombreuses allant de l'assistance aux opérateurs en temps réel jusqu'au soldat augmenté.

Les systèmes d'IA peuvent être notamment utilisés pour la surveillance et la reconnaissance. L'intégration de l'IA à la collecte et à l'interprétation des quantités considérables de données de surveillance recueillies par les satellites et les drones peut permettre d'améliorer la compréhension de la situation globale dans le cadre d'un conflit. Il est en effet nécessaire de trier toutes les informations recueillies pour en extraire ce qui est réellement utilisable. Ces données peuvent comprendre les données de missions passées, des facteurs environnementaux et cartographiques ainsi que les renseignements recueillis sur le terrain par des agents de renseignement. L'analyse de ces ensembles de données globales est essentielle pour recueillir des renseignements lors d'opérations ou planifier des scénarios militaires. L'IA permet ainsi d'analyser les données de reconnaissance, comme les images satellites, pour identifier les cibles et les menaces potentielles.

La start-up française Preligens, fondée en 2016 et rachetée récemment par Safran, édite, par exemple, des logiciels à destination du ministère des Armées et du renseignement permettant d'identifier automatiquement des matériels militaires et de repérer tout mouvement inhabituel sur des sites d'intérêts grâce à des photographies prises par satellite et analysées par des algorithmes d'intelligence artificielle.

Par ailleurs, l'intelligence artificielle peut être utilisée dans la prévention des attaques en détectant les menaces et en anticipant les actions ennemies grâce à des algorithmes de prédiction basés sur des données historiques et des analyses de comportement. La

simulation et la modélisation assistées par IA peuvent jouer un rôle crucial dans l'entraînement des forces militaires, ce qui permet de les préparer à divers scénarios complexes.

Elle peut également être utilisée pour fournir une assistance en temps réel aux commandants militaires en analysant les données et en proposant des solutions stratégiques pour optimiser les déploiements de forces sur le terrain. Elle peut être aussi utilisée pour gérer l'approvisionnement et la logistique, les itinéraires de transport, la gestion des stocks de matériel et de fournitures et les mouvements de personnel et de matériel. L'application de l'intelligence artificielle à la logistique et à la maintenance militaires devrait assurer des opérations plus fluides et une disponibilité accrue des équipements. La maintenance prédictive, exploitant des algorithmes d'IA pour analyser l'état des machines et prédire les pannes avant qu'elles ne surviennent, permettra en effet de réduire le temps d'immobilisation et d'optimiser la gestion des stocks de pièces de rechange.

L'IA peut être enfin utilisée pour contrôler les drones ainsi que les robots militaires afin d'effectuer des missions de surveillance et de reconnaissance. Elle peut également être utilisée pour développer des drones et des robots autonomes. Des essaims de drones autonomes pourraient, par exemple, être capables d'analyser en temps réel les données issues de capteurs disséminés sur le terrain et d'effectuer des attaques sans risquer de vies humaines offrant ainsi un avantage tactique.

Malgré ses nombreux avantages, l'intégration de l'intelligence artificielle dans les opérations militaires présente également des risques. En déléguant certaines tâches à des algorithmes, il y a tout d'abord le risque de dépendance vis-à-vis de ces technologies et la perte de certaines compétences humaines, ce qui peut devenir problématique en cas de panne ou de dysfonctionnement du système d'IA. Il y a également le risque du manque de transparence dans les décisions prises par les algorithmes. Il est vrai que les systèmes d'aide à la décision fondés sur l'IA ne prennent pas de décisions en soi, mais ils influencent directement les décisions des êtres humains, ce qui pose évidemment une question de responsabilité car ces décisions peuvent avoir de graves conséquences. Enfin, l'intelligence artificielle peut être utilisée à des fins militaires illégitimes, comme la guerre de l'information ou la désinformation pour diffuser de fausses informations ou pour manipuler les opinions publiques dans l'optique de déstabiliser les gouvernements et de provoquer des conflits.

La création en avril 2024 de l'Agence ministérielle pour l'IA de défense (AMIAD) par le ministre des Armées marque pourtant le début d'une ère où l'intelligence artificielle jouera un rôle central dans la stratégie de défense de la France (cf. question n° 68). L'industrie de la défense devrait donc connaître un certain essor lui permettant de continuer à trouver de nouvelles applications pour l'intelligence artificielle dans le domaine militaire.

67 La menace des robots tueurs est-elle réelle ?

Si des « robots tueurs » autonomes sont en service depuis 2006, l'utilisation en masse de l'IA à des fins létales est plus récente. Par exemple, actuellement, l'IA a permis de frapper la bande de Gaza avec une intensité inédite en générant chaque jour des centaines de victimes parmi les civils. Les médias israéliens ont, à ce titre, publié, fin 2023, une enquête sur l'IA *Habsora* (« l'Évangile », en français) : un programme informatique utilisant l'intelligence artificielle et fonctionnant comme une « usine à cibles », vingt-quatre heures sur vingt-quatre. Le 3 avril 2024, ils ont révélé l'emploi d'une autre IA à des fins militaires, un programme baptisé « Lavender », qui utilise l'IA pour identifier des cibles à Gaza, avec une certaine marge d'erreur. L'enquête est basée sur des témoignages d'agents actifs de la division d'élite du cyber-renseignement se servant de *Lavender*. Pour une personne visée à son domicile, le nombre de victimes collatérales jugées acceptable par l'armée est de 15 à 20 personnes. Si la cible est de très haute valeur, il passe de 200 à 300 personnes.

Comme l'explique +972 Magazine, les deux systèmes cohabitent : *Habsora* sert à choisir des bâtiments à frapper alors que *Lavender* est conçu pour identifier des cibles potentielles à assassiner. C'est ainsi que l'intelligence artificielle militaire israélienne a fini par pointer 37 000 personnes comme des cibles « légitimes » parmi la population de Gaza pour établir une « liste de mort ». L'enquête souligne que les militaires faisaient confiance aux décisions de *Lavender* bien que ce système commette des erreurs dans environ 10 % des cas et qu'il identifie parfois des individus qui n'ont aucun lien avec les cibles potentielles. L'armée a néanmoins approuvé toutes les frappes aériennes sur des cibles désignées par l'IA sans vérification. Ces frappes aériennes visaient d'ailleurs des personnes à leur domicile, souvent la nuit en présence des familles, car cela était jugé plus facile du point de vue du renseignement. En outre, l'enquête révèle l'utilisation d'un autre système d'intelligence artificielle appelé « *Where's Daddy* ? » (Où est papa ?) pour suivre les personnes ciblées et mener des bombardements lorsqu'elles se trouvaient à l'intérieur de leur résidence familiale. Les conséquences de ces actions ont été dramatiques, puisque des milliers de Palestiniens, majoritairement des civils, ont été tués par des frappes aériennes indiscriminées.

C'est la première fois que l'intelligence artificielle est utilisée à grande échelle par une armée. Malgré les contestations de l'ONU et des associations de défense des droits de l'homme, le recours à ce type de technologie dans les conflits va nécessairement s'accroître du fait des progrès continus dans le domaine de l'intelligence artificielle et de la robotique permis par la quatrième révolution industrielle. En effet, le développement d'armes entièrement autonomes transforme et transformera encore plus la nature de la guerre, soulevant ainsi de nombreuses questions éthiques, morales, juridiques et sécuritaires.

Ces armes entièrement autonomes, également connues sous le nom de « systèmes d'armes létales autonomes » (SALA), dits « drones tueurs » ou encore « robots tueurs » dans leur appellation plus médiatique, seraient en mesure de sélectionner et d'attaquer des cibles sans intervention humaine. Ces systèmes sont appelés à remplir plusieurs

fonctions de base dont celles d'accroître l'efficacité de destruction des cibles ennemies et de préserver la vie des soldats. Ce serait, par exemple, des drones armés de missiles qui, grâce à un logiciel de reconnaissance de cette menace, seraient, sans aucune intervention humaine, en mesure de la neutraliser. Certains drones, utilisés notamment en Ukraine, disposent, par exemple, d'un pilotage automatique prenant le relais lorsqu'ils se retrouvent en zone de brouillage radio et ne peuvent donc plus recevoir d'instructions humaines.

La question n'est pourtant pas nouvelle. Depuis le début des années 2000, les drones MALE [Moyenne Altitude Longue Endurance] MQ-1 Predator et MQ-9 Reaper (la faucheuse) ont été en effet utilisés dans les opérations menées par les forces américaines aussi bien en Afghanistan et en Irak, qu'en Somalie et en Libye. Ces drones ont profondément déjà bouleversé la manière de mener une guerre qui n'est plus seulement un affrontement sur un champ de bataille, mais est devenue surtout une guerre de technologies et de renseignement. Certains de ces systèmes existent ainsi depuis des années, mais la portée géographique et l'environnement dans lesquels ils sont utilisés restent limités. Les progrès technologiques favorisent cependant le développement de systèmes qui fonctionnent sans contrôle humain significatif et déléguant les décisions de vie ou de mort aux machines. Les fabricants d'armes ont en effet tendance à augmenter le niveau d'autonomie de leurs équipements, notamment grâce à l'intelligence artificielle, même s'il existe encore une possibilité d'intervention humaine dans le processus de frappe.

Des prototypes de ces armes sont développés par un nombre croissant d'États, et ce mouvement devrait s'accentuer dans la décennie à venir, il est donc nécessaire qu'une réflexion portant sur l'intégration de l'intelligence artificielle à la robotique militaire soit menée pour encadrer les dérives.

Le droit international humanitaire actuel ne peut traiter convenablement les questions soulevées par ces armes entièrement autonomes. Une interdiction préviendrait ainsi leur création et leur prolifération ainsi que leur utilisation. C'est, semble-t-il, la voie choisie par les États membres de l'ONU qui, lors d'une conférence sur les armes autonomes, qui s'est tenue en avril 2024 en Autriche, ont soutenu l'idée d'un traité international limitant voire interdisant ces armes qui pourraient rendre les guerres encore plus inhumaines. Les promoteurs de ce traité ont appelé à l'interdiction des systèmes d'armes autonomes qui, par nature, fonctionnent sans contrôle humain significatif ainsi qu'à des réglementations garantissant qu'aucun système ne puisse être utilisé sans contrôle humain significatif.

Un traité d'interdiction des SALA permettrait d'assurer le respect du droit international, mais encore faudrait-il qu'il soit ratifié par les principales puissances ayant la capacité industrielle et technologique de développer des tels systèmes. L'adoption d'un tel traité n'aurait en effet aucun impact réel s'il n'est pas ratifié par les États producteurs de tels systèmes.

68 L’IA entraîne-t-elle une nouvelle course à l’armement ?

Les grandes puissances mondiales actuelles, notamment les États-Unis, la Russie et la Chine, se sont lancées dans une véritable course à l’armement intégrant l’intelligence artificielle, qui n’est pas sans rappeler la course à l’armement nucléaire du XX^{e} siècle.

Dans le domaine des équipements militaires, l’IA pourrait constituer la troisième révolution majeure, après l’invention de la poudre à canon et de la bombe atomique. Elle fournit, à ce titre, une dimension supplémentaire à n’importe quelle arme : l’autonomie. Toutes sortes d’armes peuvent être transformées en systèmes autonomes, régis par des algorithmes d’IA (robots, drones, torpilles). L’IA permettrait une plus grande rapidité d’analyse et d’exécution, et donc la possibilité de localiser, de sélectionner et d’attaquer des cibles plus ou moins sans intervention humaine, en fonction du degré d’autonomie.

Le secteur militaire est celui qui reçoit le plus d’investissements dans les pays développés. La demande croissante de matériel militaire de pointe est le principal facteur de croissance de l’intelligence artificielle dans l’industrie militaire. L’enjeu est avant tout stratégique, car l’intelligence artificielle est appelée à jouer un rôle déterminant dans la supériorité technologique des armes de demain. Les dépenses militaires mondiales dans ce domaine ont connu une croissance rapide ces dernières années. Elles passent, selon *The Business Research Company*, de 8,45 milliards en 2023 à 9,86 milliards de dollars en 2024. Cette croissance va se poursuivre au cours des prochaines années jusqu’à atteindre 19 milliards de dollars d’ici 2029, selon certaines prévisions.

En France, une nouvelle agence ministérielle de l’intelligence artificielle de défense (Amiad), créée en juillet 2024, sera dotée de 300 millions d’euros par an de 2024 à 2030, soit un équivalent de 2 milliards d’euros. Elle aura pour objectif de développer l’intelligence artificielle dans le domaine de la défense afin de perfectionner l’armement, le renseignement et la planification des opérations militaires. Près de 800 personnes devraient travailler, en 2026, sur l’IA au ministère des Armées.

Le secteur de l’IA semble appelé à prendre une place importante dans les stratégies militaires des États, ce qui rend nécessaire de s’interroger sur les conséquences que pourrait avoir son développement en matière de dissuasion nucléaire. Les systèmes d’IA offrent en effet la possibilité de renforcer la dissuasion nucléaire en améliorant la précision et les performances des moyens de défense nucléaire. L’intégration de l’intelligence artificielle dans les systèmes d’armes nucléaires automatisés suscite cependant de véritables inquiétudes, car une telle évolution est susceptible d’accentuer la probabilité d’une guerre nucléaire. Personne n’ignore les risques existentiels que soulève l’utilisation de telles armes nucléaires pour l’humanité et la planète.

L’évolution de la conjoncture internationale nous indique que le risque nucléaire ne fait qu’augmenter. Les tensions géopolitiques persistantes, notamment entre les États-Unis, la Russie et la Chine, se répercutent inévitablement au niveau nucléaire. L’intensification de la menace nucléaire est d’ailleurs souvent imputée aux agissements de la Russie, plus particulièrement en raison de la guerre en Ukraine. Le Kremlin a en effet menacé à plusieurs reprises d’utiliser ses armes nucléaires en Ukraine et contre ses ennemis, éventuellement en première frappe, ce qui représente une escalade rhétorique notable.

Quant à la Chine, bien qu'elle ait aussi développé son arsenal nucléaire au cours des dernières années, elle affirme qu'elle reste engagée dans une politique de non première frappe.

Dans son principe, la dissuasion nucléaire ne repose pas sur l'emploi mais sur la menace de l'emploi de l'arme nucléaire. Elle fait cependant dépendre l'ordre international d'une menace réciproque d'anéantissement, ce qui n'est pas véritablement une situation de paix mais de non-guerre. Appliquée aux armes nucléaires, l'IA vient ajouter une couche de risque supplémentaire à un niveau de danger déjà considérable. Par ailleurs, les normes ayant régi les relations nucléaires entre grandes puissances, notamment le contrôle des armements, tendent à s'affaiblir laissant envisager une période de prolifération des armes nucléaires.

Dans une déclaration prononcée à l'occasion de l'ouverture de la session de 2024 de la Commission du désarmement, la Haute-Représentante des Nations Unies pour les affaires de désarmement, Izumi Nakamitsu, a constaté que l'environnement géopolitique actuel est marqué par une concurrence accrue en matière d'armes stratégiques et une déliquescence de la confiance entre les États dotés d'armes nucléaires. Elle a recommandé à la Commission de se concentrer sur l'élaboration de mesures visant à renforcer et à préserver l'architecture de maîtrise des armements, de désarmement et de non-prolifération afin d'empêcher tout recours à l'arme nucléaire.

Malgré les efforts déployés à l'échelle mondiale pour réfléchir à un encadrement de l'IA, les applications militaires ont été souvent absentes de nombreuses discussions. Une question centrale reste la possibilité de céder le contrôle d'une arme aussi sensible que l'arme nucléaire à une intelligence non-humaine.

L'inquiétude est bien réelle, c'est la raison pour laquelle un haut fonctionnaire américain, Paul Dean, a insisté, en mai 2024, sur le fait que Washington avait pris un « engagement clair et fort » en faveur d'un contrôle total des armes nucléaires par les humains, ajoutant que la France et le Royaume-Uni avaient fait de même. Il a également exhorté la Chine et la Russie à s'aligner sur ces déclarations selon lesquelles seuls les humains prendraient les décisions relatives au déploiement d'armes nucléaires. Une telle déclaration est-elle pour autant suffisante ? Il s'agit seulement d'un appel à la prudence et à la responsabilité dans l'utilisation de l'IA, notamment dans le domaine des armes nucléaires. Ne serait-il pas plus prudent d'interdire tout simplement l'intégration des systèmes d'intelligence artificielle dans les systèmes de contrôle des armes nucléaires ? En effet, une telle intégration augmente les possibilités d'erreurs de calcul et de fautes graves en accélérant les temps de réponse et en excluant l'intuition et l'hésitation humaines.

De manière générale, la majorité des États de la planète sont en faveur de la prohibition complète des armes nucléaires. Cent vingt-deux pays ont d'ailleurs voté, en 2017, en faveur d'un *Traité sur l'interdiction des armes nucléaires*, entré en vigueur en 2021, malgré l'opposition des grandes puissances. Bien que ce traité n'ait qu'une portée limitée en l'absence de ces dernières, il exprime l'inquiétude de nombreux pays face à la menace de l'emploi de l'arme nucléaire par une poignée d'États menaçant le monde d'un conflit nucléaire. Une réflexion sur la maîtrise des armements au sens large doit être menée au niveau mondial pour réduire le risque de la menace nucléaire.

69 L’IA peut-elle améliorer la cybersécurité ?

Toujours plus sophistiquées, les cyberattaques ont augmenté parallèlement au développement rapide de l’intelligence artificielle. Les dommages causés peuvent être très élevés et se chiffrer en centaines de milliers, voire en millions d’euros. Ces attaques se multiplient et des risques nouveaux apparaissent causant des préjudices aussi bien aux personnes et qu’aux organisations : ransomware, malware, fraudes diverses dont celle dite au président, menaces contre les données, menaces contre la disponibilité du système, etc. Lorsqu’on parle de cyberattaques, il convient de souligner qu’il ne s’agit que d’une partie de la cybersécurité correspondant à des menaces extérieures, et sont commises volontairement. L’expression qui désigne l’intégralité des risques à la cybersécurité, qu’ils soient volontaires ou non, extérieurs ou non, est celle d’incidents de sécurité.

À ce titre, la relation entre la cybersécurité et l’IA est une relation d’interdépendance. En effet, afin que l’intelligence artificielle contribue à la cybersécurité, il faudrait que celle-ci soit un outil sûr et de confiance. Une distinction doit ainsi être faite entre la cybersécurité par l’intelligence artificielle, et la cybersécurité de l’IA elle-même.

Concernant tout d’abord la cybersécurité par l’intelligence artificielle, l’IA est de fait déjà utilisée afin de renforcer la prévention, la détection et la réponse aux attaques. Dans la phase de détection des attaques et des incidents de sécurité, elle peut améliorer la détection et la réponse aux menaces d’une organisation en accélérant la réponse aux cyberattaques afin de réduire les dommages que les attaquants peuvent causer à l’organisation. Grâce à ses capacités d’apprentissage automatique et d’analyse d’un grand nombre de données, l’IA peut ainsi identifier les schémas et les comportements suspects dans les réseaux informatiques, ce qui permet de détecter rapidement les cyberattaques. Lorsqu’une menace est détectée, elle peut aussi automatiser la réponse aux incidents. Les systèmes de cybersécurité basés sur l’IA peuvent, par exemple, arrêter ou ralentir les processus suspects et lancer des analyses de sécurité approfondies sans intervention humaine.

Elle peut également prioriser les tentatives d’intrusion détectées, ce qui en fait un véritable atout pour les opérateurs. Dans la détection comme dans l’analyse des incidents de sécurité, l’IA peut affiner l’étude des menaces grâce à des corrélations permettant ainsi d’accélérer la réponse à ces menaces. Elle n’est cependant pas infaillible et peut parfois générer des faux positifs, c’est-à-dire des alertes de sécurité entraînant une surcharge de travail pour les équipes de sécurité et la découverte tardive de véritables menaces.

Parmi ces menaces, nous pouvons citer les ransomwares qui constituent l’un des problèmes les plus pressants. Ces logiciels malveillants conçus pour chiffrer les fichiers d’une victime, rendant les données inaccessibles jusqu’à ce qu’une rançon soit payée. Ces attaques peuvent paralyser les entreprises, les hôpitaux et même les gouvernements, entraînant d’importants préjudices. Les algorithmes d’apprentissage automatique peuvent être entraînés à reconnaître les signes avant-coureurs d’une attaque de ransomware, tels que des tentatives de connexion inhabituelles ou des transferts de données anormaux.

Outre la détection et de la réponse, l'IA joue aussi un rôle dans le renforcement de la résilience des systèmes d'information. Elle peut aider à prédire les vulnérabilités futures et à développer des stratégies de défense proactives. Par exemple, en analysant les codes sources pour identifier les failles de sécurité potentielles avant qu'elles ne soient exploitées par des cybercriminels. Elle est également capable de générer une cartographie des éléments touchés par l'attaque. Enfin, elle peut proposer des plans de remise en marche des systèmes et des réseaux.

Malheureusement, si cette technologie est déjà utilisée par les cyberdéfenseurs, elle l'est également par les cyberattaquants. Dans un rapport du 24 janvier 2024, le centre national britannique de la cybersécurité (*National Cyber Security Centre*, NCSC) indique que l'IA est déjà utilisée dans le cadre de cyberactivités malveillantes et qu'elle augmentera très certainement le volume et l'impact des cyberattaques, y compris des ransomwares. Le rapport souligne que cette technologie permet à des attaquants relativement peu qualifiés de mener plus aisément ce type d'opérations. Le recours à l'IA permet également de démultiplier le pouvoir de nuisance de ces attaquants qui seraient en mesure de conduire des attaques plus furtives, plus automatisées et à bien plus grande échelle.

L'intelligence artificielle est donc elle-même sensible à toutes les cyberattaques répertoriées et à différentes attaques plus spécifiques, telles que l'empoisonnement des données (envoi massif de données de nature à fausser les résultats), l'empoisonnement du modèle (corruption des modèles pré-entrainés dans le but de contourner le processus de décision), l'extraction de données, l'extraction de modèle (créer une copie du modèle en disposant d'un jeu de données d'entrée/sortie afin d'étudier les solutions de contournement du modèle), l'évasion (communication de modifications dans les données entrantes qui vont modifier la classification), etc. Les conséquences de ces attaques pourront atteindre la disponibilité du système, son intégrité ou la confidentialité des données. Par conséquent, la question de la cybersécurité de l'IA doit être impérativement intégrée dans les phases de conception de ces systèmes. Le cadre juridique de la cybersécurité est d'ailleurs en constance évolution afin d'intégrer ces nouvelles menaces.

Depuis 2013, la France a déjà développé une politique de cybersécurité pour les entités critiques qualifiées d'opérateurs d'importance vitale (OIV). L'Union européenne s'est saisie de cette question en 2016 avec la directive NIS (*Network & Information Security*) qui s'applique aussi bien aux organisations publiques que privées. Ces règles ont été mises à jour par la directive NIS 2, entrée en vigueur en 2023, qui a modernisé le cadre juridique existant afin de le renforcer afin de tenir compte de la généralisation de la numérisation et des multiples interconnexions. Elle devra être transposée par chaque État membre de l'Union européenne au plus tard en octobre 2024. En élargissant ses objectifs et son champ d'application, cette directive anticipe les nouvelles formes d'attaques et marque ainsi un changement de paradigme, passant d'une approche réactive à une stratégie proactive.

70 Quels sont les dangers des algorithmes de reconnaissance faciale ?

Dans son roman 1984, George Orwell imaginait une société dans laquelle les citoyens seraient surveillés en permanence par l'œil de *Big Brother*. Ce classique de la science-fiction n'a pas empêché le monde de converger vers un modèle similaire à celui prédit par l'écrivain britannique. La reconnaissance faciale peut s'apparenter à de la science-fiction, mais elle est de plus en plus présente dans notre vie quotidienne. La technologie de reconnaissance faciale s'est en effet considérablement développée ces dernières années et a trouvé de nombreuses applications, notamment dans les domaines de la surveillance et de la sécurité.

La reconnaissance faciale est une technologie utilisée pour identifier une personne sur une photo ou une vidéo en comparant son visage avec ceux sauvegardés dans une base de données. Il s'agit d'une technologie combinant les techniques biométriques, l'intelligence artificielle, la cartographie 3D et le *deep learning* pour comparer et analyser le visage d'une personne afin de l'identifier. Cette technologie repose sur la capacité des outils à reconnaître les motifs et structures des caractéristiques faciales humaines, telles que les yeux, le nez, la bouche ou encore les contours du visage. Il est ainsi possible d'identifier un visage dans une image ou une vidéo.

Il convient d'effectuer une distinction entre l'identification et l'authentification, toutes deux opérées grâce à la reconnaissance faciale mais pour lesquelles les enjeux posés ne sont pas les mêmes. L'identification sert à retrouver une personne au sein d'un groupe d'individus, dans un lieu, une image ou une base de données. Cette pratique bénéficie, par exemple, au fichier TAJ (Traitement d'antécédents judiciaires) que les enquêteurs interrogent avec une image pour identifier un suspect.

Quant à l'authentification, il s'agit de vérifier si le visage identifié correspond à celui d'une personne déjà connue, présente dans une base de données. De nombreux téléphones utilisent déjà cette technologie pour accéder au système, comme pour Face ID pour le déverrouillage d'un iPhone. Des portiques comparant le visage d'un passager à celui stocké dans son passeport sont également utilisés pour le passage de la frontière dans les aéroports.

En pratique, la reconnaissance peut être réalisée à partir de photos ou d'enregistrements vidéo et se déroule en plusieurs phases. Dans un premier temps, le visage d'un individu est capté sur une photo ou une vidéo à l'aide de caméras de surveillance, de caméras de sécurité, de webcams ou même de smartphones. Vient ensuite l'analyse de l'image captée. La plupart des systèmes identifient 80 caractéristiques du visage appelées points nodaux. Parmi ces caractéristiques, on compte la longueur ou la largeur du nez, la distance entre les yeux, la forme des joues, la profondeur des orbites, ou encore la largeur de la mâchoire. Les informations analogiques sont ensuite converties en code numérique appelé *faceprint* ou empreinte faciale. Cette représentation numérique du visage est alors comparée à une base de données de visages.

Le système de reconnaissance faciale permet d'identifier rapidement et précisément les individus cibles lorsque les conditions sont favorables. Toutefois, sa fiabilité peut diminuer si le visage est endommagé ou modifié par chirurgie, partiellement obscurci ou de profil

plutôt que de face.

La reconnaissance faciale présente plusieurs avantages dont le fait d'accroître le niveau de sécurité lorsqu'elle est couplée avec la vidéosurveillance. Elle permet aussi de simplifier le processus d'authentification sur les appareils électroniques tout en augmentant le niveau de sécurité.

Elle soulève pourtant de nombreuses craintes et inquiétudes. Ses détracteurs pointent le manque de fiabilité des algorithmes et le risque de biais. Les réseaux de neurones étant principalement entraînés sur des photos d'hommes caucasiens, ils se révèlent par la suite moins performants pour identifier les femmes ou les personnes de couleur.

La reconnaissance faciale peut-être malheureusement aussi utilisée pour cibler délibérément un groupe de personnes et porter ainsi atteinte à ses libertés fondamentales. En Chine, elle est utilisée pour contrôler le crédit social des personnes, dans certaines zones publiques. Autre exemple, Amnesty International a publié un rapport sur l'apartheid automatisé en mai 2023 qui mettait en évidence l'utilisation par l'armée israélienne de technologies de reconnaissance faciale pour renforcer son contrôle sur les Palestiniens en Cisjordanie et à Jérusalem-Est et automatiser les restrictions de leur liberté de mouvement, l'aidant ainsi à maintenir son système d'apartheid. Plus récemment, une enquête du *New York Times* de mars 2024 a révélé que les services de renseignement militaire israéliens utilisent la reconnaissance faciale pour effectuer une surveillance généralisée et ficher les visages des Palestiniens de Gaza sans leur consentement.

L'utilisation de la technologie de reconnaissance faciale dans une zone de conflit a suscité un débat mondial, car il est peu probable que les soldats israéliens remettent en question la technologie lorsqu'elle identifie une personne comme faisant partie d'un groupe militant, même s'il s'agit d'une erreur. Il y a donc un risque évident de déshumanisation d'un groupe de personnes.

Le véritable danger de la reconnaissance faciale est qu'il s'agit d'une technologie sans contact qui peut s'utiliser à distance et à l'insu des personnes. L'on passe alors d'une surveillance ciblée à une véritable surveillance de masse, ce qui constitue un véritable changement de paradigme.

VIII. L'INTELLIGENCE ARTIFICIELLE ET LES TRANSPORTS

71 Qu’est-ce qu’un système autonome ?

Voitures autonomes, sondes autonomes, robots autonomes ou encore armes autonomes sont des sujets de prédilection de la presse grand public. L’autonomie consiste à conférer à la machine la capacité d’agir sans supervision humaine ou dans des environnements plus ou moins complexes et variables. Comme il n’existe pas d’humain capable d’agir idéalement dans toutes situations, les machines ont leur propre champ d’action et d’autonomie. L’on parle de niveau ou de degré d’autonomie. L’autonomie de ces engins reste un défi scientifique majeur dans la mesure où il s’agit de substituer à la décision humaine une décision numérique. Concrètement, afin qu’une machine soit autonome, elle doit disposer de programmes ayant la capacité de sélectionner et d’organiser les actions à effectuer pour atteindre des objectifs sur la base de la connaissance de l’environnement dans lequel elle évolue et de son état général. Contrairement à la décision humaine, cette décision numérique est fondée sur des calculs mathématiques. Or ces derniers ne peuvent reproduire à la perfection certaines caractéristiques humaines telles que l’expérience, le jugement, l’intuition, ou encore le réflexe qui interviennent dans un processus de décision. Par ailleurs, la décision ne se réduit pas à une opération rationnelle, totalement maîtrisée, de l’esprit. En pratique, la décision fait face aussi à des incertitudes et des imprévus.

D’un point de vue philosophique, l’autonomie est la capacité à se donner ses propres lois. Autonomie vient en effet du grec *autos*, soi-même, et *nomos*, loi, règle. Est autonome ce qui se donne ses propres lois. Les machines n’en sont aujourd’hui pas encore capables. L’humain apparaît en effet dans la conception, la supervision et le renforcement des systèmes d’IA. Il ne faudrait pas oublier que ce sont les hommes qui fournissent données et objectifs à l’algorithme. S’il est certain que les systèmes d’IA basés sur l’apprentissage automatique sont effectivement capables de définir seuls une partie du paramétrage de leur modèle lors de la phase d’apprentissage, il faut cependant relativiser le caractère autonome de cet apprentissage. C’est en effet l’humain qui crée l’infrastructure initiale d’un réseau de neurones, procède à son entraînement en sélectionnant les jeux de données et qui participe à l’évolution du système d’IA tout au long de son cycle de vie. Par exemple, aucun système d’IA ne serait capable de dire qu’un chien est présent sur une photo sans aucune intervention humaine initiale. Lorsque l’on parle de l’autonomie de la machine ou de sa « semi-autonomie », se pose pourtant la question de la limitation de l’autonomie de l’homme, et par voie de conséquence de la responsabilité humaine. Cette dernière se diluerait d’une certaine façon devant les machines intelligentes et autonomes.

Jacques Ellul, professeur d’histoire du droit et sociologue, écrivait déjà en 1954 qu’« *il n’y a pas d’autonomie de l’homme possible en face de l’autonomie technique* ». Pourfendeur inlassable de la modernité technologique et théoricien de son autonomisation, la question de la technique occupe, à ce titre, une place centrale dans sa réflexion dans la mesure où la technique moderne constituerait la principale menace pour la liberté de l’homme en raison de sa subordination complète et définitive aux moyens qu’il s’était donné pour se libérer des contraintes naturelles. Sa thèse est que la technique a cessé

d’être un instrument neutre ou un simple moyen pour devenir un principe autonome d’organisation des sociétés. Selon lui, cette technique ne répondrait qu’à sa propre loi d’efficacité, elle s’auto-engendrerait faisant, en quelque sorte, des humains des objets de sa progression autonome. Loin d’être neuve, l’idée que les techniques modernes tendaient à l’autonomie a en réalité accompagné l’histoire des sociétés industrielles depuis le début du XIX^{e} siècle. Autrement dit, l’homme perd son autonomie au profit de la technique qu’il ne peut plus ni limiter, ni même orienter.

Sans aller jusqu’à considérer que la technique n’est plus dirigée ni orientée par quiconque, il est certain qu’elle affecte la responsabilité de l’humain. Que reste-t-il de l’autonomie de l’humain face à une machine autonome, et de ses capacités intrinsèques face à des capacités numériques de plus en plus performantes ?

L’émergence des systèmes autonomes de toute nature n’est pas exempte de défis techniques, mais aussi de défis tant juridiques qu’éthiques. L’IA n’est en effet pas seulement novatrice par sa technologie, elle l’est aussi par ses fonctions. Elle constitue un dispositif inédit ayant vocation à se substituer au fait intellectuel de l’homme en étant plus rationnel et performant que lui dans certains secteurs, comme la machine-outil a remplacé, à une certaine époque, son fait manuel en étant plus efficace que lui sur les chaînes de production. Contrairement à d’autres outils, l’IA dispose du pouvoir de prendre une décision affectant le réel sans qu’aucun stimulus externe ne lui soit imposé.

Or, les systèmes d’IA demeurent des systèmes informatiques comme les autres pouvant présenter des failles et subir des dysfonctionnements. Du fait de leurs décisions et des actions qu’ils réalisent, ils peuvent être source de dommages de toutes sortes. Étant une technologie complexe, l’intelligence artificielle pose de très nombreuses questions éthiques et juridiques qui restent encore pour le moment en suspens. Il est souvent prématuré d’essayer d’adopter un nouveau cadre juridique pour encadrer des technologies dont on commence à percevoir les implications. La question du régime de responsabilité en cas de dommage causé par une IA constitue toutefois un point crucial qui devra être clarifié afin de poursuivre le déploiement de l’intelligence artificielle en toute sécurité.

72 L’IA va-t-elle révolutionner nos moyens de transport ?

L’intelligence artificielle marque le début d’une nouvelle ère pour le secteur des transports. L’un des sujets les plus connus de l’impact de l’intelligence artificielle dans les transports est celui de la mise en place des véhicules autonomes. Apparu d’abord dans les séries de science-fiction, notamment la série K2000 datant des années 1980, le véhicule autonome est désormais une réalité.

Toutes les formes de transport sont toutefois concernées, y compris les trains, les navires et les avions. On parle d’ailleurs de mobilité du futur alliant objets connectés, nouvelles formes d’énergie et intelligence artificielle. La mobilité du futur sera ainsi technologique. Les transports, et en particulier les véhicules individuels, seront métamorphosés pour être plus autonomes et ultra connectés.

Un système de conduite autonome est un système permettant à un véhicule d’utiliser un ensemble de capteurs, de caméras, de radars et d’intelligence artificielle pour se déplacer entre des destinations sans un opérateur humain. Afin d’être qualifié de totalement autonome, un véhicule doit être capable de circuler du point de départ à destination sans aucune intervention humaine, avec un niveau de compétence comparable ou supérieur à celui d’un conducteur humain. À l’heure actuelle, ce niveau d’autonomie n’a pas encore été atteint. Le véhicule intégralement autonome n’existe pas encore, mais les technologies disponibles intègrent des systèmes qui autonomisent progressivement la conduite. Les voitures autonomes présentent d’ailleurs de nombreux avantages, notamment leur capacité d’améliorer considérablement la fluidité du trafic autoroutier en réduisant les embouteillages – leurs rapides temps de réaction aux changements de file et de rythme de conduite surpassant ceux de la plupart des conducteurs humains permettent en effet d’éviter les effets « accordéon » qui provoquent un ralentissement général des vitesses en situation de trafic chargé – et en optimisant les trajets.

La planification d’itinéraire est en effet l’un des domaines de l’IA les plus largement appliqués dans le domaine du transport, y compris le transport public. En effet, les algorithmes d’apprentissage automatique ont la capacité d’analyser d’importantes quantités de données en temps réel, en tenant compte de divers facteurs tels que la circulation et les conditions météorologiques, afin de définir les itinéraires les plus efficaces. L’IA dans la planification des itinéraires aide ainsi à réduire les temps d’attente, les temps de trajet et la consommation de carburant et à optimiser l’usage des voies de circulation. Elle peut aussi améliorer le transport public en ajustant les fréquences de service en fonction de la demande. Plus de véhicules peuvent, par exemple, être déployés aux heures de pointe qu’aux heures creuses, ce qui permet de réduire la consommation d’énergie.

Le transport routier est aussi en phase de transformation. Le marché des camions autonomes est certes encore embryonnaire, mais des tests ont déjà lieu, notamment aux États-Unis. Dans un contexte réglementaire moins contraignant que celui de l’Europe, des camions autonomes roulent déjà, à titre expérimental, sur certains axes notamment dans les États du sud, de l’Arizona à la Floride. Trois entreprises spécialisées dans les camions autonomes prévoient d’ailleurs de faire rouler ces camions sans conducteurs à bord dès la

fin de l'année 2024, malgré le risque déraisonnable que leur présence sur les mêmes axes de circulation peut représenter pour les automobilistes. La transformation potentielle du transport routier dépendra de la réussite de ces premiers trajets sans conducteur. D'ailleurs, l'intelligence artificielle ne se limite pas aux transports routiers, elle intègre aussi les transports aérien, ferroviaire, maritime et naval. Ces technologies ne cessent de s'améliorer et des systèmes de plus en plus autonomes ont vu et voient le jour.

L'intelligence artificielle peut également améliorer la maintenance des véhicules dans le secteur du transport, notamment celui du transport public. En analysant les données des capteurs installés sur les véhicules et l'infrastructure, l'IA peut anticiper les pannes potentielles avant qu'elles ne surviennent. Cette approche réduit les temps d'arrêt et prévient les pannes coûteuses. Par exemple, la RATP a développé une intelligence artificielle qui permet de mieux prédire les risques de panne de différents composants des matériels roulants, pour permettre de les réparer avant qu'ils ne tombent en panne et occasionnent des retards.

Par ailleurs, la sécurité est bien évidemment une préoccupation majeure pour le transport des personnes et des marchandises. L'IA peut, à ce titre, améliorer la sécurité des transports en détectant les comportements de conduite dangereux, tels que la distraction du conducteur ou sa somnolence. Pour ce faire, les systèmes d'intelligence artificielle peuvent surveiller en permanence le comportement du conducteur et alerter le conducteur ou le centre de contrôle en temps réel si un comportement dangereux est détecté.

Enfin, l'intelligence artificielle peut aussi améliorer l'expérience client dans les transports en commun à travers des chatbots qui fournissent des informations en temps réel et qui répondent aux questions des voyageurs. Les algorithmes peuvent en outre analyser leurs préférences individuelles afin de leur adresser des recommandations personnalisées de services (musique, activités, etc.) basées sur leur historique de voyage.

Des véhicules autonomes à la gestion logistique intelligente, l'intelligence artificielle transforme progressivement les moyens de transport des personnes et des marchandises. Si ces innovations semblent augmenter leur efficacité et leur rentabilité, il ne faudrait pas que cette évolution se fasse au détriment de la question fondamentale de la sécurité. Il convient donc d'intégrer avec prudence ces systèmes d'intelligence artificielle dans les différents moyens de transport tant que ces technologies ne seront pas totalement maîtrisées.

73 Comment fonctionne une voiture autonome ?

Sous le capot d'une voiture autonome, il y a des systèmes qui fonctionnent selon le triptyque : détection, planification et action. Le problème que pose la conception d'une voiture sans conducteur est qu'il s'agit de résoudre un nombre important de petits problèmes (reconnaissance de la signalisation et des obstacles, etc.) afin que toutes les solutions fonctionnent ensemble. Que les véhicules soient partiellement automatisés ou complètement autonomes, leur mode de fonctionnement n'est pas dans les faits si différents.

La première étape pour qu'une voiture autonome conduise à la place du conducteur est de l'équiper de capteurs (caméras, radars, lasers, capteurs ultrasons, etc.) qui servent à collecter des données sur l'environnement immédiat. Les voitures autonomes sont équipées d'un certain nombre de capteurs différents ayant chacun une fonction spécifique. Tous ces capteurs sont d'une certaine façon les « yeux » de la voiture qui collectent en permanence les informations extérieures utiles à la conduite telles que le trafic routier ou les panneaux de signalisation. Le capteur le plus important est, à ce titre, le Lidar (*Light Detection and Ranging*), qui utilise des rayons laser pour estimer la distance entre le capteur et l'objet atteint. Le Lidar permet à la voiture autonome de créer une carte en 3D de son environnement. Ces capteurs envoient des informations à des algorithmes qui les analysent et les traitent.

Les voitures autonomes sont également équipées de caméras servant à détecter les panneaux de signalisation, les feux de circulation, les marquages routiers et d'autres véhicules sur la route. Les données collectées sont traitées par des algorithmes de manière à obtenir des informations sous forme de détection ou de classification (couleur des feux ou type de signalisation). Enfin, les voitures autonomes sont également équipées de radars qui détectent les obstacles à longue portée ainsi que de capteurs à ultrasons détectant les obstacles à courte portée.

Les informations collectées par tous les capteurs sont ensuite traitées par un logiciel informatique intégrant une intelligence artificielle qui analyse, recoupe, classe et donne un sens aux données en temps réel. Autrement dit, c'est le système d'IA complexe du véhicule qui traitera toutes les données fournies par les différents capteurs et qui les transmet à un ensemble de logiciels afin de permettre à la voiture de fonctionner de manière totalement indépendante de son conducteur. Différentes applications vont par la suite assurer le contrôle du véhicule en fonction de son environnement immédiat.

En fonction du résultat de l'analyse des données par l'intelligence artificielle, la voiture totalement autonome peut prendre une décision de conduite. Les algorithmes de prise de décision utilisent notamment des modèles de prédiction pour anticiper le comportement des autres usagers de la route. Par exemple, les capteurs du véhicule détectent une forme. L'algorithme, qui analyse les données, identifie une apparence humaine qu'il interprète comme étant un piéton s'engageant sur le passage piéton ou un enfant surgissant entre deux voitures garées sur le côté à la poursuite d'un ballon. Il prend alors la décision d'activer les freins pour arrêter le véhicule et, si une collision est inévitable, décide de la manière de limiter les dégâts humains et matériels, par exemple en

choisissant de percuter le ballon ou un véhicule garé sur le côté plutôt que l'enfant qui court. Les conducteurs humains s'efforcent de déterminer ce que les autres conducteurs ou les piétons sont sur le point de faire afin d'éviter les accidents, c'est aussi ce que doit réaliser la voiture autonome grâce à la prédiction par l'apprentissage automatique. Elle anticipe ce qui va se produire afin de décider rapidement d'une manœuvre et de disposer du temps nécessaire pour l'exécuter. Les algorithmes de prise de décision permettent également d'optimiser la conduite en fonction des conditions de la route (par exemple, ajuster la vitesse en cas d'embouteillages).

Les voitures autonomes utilisent également des algorithmes afin de planifier une trajectoire optimale en fonction de l'environnement immédiat et de la destination souhaitée. Il s'agit de déterminer la manière dont le véhicule va exécuter les tâches requises, sachant qu'il existe parfois plusieurs manières d'atteindre un objectif. Les algorithmes de planification de trajectoire tiennent aussi compte des obstacles ainsi que des conditions de la route telles que les feux de signalisation.

Après ces différentes étapes, le véhicule peut enfin agir en freinant, accélérant ou en orientant sa direction. Toutes ces étapes se déroulent naturellement en une fraction de seconde. Détection, planification et action représentent les trois phases d'un cycle qui se répète jusqu'à ce que le véhicule ait atteint sa destination.

Le fait de permettre à un système informatique de gérer la totalité de la conduite d'un véhicule exige bien évidemment la mise en place de systèmes de cybersécurité particulièrement performants afin d'éviter la prise de contrôle – parfois à distance – des véhicules autonomes par un tiers malveillant. Le véritable enjeu demeure cependant la détection, car, sans elle, il ne peut y avoir de planification ni de manœuvre possible. Les constructeurs prennent un soin particulier à perfectionner ce matériel de détection afin de renforcer la sécurité des véhicules. Un autre enjeu consiste aussi à apprendre au logiciel à analyser pertinemment la quantité d'importantes de données qu'il reçoit en temps réel, à les trier, les recouper et les hiérarchiser pour construire une représentation en 3D de l'environnement de la voiture. Il est également nécessaire que des cartes haute définition très précises soient embarquées dans la voiture afin de disposer d'une description détaillée de l'environnement traversé par le véhicule : géométrie des voies, signalisation horizontale et verticale, trottoirs, limitations de vitesse, trajectoires, lois de circulation applicables, etc. Ces différentes technologies sont essentielles pour le bon fonctionnement d'une voiture autonome. Grâce à elles, cette dernière est capable de se repérer dans l'espace, et reconnaître les obstacles. Mais, elles ont aussi leurs limites. En effet, la voiture autonome repose sur un système informatique, il est donc impératif que ce système soit protégé contre les dysfonctionnements et les piratages.

74 Quels sont les différents niveaux d'autonomie ?

Les différents constructeurs prétendent avoir des voitures autonomes mais les niveaux d'autonomie peuvent varier de manière importante. Il existe différentes classifications techniques dont la plus connue demeure la norme SAE. L'association *Society of Automotive Engineers* (SAE), composée d'experts et d'ingénieurs dans le domaine de l'automobile et de l'aérospatiale, a créé une échelle à cinq niveaux pour répertorier le degré d'autonomie des IA embarquées dans les voitures autonomes. L'autonomie d'un véhicule se divise en 6 niveaux, allant du niveau 0 au niveau 5 en toutes circonstances et sur tout type de route.

Concernant tout d'abord le niveau 0, il n'y a dans ce cas aucun système automatisé. Le conducteur effectue dès lors seul toutes les actions de conduite. Ce niveau constitue un simple point de référence.

Le niveau 1 décrit une assistance à la conduite : le contrôle du véhicule est encore assuré par le conducteur, mais le véhicule peut prendre en charge certaines actions simples telles que le contrôle de la vitesse. L'exemple le plus parlant est le régulateur de vitesse ou encore le freinage automatique d'urgence, dont sont équipés l'immense majorité des véhicules modernes de nouvelle génération.

Le niveau 2, quant à lui, traduit une automatisation partielle de la conduite. Dans ce cas, le véhicule peut remplacer le conducteur plus souvent, sachant qu'il peut accélérer, freiner et changer de direction en cas de besoin. Le conducteur doit cependant rester attentif et garder le contrôle du véhicule mais il peut lâcher temporairement le volant tant qu'il reste vigilant à son environnement de conduite. Le freinage automatique lorsque le véhicule détecte un risque immédiat de collision sur sa trajectoire (par exemple, un piéton qui traverse) en est un exemple. Il peut aussi s'agir des systèmes de maintien du véhicule au centre de la voie de circulation, de régulateurs de vitesse intelligents adaptant la vitesse à la situation, ou encore d'assistants de stationnement. Ce niveau d'automatisation existe depuis 2013 sur des modèles commercialisés auprès du grand public. Les véhicules qui assistent les conducteurs (niveaux 1 et 2 d'automatisation) sont, d'ailleurs, déjà disponibles sur le marché européen.

L'on parle ensuite d'une automatisation conditionnelle lorsqu'on atteint le niveau 3. Avec le niveau 2, le conducteur doit être prêt à prendre immédiatement en main la tâche de conduite si le système rencontre un problème qu'il ne peut pas gérer. Avec le niveau 3, le système doit pouvoir gérer la conduite dans la mesure où cela relève de son domaine de conception opérationnelle. Cela signifie que le rôle de l'homme doit être un repli. Dans ce cas, le conducteur peut déléguer la conduite du véhicule au système de conduite automatisé dans certaines conditions (par exemple, uniquement sur une autoroute, ou sur une route unidirectionnelle), mais doit rester vigilant et prêt à reprendre le véhicule en main au besoin. Si les conditions sont réunies, la voiture est complètement autonome : elle freine, accélère et gère la direction de la voiture.

Concernant le niveau 4, on atteint une automatisation élevée de la conduite. Dans ce cas, le système peut exécuter toutes les tâches de conduite (direction, accélération et freinage), surveille et prend en compte tous les changements intervenant dans les

conditions du trajet, du point de départ à la destination sans intervention du conducteur. Ce niveau d'automatisation n'est possible que sur certaines routes telles que les autoroutes et à des conditions météorologiques classiques. Les véhicules de niveau 4 dispensent le conducteur de tout devoir de vigilance et lui permettent d'effectuer d'autres tâches que la conduite. Ces systèmes sont aussi appelés « *mind off* ». Les voitures de niveau 4 ne circulent pas encore sur les routes et sont encore à l'état de prototype.

Quant au niveau 5, on atteint l'automatisation totale de la conduite. Le véhicule peut réaliser toutes les tâches de conduite en toutes circonstances, et sans aucune intervention du conducteur qui devient un passager. Quelques constructeurs, y compris Tesla et BMW, avaient affirmé qu'ils disposeraient de systèmes de niveau 5 dans quelques années, mais de nombreux experts estiment qu'il sera difficile d'atteindre ce niveau d'autonomie.

Malgré les investissements et l'enthousiasme, il existe des freins réels au développement de cette automatisation totale. D'une part, il peut être difficile de faire circuler une voiture autonome dans un environnement urbain aux côtés de voitures non automatisées. D'autre part, il existe aussi un frein humain, car il convient de créer la confiance envers ces véhicules totalement autonomes. C'est la raison pour laquelle de nombreux spécialistes ne s'attendent pas à voir de tels véhicules avant 2040-2050.

Si cette automatisation totale de la conduite du véhicule individuel peine à se concrétiser, le déploiement de transports autonomes collectifs semble plus prometteur. En effet, il est d'ailleurs au cœur de la Stratégie nationale 2023-2025 de développement de la mobilité routière automatisée et connectée. Des projets de transports collectifs autonomes sont lancés dans de nombreuses villes. La ville de Châteauroux prévoit, par exemple, de mettre en service en 2026 six minibus électriques automatisés de niveau 4, c'est-à-dire sans opérateur de sécurité à bord.

Concernant les trains entièrement autonomes, s'ils avaient aussi été annoncés en 2021, puis en 2023, il existe de nombreux freins dont notamment la présence de plusieurs types de trains sur le réseau ferroviaire (fret, TGV, TER). Le premier train de fret sans conducteur a été mis en service en Australie depuis 2019, mais sans autre trafic sur cette voie ferrée. Le frein psychologique est aussi présent dans ce cas de figure, car les opérateurs n'imaginent pas laisser un train rouler sans présence humaine pour des problèmes de sécurité, y compris pour le fret.

Lorsqu'on sait que plus de 90 % des accidents de la circulation impliquent un facteur humain, il est assez logique qu'il y ait une certaine réticence de la part des usagers à l'idée d'abandonner la conduite à un véhicule autonome, malgré les importants progrès réalisés dans ce domaine. C'est la raison pour laquelle les constructeurs développent davantage des systèmes d'assistance à la conduite qui semblent plus rassurants pour les usagers.

75 La voiture autonome cause-t-elle moins d'accidents ?

90 % des accidents sont causés par l'erreur humaine, notamment l'ébriété, l'inattention, le non-respect volontaire des règles de circulation et la fatigue. Une voiture autonome a normalement moins de risques d'être distraite qu'un humain et est conçue pour limiter les accidents et respecter toujours le Code de la sécurité routière. En supprimant l'inévitable erreur humaine, certaines études prévoient que la voiture autonome pourrait réduire par dix le nombre d'accidents. Aucun élément concret n'accompagne pour autant ces estimations. Ces études se fondent sur l'idée que ces réductions seraient obtenues lorsque ne circuleront plus que des véhicules autonomes de niveau 4 et 5 fiables.

Le défi restera toutefois de faire cohabiter les voitures autonomes avec les conducteurs, cyclistes ou piétons dont les comportements sont parfois imprévisibles. Des chercheurs des universités de Copenhague, Linköping et du King's College ont démontré en analysant des vidéos de situations réelles que les véhicules sans conducteur sont incapables de gérer des contextes ambigus que les humains savent régler, ce qui peut nuire à la circulation voire provoquer des accidents.

Les accidents de voitures autonomes font régulièrement la une de l'actualité. Le premier, et le plus connu a eu lieu en 2018, a été causé par la Volvo XC90 testée par Uber qui, de nuit, a renversé et tué une femme traversant la route en dehors d'un passage pour piétons et poussant son vélo. Un conducteur était à bord du véhicule qui roulait en mode autonome. Le véhicule aurait dû avoir une remontée d'information des capteurs et une réaction immédiate pour éviter l'accident. Le conducteur, quant à lui, n'était pas assez attentif au moment des faits et n'a pas pu freiner à temps. La défaillance trouve sa source dans une combinaison de facteurs : les capteurs hésitant dans l'analyse de cet obstacle soudain et son mouvement dans un environnement sans perturbations.

Malgré cet accident, et d'autres qui se sont multipliés au fil des ans, les tests sur ces véhicules se poursuivent. Le plus souvent, il s'agit de systèmes d'aide à la conduite semi-autonome et non d'une conduite automatisée, ce qui peut parfois prêter à confusion pour les automobilistes augmentant ainsi leur temps de réaction. S'agissant de véhicules expérimentaux, il faut s'attendre à déplorer quelques accidents matériels, mais la multiplication d'accidents mortels incite à la prudence.

Dans une étude récente publiée en 2024 dans la revue *Nature Communication*, des chercheurs de l'université de Floride centrale ont pourtant conclu que la conduite autonome est généralement plus sûre que la conduite humaine dans des conditions normales. Il s'agit d'une des plus importantes études comparatives sur les accidents de la route, comparant les données d'accidents de 2 100 véhicules autonomes et de 35 113 véhicules conduits par des humains en Californie entre 2016 et 2022, où des essais sur route de véhicules autonomes ont lieu depuis plusieurs années. Les chercheurs ont également tenté de comparer les accidents en fonction des conditions de la conduite : météo, moment de la journée, à une intersection ou dans une ligne droite, etc.

Cette recherche a révélé que les véhicules autonomes, ou à conduite autonome, présentent des taux d'accidents inférieurs à ceux des humains dans presque tous les scénarios. Cependant, elle indique que les véhicules autonomes semblent plus sujets aux

accidents dans des situations spécifiques. En effet, les véhicules autonomes seraient cinq fois plus susceptibles d'avoir un accident que les véhicules conduits par des humains dans des conditions de faible luminosité, à l'aube ou au crépuscule. De même, le taux d'accident est presque deux fois plus élevé que celui des conducteurs humains lorsque les véhicules autonomes prenaient des virages.

Les auteurs de l'étude soulignent, à ce titre, que l'un des obstacles à leur étude est que la base de données sur les accidents des véhicules autonomes est relativement limitée. Il est donc nécessaire de disposer de ces données en améliorant les rapports sur les accidents des véhicules autonomes. Certains accidents impliquant les véhicules autonomes ne sont en effet pas signalés à la police s'il ne s'agit, par exemple, que d'accrochages mineurs. Le défaut d'exhaustivité de ces données doit ainsi conduire à considérer avec une certaine réserve l'idée selon laquelle les véhicules autonomes seraient plus sûrs que les conducteurs humains, d'autant plus qu'ils circulent souvent dans des zones et des environnements particuliers, ce qui rend difficile la généralisation des résultats actuels. Par ailleurs, ces véhicules autonomes vont devoir pendant un certain temps cohabiter avec les véhicules conduits intégralement pas l'humain, ce qui peut faire apparaître de nouveaux types d'accidents résultant de l'interaction entre des véhicules avec des modes d'autonomie de niveaux différents et des véhicules conventionnels. Il est à noter également que les routes ne sont pas équipées de suffisamment de capteurs, d'antennes et autres éléments permettant d'assurer une communication en continu entre les différentes automobiles. Ces équipements sont pourtant importants pour faciliter le trajet de tels véhicules.

L'on retrouve aussi quelques études issues de l'industrie, dont celle menée en 2022 par Swiss Re et Waymo s'appuyant sur 3,8 millions de kilomètres parcourus à travers San Francisco et Phoenix par des véhicules autonomes, sans conducteur derrière le volant. Cette étude avait conclu à 76 % de réduction de la fréquence des sinistres liés aux dommages matériels par rapport aux standards avec conducteur humain.

Face aux promesses de baisse de l'accidentalité, il y aura aussi parallèlement une exigence plus forte de la population en termes de sécurité. Certains accidents médiatisés de voitures autonomes ont d'ailleurs déjà conduit à ce que certains riverains vandalisent, voire incendient ces véhicules. Les victimes d'accidents risquent aussi d'avoir plus de mal à accepter ces accidents pour lesquels il sera plus difficile de rechercher les responsabilités. Jusqu'à présent, il était en effet commode de designer le conducteur comme l'unique responsable. Par conséquent, l'on peut penser que le déploiement de ces véhicules plus ou moins autonomes dans la circulation va peut-être réduire l'accidentalité, mais probablement pas dans les proportions annoncées.

76 Quels sont les enjeux éthiques liés à la collecte et au traitement des données à caractère personnel ?

Les véhicules à conduite automatisée perçoivent leur environnement en permanence grâce à différents capteurs visuels et de distance pour leur besoin de navigation. Ils collectent aussi des données sur leurs utilisateurs ainsi que sur les autres usagers de la voie publique. Ces traitements de données posent des différentes questions liées à la protection des données à caractère personnel.

En effet, plus la voiture atteint un niveau d'autonomie élevé, plus ses capacités de traitement des données sensibles de son utilisateur augmentent. Pour atteindre d'ailleurs le niveau 5, le traitement maximal des données à caractère personnel demeure une pierre angulaire. C'est en effet le traitement, par des outils d'intelligence artificielle, des données collectées par la voiture de son environnement externe et interne qui permet de l'automatiser. Il ne peut donc exister de voiture autonome sans collecte et traitement des données personnelles du conducteur. Une voiture entièrement automatisée, de niveau 5, sera ainsi en mesure de géolocaliser en temps réel son utilisateur, opérer un traçage de ses habitudes de vie, voire de détecter le nombre de passagers dans le véhicule. Plus généralement, la collecte de ces données risque de conduire au développement graduel d'une surveillance généralisée et permanente.

Sans atteindre le degré d'autonomie qui caractérise la voiture 5, la voiture actuelle collecte et stocke déjà une quantité importante de données. Le GPS enregistre, par exemple, les trajets quotidiens du conducteur. De même, le système à bord peut enregistrer les données relatives au téléphone, s'il est synchronisé en Bluetooth au système, pour diffuser de la musique ou passer des appels en mode mains-libres. Les conducteurs et les passagers ne sont pas toujours correctement informés du traitement des données effectué dans un tel véhicule ou par l'intermédiaire de celui-ci.

Le risque avec la voiture autonome s'accentue au regard de la nature des données collectées qui ne se réduisent pas aux données personnelles de l'utilisateur, mais concernent aussi le flux continu des données échangées en temps réel entre la voiture autonome et son environnement. Ces données sont relatives aux conducteurs (vitesse moyenne, durée de la conduite, état de fatigue…), à celle des autres véhicules mais également celles relatives au temps, à la route, etc. Les exemples les plus emblématiques sont très certainement les différents modèles de Tesla dans lesquels le rôle de la donnée est central pour le fonctionnement du véhicule. De plus, la voiture ne se limite plus à traiter les données en interne, elle peut les transférer aussi à des acteurs extérieurs.

Dans la mesure où ces systèmes permettent, voire nécessitent, le traitement de données, ils doivent satisfaire aux exigences juridiques en matière de protection de données personnelles, notamment au RGPD. Les constructeurs des voitures autonomes ne pourront ainsi pas collecter une quantité maximale de données, car ils seront bridés par les dispositions du règlement qui consacrent le principe de la minimisation de la collecte ou encore les concepts de *privacy by design* et *by default*, lesquels supposent une pseudonymisation ou une anonymisation des données. Ainsi, les principes du *privacy by*

design ainsi que celui de la minimisation de la collecte de données imposeront de prendre en compte ces exigences dès la conception des radars, caméras ou autres capteurs capables de collecter ces informations.

Après la protection des données à caractère personnel mise en œuvre par le règlement européen sur la protection des données en 2018, l'Union européenne s'est aussi intéressée aux données industrielles et à celles produites par l'internet des objets. Ce qui a conduit à l'adoption du règlement relatif à des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données, également connu sous le nom de règlement sur les données ou *Data Act*, du 13 décembre 2023 qui est entré en vigueur le 11 janvier 2024. Il sera applicable à partir du 12 septembre 2025. Il s'agit du premier règlement européen sur les données de tous les appareils connectés, y compris les automobiles. Cette nouvelle réglementation vise à faciliter le transfert continu de données entre les détenteurs de données et les utilisateurs de données tout en préservant leur confidentialité.

Depuis 2016, la CNIL s'intéresse également à cette problématique. Elle a publié un référentiel sectoriel en 2017 permettant aux professionnels de se mettre en conformité avec le RGPD, intitulé pack de conformité « véhicules connectés et données personnelles ». Il constitue pour ces professionnels une boîte à outils leur permettant d'intégrer la protection des données personnelles dès la phase de conception des produits et d'assurer la maîtrise par les personnes de leurs données. Le concepteur du véhicule doit ainsi s'interroger sur le traitement qui va être fait des données de l'utilisateur. Pour l'y aider, le pack envisage trois scénarios différents : le traitement local des données ; la transmission des données à un tiers à l'extérieur du véhicule pour fournir un service à la personne concernée ; la transmission des données à l'extérieur pour déclencher une action automatique dans le véhicule. À chaque scénario correspond un régime de traitement spécifique, au sein duquel la CNIL met en garde contre les éventuels dangers ou abus et dresse une liste de recommandations et bonnes pratiques adaptées au scénario retenu. Ce pack de conformité ainsi que les lignes directrices du Comité européen de la protection des données (CEPD) sur les véhicules connectés et les applications liées à la mobilité, publiés en 2020, constituent les documents de référence dans le secteur servant à guider les constructeurs dans leur démarche de conformité au RGPD.

En mars 2023, la CNIL a d'ailleurs créé un « club conformité » dédié aux acteurs du véhicule connecté et de la mobilité. En regroupant les principales parties prenantes en France, elle souhaite promouvoir une utilisation responsable des données personnelles.

77 Qui est responsable en cas d'accident d'une voiture autonome ?

Le développement des voitures autonomes soulève un problème de qualification des personnes qui les utilisent. Les textes français accordent en effet un rôle central au « conducteur » du véhicule. C'est sur ce dernier que pèse la responsabilité civile en cas d'accident.

La Convention de Vienne sur la circulation routière du 8 novembre 1968 définit, à cet effet, très largement la notion de conducteur. Cette notion est définie à son article premier comme étant « *toute personne qui assume la direction d'un véhicule, automobile ou autre (cycle compris), ou qui, sur une route, guide des bestiaux, isolés ou en troupeaux, ou des animaux de trait, de charge ou de selle* ». Cette définition renvoie à l'idée de contrôle et de maîtrise du véhicule par une personne physique. La localisation du conducteur n'a pas été précisée, car celui-ci était *a priori* dans son véhicule comme le conducteur d'une voiture à cheval était supposé être sur le siège de la voiture ou sur le cheval (moteur) la propulsant. À noter que guider des bestiaux peut techniquement se faire à distance, notamment par le truchement de chiens spécialisés opérant à proximité immédiate du troupeau, ce qui présente certaines similitudes potentielles avec le fonctionnement des véhicules autonomes. L'on pourrait arguer que l'IA du véhicule joue ainsi le rôle du chien et que son maître – et conducteur – est celui lui ayant donné des ordres.

On retrouve cette idée de maîtrise également dans la jurisprudence française. Il n'existe pas, à ce titre, de définition légale du conducteur dans le Code de la route, mais il paraît résulter de la jurisprudence que la qualification de conducteur suppose que la personne concernée conserve une certaine maîtrise du véhicule. La personne physique désignée comme conducteur est ainsi celle qui, au moment de l'accident, a la possibilité de maîtriser le véhicule et de le contrôler en disposant des pouvoirs de commandement.

Si l'autonomie du véhicule demeure faible, le conducteur conservera *a priori* une maîtrise suffisante du véhicule. À l'inverse, lorsque le niveau d'autonomie est plus élevé pour le décharger totalement ou partiellement de la conduite du véhicule, l'on peut se demander s'il peut toujours être qualifié de conducteur. Si la notion de conducteur peut encore avoir un sens en présence d'une voiture autonome de niveau 3, elle peut sembler au contraire inadaptée à la voiture autonome de niveau 4 ou 5.

Chacun comprend que l'identité du responsable pénal constitue un enjeu crucial pour le développement des voitures autonomes. La France est d'ailleurs loin d'être en retard sur cette question. L'ordonnance du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation et son décret d'application du 29 juin 2021 ont ainsi adapté le Code de la route pour permettre la circulation sur les routes françaises des véhicules équipés de système à délégation de conduite, dès leur homologation, et des systèmes de transport routier automatisés sur parcours ou zones prédéfinis dès le 1er septembre 2022.

Depuis cette date, est ainsi autorisée la circulation des véhicules autonome, c'est-à-dire ceux dont le « contrôle dynamique » peut être délégué, durant une période prolongée, à un système de conduite automatisé, ce qui correspond au niveau 3. Les différents

capteurs, caméras et radars présents dans le véhicule autonome permettront à un logiciel de diriger le véhicule en temps réel, sans intervention nécessaire du conducteur. Si ces dispositions permettent théoriquement la circulation de véhicules dotés du niveau 3 sur nos routes, des conditions contraignantes doivent cependant être réunies. En effet, la conduite autonome du véhicule est permise uniquement dans les conditions imposées par un règlement de l'ONU relatif à l'homologation des véhicules automatisés du 22 janvier 2021. Ce dernier exige que le véhicule se trouve sur une route où les piétons et cyclistes sont interdits et qui est équipée d'une chaussée séparée sachant que sa vitesse ne doit pas dépasser 60 km/h. Concrètement, peu de modèles correspondant à ce niveau d'autonomie ont été homologués et commercialisés. À l'heure actuelle, l'autorisation de circulation des véhicules autonomes est donc essentiellement d'ordre théorique, mais ce cadre juridique pose les bases pour les années à venir.

Par ailleurs, l'ordonnance du 14 avril 2021 est aussi venue préciser comment fonctionnera le partage de responsabilité en cas d'infraction ou d'accident entre le constructeur et le conducteur. Le nouvel article L. 123-1 du Code de la route prévoit, à ce titre, que le conducteur n'est pas responsable des infractions pénales constatées lors de la conduite du véhicule lorsque le système automatisé exerçait, au moment des faits, le contrôle dynamique du véhicule. C'est sur le constructeur du véhicule que pourrait se reporter, le cas échéant, la responsabilité pénale ou le paiement de l'amende encourue. Le législateur désigne désormais le constructeur du véhicule ou son mandataire comme pénalement responsable des délits d'atteinte involontaire à la vie ou à l'intégrité de la personne. L'article L. 123-2 du Code de la route dispose ainsi que le constructeur est pénalement responsable des délits d'atteinte involontaire à la vie ou à l'intégrité d'une personne, à condition que puisse lui être imputée une faute personnelle.

En revanche, le conducteur redevient pénalement responsable lorsqu'il reprend la main du système de conduite automatisée. En outre, ce dernier doit se tenir constamment en état et en position de répondre à une demande de reprise en main. Le conducteur d'un véhicule à délégation de conduite est en effet chargé de superviser ce véhicule et son environnement de conduite. À bord comme à l'extérieur du véhicule, il doit être prêt à tout moment à prendre le contrôle du véhicule afin d'effectuer les manœuvres nécessaires à la mise en sécurité du véhicule, de ses occupants et des usagers de la route.

Si cette nouvelle législation spéciale n'est certes pas sans susciter quelques interrogations, elle a cependant le mérite d'exonérer clairement le conducteur de toute responsabilité pénale lorsqu'il avait délégué, au moment des faits, le contrôle dynamique du véhicule. Afin d'établir cette responsabilité pénale du constructeur, il est indispensable de pouvoir accéder aux données du dispositif d'enregistrement des données du système de délégation de conduite. Pour ce faire, le conducteur et les forces de l'ordre doivent être en mesure d'accéder à la « boîte noire » obligatoirement présente à bord du véhicule autonome. Si des évolutions en matière de droit pénal sont encore attendues pour adapter le régime de responsabilité aux véhicules autonomes, il est certain que ces nouvelles dispositions ouvrent la voie progressivement à la voiture autonome.

78 Quel est le régime de responsabilité civile applicable ?

La victime d'un accident de la circulation impliquant un véhicule terrestre à moteur dispose d'un droit à indemnisation à l'encontre du gardien ou conducteur dudit véhicule. Sur le plan de la responsabilité civile, la loi du 5 juillet 1985, dite « *loi Badinter* », a en effet pour une grande part écarté le schéma classique de la responsabilité civile en prévoyant que l'indemnisation des victimes d'accidents de la circulation aurait lieu sans recherche de responsabilité, au moyen d'un critère de désignation objectif des débiteurs de l'indemnisation, à savoir l'« implication » du véhicule dans l'accident. Cette loi instaure ainsi une responsabilité sans faute applicable dès lors que survient un accident de la circulation dans lequel se trouve impliqué un véhicule terrestre à moteur. Sans devoir nécessairement rechercher une faute, l'imputabilité impose d'identifier un responsable. De ce fait, le droit français organise un système de responsabilité spéciale particulièrement favorable aux victimes en cas d'accident de la circulation dans lequel est impliqué un véhicule terrestre à moteur : les victimes ont vocation à être indemnisées rapidement et simplement par un assureur. Pour faciliter l'indemnisation des victimes, celles-ci ne peuvent en principe se voir opposer la force majeure ou le fait d'un tiers par le conducteur ou le gardien du véhicule, tout comme elles ne peuvent se voir opposer leur propre faute, à l'exception de leur faute inexcusable si elle a été la cause exclusive de l'accident.

Les différentes exigences de la loi Badinter ne semblent pas de nature à faire obstacle à l'indemnisation de la victime d'un accident causé par une voiture autonome. En prévoyant la présence d'un conducteur à bord des véhicules autonomes, la mise en œuvre de cette loi ne devrait pas susciter de difficultés particulières. Il ne fait pas de doute que le véhicule autonome est bien un véhicule terrestre à moteur dont les dommages entrent dans le champ de la loi de 1985 chaque fois qu'un fait de circulation peut être caractérisé. Aussi, si un véhicule autonome est impliqué dans un éventuel accident de la circulation, son conducteur sera responsable, peu important que le contrôle dynamique du véhicule ait été délégué au système automatisé lors de la survenue de l'accident. L'implication est normalement établie en cas de contact entre le véhicule et le siège du dommage, mais rien n'empêche qu'elle soit aussi retenue en l'absence de collision, dès lors que le véhicule est intervenu à quelque titre que ce soit dans la réalisation de l'accident.

Contrairement au droit pénal, une telle délégation ne devrait pas permettre au conducteur d'échapper à sa responsabilité civile, cette modalité de conduite automatisée ne figurant pas dans la liste des causes exonératoires de responsabilité énumérées par la loi Badinter. Si cette loi semble capable de s'accommoder de la voiture autonome de niveau 3, elle paraît inadaptée à la voiture autonome de niveau 4 ou 5. La qualification de conducteur peut-elle être retenue lorsque la personne ayant enclenché la conduite automatique sur autoroute de son véhicule de niveau 3 en profite, par exemple, pour répondre à ses courriels ? Il en va de même pour la conduite d'un véhicule autonome de niveau 4, peut-on considérer comme conducteur la personne qui décide de dormir pendant le trajet préalablement enregistré ? Autrement dit, plus le niveau d'autonomie est développé, plus la notion de conducteur devient floue. Pour les véhicules autonomes

de transport de passagers circulant sans aucun être humain contrôlant la conduite pendant le voyage, il conviendra, avant leur arrivée en Europe, de prévoir une adaptation de la notion de conducteur pour tenir compte de cette évolution.

De plus, dans la mesure où l'assurance du conducteur paie en toutes circonstances, il faut prendre en compte le fait que, dans une voiture autonome de niveau 3, le mode d'autonomie risque d'être souvent enclenché. Cette situation aura des incidences sur les primes d'assurance payées par le conducteur qui seront plus élevées pour un véhicule de cette catégorie. Le système bonus/malus sera aussi remis en cause. Du point de vue de l'assureur du conducteur ayant indemnisé la victime, celui-ci ne devrait pas nécessairement supporter définitivement le poids de la dette.

Il bénéficie en effet d'une subrogation lui permettant d'exercer un recours contre le constructeur du véhicule autonome si l'accident a pour origine, entièrement ou partiellement, une défaillance du système de conduite automatisé, notamment sur le fondement d'une responsabilité du fait des produits défectueux. Cette situation risque d'entraîner des recours en cascade qui encombreront les tribunaux.

Si le droit positif apparaît pour le moment adapté au déploiement des véhicules partiellement autonomes, il n'en va pas de même s'agissant des véhicules totalement automatisés dans lesquels la présence d'un conducteur pourrait n'être qu'une simple option. Lorsqu'une telle évolution verra le jour, il sera alors nécessaire de repenser et adapter les règles en vigueur.

79 Une voiture autonome doit-elle faire des choix moraux ?

Les véhicules autonomes pourraient devoir décider qui vit et qui meurt en cas de défaillance des freins ou de collision inévitable. Percuter un enfant ou heurter un mur en risquant de tuer les passagers ? L'ordre doit-il être de percuter cet enfant si la tentative de son évitement provoque un accident mortel pour les passagers du véhicule ? Il s'agit là du genre de dilemme auquel sont confrontées les voitures autonomes, plus généralement ceux qui les programment. Ces choix peuvent s'avérer cornéliens puisqu'il impliquerait le sacrifice injustifié d'une personne.

C'est afin de répondre à ce dilemme moral (connu sous le nom du « dilemme du Tramway ») que les chercheurs du MIT ont lancé en 2016 une expérience participative appelée *The Moral Machine* en vue d'interroger les internautes sur les réactions qu'ils espèrent de la part des voitures autonomes dans différents scénarios. Les chercheurs ont ainsi pu récolter 40 millions de décisions auprès de millions d'internautes du monde entier. L'expérience en ligne se présentait sous forme de jeu confrontant les internautes du monde entier à des choix moraux impliquant, d'une part, voiture et passagers et piétons de tout type, de l'autre : enfant, adulte, chat, chien, délinquant, homme, femme, femme enceinte, docteur, sans-abri, gros, cadre, piéton traversant hors des clous. Les réponses apportées ont également fait l'objet d'une étude, publiée en 2018 dans la revue *Nature* intitulée « *L'expérience de la Machine morale* ».

Cette expérience a démontré que les différences culturelles rendent difficile l'établissement d'une éthique universelle, car les décisions sont clairement différentes selon les pays d'appartenance. Ces variations entre les pays s'expliquent potentiellement par des facteurs culturels et économiques. D'après l'étude, il existerait trois principaux groupes de pays au sein desquels les choix des internautes sont proches : groupe ouest (Amérique du Nord, pays européens mais pas tous), groupe est (Asie, monde musulman) et groupe sud (Amérique Latine, la France et les pays marqués historiquement par son influence). L'analyse de ces données a permis d'identifier trois critères moraux principaux : sauver des vies humaines plutôt que des animaux, sauver le plus grand nombre de vies et sauver les vies des plus jeunes plutôt que celles des personnes âgées.

Trois tendances globales se dégagent : entre des humains et des animaux, les internautes épargnent les premiers ; ils privilégient l'action préservant le plus de vies possibles ; ils sacrifient plus souvent les personnes âgées quand la vie d'enfants est en jeu. La préférence pour sauver les jeunes plutôt que les vieux est plus prononcée, par exemple, dans les pays du groupe sud, dont fait partie la France, que le groupe est. Les profils les plus sauvés dans les situations proposées sont d'ailleurs les bébés en poussette, les enfants et les femmes enceintes.

Les résultats montrent également des préférences morales plus discutables : les personnes en surpoids ont plus de probabilités de se faire tuer face à des personnes athlétiques tout comme les pauvres face aux riches. Les personnes ne respectant pas les feux de signalisation ont aussi tendance à être plus facilement sacrifiées.

Dans la mesure où les hommes ne peuvent s'entendre sur des principes éthiques communs, certains s'inquiètent de la vitesse à laquelle les véhicules autonomes sont déployés. Pour l'heure, ces décisions sont surtout prises par les constructeurs de véhicules autonomes. Le Comité national pilote d'éthique du numérique, pérennisé et dénommé le Comité consultatif national d'éthique du numérique par un décret du 23 mai 2024, avait rendu, à ce titre, un avis le 20 mai 2021 intitulé « Le véhicule autonome : enjeux d'éthique ». Il précise dans cet avis que les débats relatifs aux dilemmes se fondent sur l'idée que le calcul automatique d'une décision peut être calqué sur le raisonnement moral d'un être humain alors qu'ils sont de nature différente. En effet, les actions du véhicule sont déterminées par des algorithmes prédéfinis par le concepteur ; cela ne fait pas du véhicule un agent moral. Le Comité insiste sur le fait que l'utilisation d'un vocabulaire pour décrire des caractéristiques humaines (« décision de tuer ») est susceptible de projeter, de manière contestable, une moralité sur le véhicule. Il semble ainsi indiquer qu'il serait préférable de soustraire les véhicules aux projections de moralité, ce qui signifie de ne pas intégrer d'algorithme de choix de la victime qui serait géré par le véhicule.

Le risque étant d'avoir des critères de programmation différents suivant les constructeurs, il est donc nécessaire que les pouvoirs publics s'emparent de cette question afin de définir un cadre sécurisé et uniforme pour le déploiement de ces véhicules.

80 La voiture autonome est-elle écologique ?

Les voitures autonomes soulèvent des problématiques environnementales non négligeables. Elles sont pourtant présentées comme le futur de la mobilité motorisée, que ce soit en termes de sécurité, de services et d'écologie. Les promesses écologiques sont souvent mises en avant lorsqu'on parle du véhicule autonome. Ses promoteurs mettent en effet en avant la conduite plus sobre d'un tel véhicule, la réduction du poids liée à la suppression d'éléments de sécurité dans la carrosserie ainsi que la complémentarité de ces véhicules avec les transports en commun.

Une étude du Forum Vies Mobiles et de La Fabrique écologique de mars 2021 démontre cependant que cette innovation n'aidera pas à la décarbonisation des transports, voire pourrait aggraver la situation. Ce rapport réfute cet argument en démontrant que ces véhicules pourraient augmenter la consommation énergétique, car étant équipés de caméras et de divers capteurs, ces voitures doivent effectuer d'innombrables opérations en temps réel pour analyser les conditions de circulation et piloter le véhicule. L'augmentation de la consommation énergétique serait ainsi plus probable que sa réduction.

Pour analyser plus finement l'impact écologique du véhicule autonome, trois scénarios ont été étudiés : une mobilité individuelle avec des voitures à usage privé ; une mobilité à la demande s'appuyant sur des flottes de robots-taxis ; une mobilité collective avec des navettes autonomes pour le transport collectif.

S'agissant tout d'abord du véhicule autonome individuel, le rapport relève que libéré de la conduite, l'automobiliste pourrait utiliser la voiture autonome bien plus souvent qu'une voiture conventionnelle, car il serait plus disponible pour développer d'autres activités à bord du véhicule. Du fait de l'absence de conducteur lorsque le véhicule sera intégralement autonome, un changement dans les modes de vie risque de s'opérer avec une utilisation plus intensive du véhicule, des déplacements plus nombreux et plus lointains puisque moins contraignants. Cette utilisation intensive pourrait engendrer une augmentation des émissions de gaz à effet de serre et de la pollution, et ce malgré l'usage de voitures électriques. Par conséquent, ce scénario présente le risque d'une augmentation croissante de véhicules sur les routes entraînant une augmentation de la production de véhicules et d'énergie pour les alimenter.

Selon une étude du MIT publiée en 2023, si un milliard de voitures autonomes conduisent une heure par jour, celles-ci généreraient la même quantité de gaz à effet de serre que tous les datacenters mondiaux, principalement en raison de l'électricité nécessaire pour alimenter leurs systèmes de guidage automatique. Le MIT reconnaît toutefois que ces prédictions pourraient rapidement devenir obsolètes si la technologie parvient à limiter l'impact environnemental. Ainsi, le véhicule autonome n'apparaît pas comme une solution idéale dans un contexte de transition énergétique.

Le second scénario concerne celui des robots-taxis en zone urbaine dans lequel l'utilisateur serait le client d'une entreprise privée. Ce scénario soulève plusieurs difficultés dont le tarif du service qui serait appliqué par l'entreprise. Il est en effet possible que le prix de ces robots-taxis ne soit pas abordable pour une partie de la

population, ce qui pourrait accroître non seulement les inégalités sociales, mais aussi territoriales dans la mesure où les territoires desservis seront sélectionnés pour leur rentabilité. Se pose également la question de la protection des données à caractère personnel que l'utilisateur doit communiquer à l'opérateur privé, celles-ci risquant d'être exploitées par les acteurs du numérique proposant des offres de mobilité autonome. Comme il a été démontré pour le véhicule autonome individuel, le recours à un service de robots-taxis pourrait augmenter les distances jugées acceptables pour les trajets du quotidien, car le temps de conduite pourrait être remplacé par d'autres activités. Le partage du taxi avec d'autres usagers diminue cependant son attrait par rapport à l'utilisation d'une voiture individuelle, même s'il peut conduire à la fin de la propriété individuelle d'un véhicule. Le fait de ne plus posséder de véhicule personnel peut engendrer pourtant des effets pervers, car cela suppose de vivre dans des zones denses qui garantissent la disponibilité de tels services.

Concernant enfin le scénario du développement de navettes autonomes pour le transport collectif des voyageurs, il s'agit du modèle qui est souvent mis en avant dans la communication des pouvoirs publics qui suppose une intervention forte de l'État, notamment en termes d'investissement, pour développer un tel modèle. Ces solutions de mobilité pourraient certes faciliter l'utilisation des transports en commun avec pour conséquence le recul de la place accordée à la voiture individuelle. Cela entraînerait toutefois le même effet pervers qui est celui de l'augmentation des habitations en zone dense ou desservies par les transports en commun pour bénéficier du réseau de transports.

Même s'il est difficile de prédire ce que sera la mobilité autonome, il semblerait que le véhicule autonome ne peut être écologique que si sa production est limitée. En effet, son déploiement pourrait augmenter les émissions de carbone liées au transport, mais aussi, la production massive de véhicules, de matériel électronique et d'infrastructures, ainsi qu'au traitement d'une importante quantité de données. Il ressort de cette étude que la mobilité partagée, à savoir les navettes autonomes, ou dans une certaine mesure les flottes de robots-taxis, semble être le meilleur moyen d'améliorer la mobilité tout en maintenant les objectifs de décarbonisation. Dans une perspective écologique, les pouvoirs publics doivent donc concevoir et déployer un nouveau système de mobilité combinant différents modes de transports afin de faciliter les déplacements et limiter les temps de trajet.

IX.
LES ŒUVRES ET LES INVENTIONS DE L'INTELLIGENCE ARTIFICIELLE

81 Quel est l'impact de l'intelligence artificielle sur la création artistique ?

Plus d'une trentaine d'intelligences artificielles sont désormais capables de générer des œuvres d'art à partir d'une simple phrase fournie à la machine, appelée un « prompt ». Les œuvres créées peuvent s'inspirer de nombreux styles artistiques en puisant leur inspiration dans une vaste base de données.

Les algorithmes d'apprentissage automatique analysent en effet des milliers d'œuvres d'art pour comprendre les motifs, les styles et les techniques afin de créer de nouvelles pièces en s'inspirant de ces données, tout en innovant. Un exemple notable est celui du célèbre tableau du peintre Johannes Vermeer « *La Jeune fille à la perle* », interprété par Julian van Dieken, qui l'a réalisé à l'aide de MidJourney et Photoshop, et qui est désormais exposé au Musée Mauritshuis.

Pour autant, il est difficile de parler de création au sens strict, il s'agirait plutôt d'une production à partir de données. En effet, l'intelligence artificielle peut assimiler le style d'un peintre. Par exemple, une photo ordinaire peut être transformée en une œuvre d'art à la manière de Van Gogh ou de Picasso. En analysant les motifs, les couleurs et les techniques de l'artiste, l'intelligence artificielle peut aussi générer une peinture qui ressemble à une véritable œuvre d'un peintre célèbre. Ce fut le cas en 2016 avec le projet *The Next Rembrandt* qui a pu produire une œuvre à la manière du maître mais totalement inédite.

Les artistes peuvent ainsi utiliser l'intelligence artificielle pour générer des éléments, tels que des compositions musicales ou des éléments visuels, afin de les intégrer à leur œuvre. Cette collaboration entre l'homme et la machine ouvre de nouvelles perspectives pour l'art en permettant aux artistes d'expérimenter des idées qui auraient été impossibles à réaliser sans cet outil. Par exemple, l'artiste Refik Anadol utilise l'IA pour créer des installations d'art numérique dynamiques. Ses œuvres utilisent des données environnementales en temps réel pour générer des visuels en constante évolution afin de créer une expérience immersive évoluant constamment. Loin de remplacer les artistes, l'intelligence artificielle leur permet au contraire de repousser les limites de leur créativité. Il est en effet important de rappeler que l'intelligence artificielle ne se substitue pas à la créativité humaine ; les algorithmes ne pouvant pas remplacer l'émotion et l'expression humaine qui sont au fondement de la création artistique. Les artistes doivent toujours apporter leur propre esthétique à leurs œuvres.

Au-delà de la simple transcription de « prompt », l'intelligence artificielle peut avoir d'autres d'applications dans le domaine artistique. Elle peut en effet être utilisée pour authentifier des œuvres. *Art Recognition*, une intelligence artificielle mise au point par une entreprise suisse, a réussi, par exemple, en 2021, à déterminer l'origine d'un tableau de Renoir dont l'authenticité était contestée. Par ailleurs, l'intelligence artificielle peut aussi établir des liens entre différentes créations issues de diverses époques afin de soutenir le travail des historiens de l'art. Elle peut enfin analyser les émotions suscitées par un tableau ou encore prédire les sentiments qu'un spectateur pourrait ressentir face à une œuvre. C'est le cas de l'intelligence artificielle nommée *ArtEmis* qui sait reconnaître

une émotion et motiver par écrit ce qu'elle « ressent » quand elle observe une œuvre. Ses réponses se basent sur les émotions recueillies, dans le cadre d'une enquête, auprès de milliers de personnes ayant observé des milliers d'œuvres sur WikiArt. Cette intelligence artificielle a été conçue pour accompagner le travail des artistes en leur permettant d'évaluer la manière dont le public pourrait percevoir leurs œuvres afin de vérifier si cet impact est bien celui escompté.

L'intelligence artificielle joue également un rôle important dans le domaine musical. Elle peut composer de la musique, l'arranger et jouer des instruments. Par exemple, AIVA (*Artificial Intelligence Virtual Artist*) est un système d'IA qui utilise l'apprentissage automatique pour composer de la musique classique pour films, jeux vidéo et autres médias. L'outil analyse différentes formes et genres de musique pour créer de nouvelles mélodies. Des plateformes comme *Jukedeck* permettent également de créer de la musique personnalisée basée sur les préférences, le genre et l'ambiance souhaités par les utilisateurs afin qu'ils les intègrent à leurs vidéos. L'IA peut aussi être utilisée pour la restauration d'enregistrements historiques et même l'enseignement musical. Le faux titre des chanteurs canadiens *Drake* et *The Weeknd*, réalisé grâce à une intelligence artificielle est, à ce titre, un exemple du potentiel de l'IA dans la création musicale. Cet exemple a aussi suscité la crainte de voir n'importe quel individu avec les bons outils en main recréer les voix d'artistes célèbres sans leur consentement. Ces outils offrent certes de nouvelles possibilités créatives, mais il convient de rester vigilant puisqu'ils soulèvent aussi des questions éthiques et juridiques.

Plus généralement, l'art généré par l'intelligence artificielle soulève des questions sur la nature de l'art et de la créativité, et sur la place de l'homme dans le processus de création artistique. La généralisation de l'utilisation de ces outils dans le processus de création artistique pourrait conduire à une dévaluation du travail de l'artiste du fait de la culture de la création rapide et une absence d'originalité des créations générées par l'intelligence artificielle. Si l'avenir de l'art est étroitement lié à l'évolution de l'IA, il est nécessaire qu'une réflexion soit menée afin que cet outil reste seulement au service de la créativité humaine.

82 Une intelligence artificielle peut-elle être considérée comme un auteur ?

Le développement de l'intelligence artificielle bouleverse de nombreux domaines, y compris celui de la création artistique. En effet, se développe ce que l'on qualifie d'« art génératif », ou encore d'« art algorithmique » avec des outils tels que ChatGPT, Midjourney, Stable Diffusion et autres qui ont ouvert de nouvelles perspectives créatives. Ces IA génératives fonctionnent grâce à l'apprentissage automatique, plus spécifiquement au *machine learning* qui est essentiel, car il leur permet de perfectionner leurs algorithmes. Ces systèmes d'IA sont programmés pour créer de manière autonome des contenus variés (textes, sons, images).

L'art assisté par ordinateur a pourtant débuté dans les années 1970. Il s'inscrit dans la continuité des genres traditionnels tels que la peinture, la musique ou la vidéo, mais le pinceau, les instruments et la composition sont remplacés par des programmes, les images ou les notes de musique deviennent numériques.

L'art algorithmique est une forme d'art numérique créé à l'aide de codes informatiques et d'algorithmes. Pour la création d'un projet d'art algorithmique, le programmeur définit les paramètres de l'algorithme qui peuvent inclure des éléments tels que la taille, les formes, la luminosité, etc. L'une des premières tentatives a été réalisée par Harold Cohen ayant écrit un programme connu sous le nom d'AARON en 1973 qui a suivi un ensemble de règles par son auteur pour produire une œuvre d'art.

Au cours de la dernière décennie, les algorithmes avancés ont pu apprendre à analyser des masses de données, extraire des schémas et produire de nouvelles œuvres imitant le style et la qualité de celles créées par des artistes humains. Il s'agit désormais de créations réalisées pratiquement intégralement par un algorithme, nourries de toutes sortes de données culturelles. Les nouveaux algorithmes ne sont en effet pas écrits pour suivre un ensemble de règles, ils analysent et apprennent un esthétisme spécifique en scannant des milliers d'images. Par exemple, des programmes tels que DALL-E pour la création d'images prouvent que l'intelligence artificielle peut être formée pour générer des œuvres qui semblent originales.

La capacité des systèmes d'IA à générer des œuvres originales soulève des questions complexes liées à la titularité des droits d'auteur sur ces œuvres. En effet, le droit d'auteur français est dit personnaliste, c'est-à-dire centré sur la personne de l'auteur. Il s'agit d'un ensemble de prérogatives exclusives attribuées au créateur d'une œuvre de l'esprit pour l'utilisation et la diffusion de celle-ci. C'est le droit pour son auteur de jouir des produits issus de la reproduction, de l'exécution ou de la représentation de ses œuvres.

Afin qu'une création soit protégée par le droit d'auteur, elle doit être originale, c'est-à-dire qu'elle doit traduire l'empreinte de la personnalité de son auteur. C'est cette originalité qui déclenche le monopole de l'auteur sur son œuvre ayant pris forme. En principe, l'auteur de l'œuvre doit être une personne physique, soit un humain, pour détenir des droits moraux et patrimoniaux sur sa création. Aussi, si l'intelligence artificielle est utilisée par un individu pour l'assister dans la création d'une œuvre, il aura

la qualité d'auteur d'une œuvre de l'esprit si celle-ci est originale. C'est le cas, par exemple, d'un musicien qui utilise un logiciel de composition musicale assistée par IA pour créer une œuvre musicale.

En l'état actuel du droit, une œuvre créée par une IA de manière autonome n'est pas protégée par le droit d'auteur, car l'essence du droit d'auteur repose sur une œuvre de l'esprit créée par un humain. La plupart des systèmes juridiques actuels, y compris la législation américaine, requièrent qu'une œuvre soit le fruit d'une création intellectuelle humaine pour être protégée par le droit d'auteur. Ce critère exclut de fait les œuvres générées entièrement par une intelligence artificielle, car elles ne sont pas considérées comme ayant un auteur humain. Autrement dit, l'IA n'étant pas dotée de la personnalité juridique, elle ne peut pas être titulaire de droits d'auteur.

Une autre solution ne faisant pas l'unanimité existe pourtant au Royaume-Uni. En effet, l'article 9.3 du *Copyright, Designs and Patents Act* du 15 novembre 1988, attribue les droits à la personne qui a pris les dispositions nécessaires pour créer ladite œuvre au moyen d'un ordinateur. Le concepteur d'une intelligence artificielle est ainsi susceptible de se voir attribuer la paternité de l'œuvre. Il s'agit là d'un auteur indirect, plus éloigné. En pratique, cette solution est difficilement applicable dans la mesure où le concepteur ne peut pas faire respecter ses droits sur un instrument détenu par un tiers, à moins d'attribuer cette qualité à l'utilisateur.

C'est au législateur que reviendra l'initiative de placer ces créations algorithmiques dans le champ de la protection du droit d'auteur. Il n'est pas exclu que le droit évolue à l'avenir pour créer un nouveau régime de protection pour ce type d'œuvres.

83 Comment protéger les œuvres créées par des robots ?

Avec l'intelligence artificielle, les robots commencent à produire des œuvres en témoignent les prouesses artistiques de Sophia et *Ai-Da*, les deux premiers robots capables de dessiner et de peindre. Ces robots sont des robots humanoïdes dont l'apparence ressemble étroitement à un humain, plus précisément il s'agit de « gynoïdes », c'est-à-dire de robots humanoïdes féminins. Le terme « gynoïde » a été utilisé pour la première fois par Isaac Asimov en 1979, comme l'équivalent féminin du mot androïde.

Sophia est un robot humanoïde, développé par Hanson Robotics en 2015, une société basée à Hong Kong. Elle a été acquise par l'Arabie saoudite en 2017 et s'est même vu accorder la citoyenneté saoudienne en 2017, ce qui a suscité un débat sur l'octroi de la personnalité juridique aux robots. Il s'agit d'un robot humanoïde extrêmement réaliste capable d'afficher des expressions faciales semblables à celles des humains et d'interagir avec ceux-ci de façon autonome. Elle ressemble aux robots conscients de la série Westworld, mais n'est pas pour autant dotée d'une intelligence artificielle générale ou d'une intelligence polyvalente comparable à celle des humains. Des réseaux neuronaux profonds permettent en effet au robot de percevoir les émotions d'une personne selon le timbre de sa voix et son expression faciale pour y réagir. Au début, il s'agissait d'un chatbot, mais en novembre 2019 elle fit sa première « prestation d'artiste », en l'occurrence un portrait à main levée d'après nature. En 2021 un autoportrait d'elle, baptisé *Sophia Instantiation*, fut vendu au prix de 688 888 dollars. Disponible sous la forme d'un jeton non fongible (NFT), qui certifie la valeur d'objets numériques et de créations artistiques, le lot comprend un fichier vidéo de 12 secondes d'un portrait initial de l'humanoïde réalisé par l'artiste italien Andrea Bonaceto se transformant en une peinture numérique. C'est donc en collaboration avec un artiste qu'elle a créé des œuvres en se basant sur le travail de son homologue humain.

Ai-Da, développée conjointement par une entreprise de robotique britannique et des informaticiens d'Oxford, s'inscrit dans la même démarche. Son créateur, l'inventeur et galeriste Aidan Meller, la présente comme « *la première artiste robot humanoïde ultra réaliste au monde* ». Créé en 2019, ce robot possède, à ce titre, un bras qui lui permet d'utiliser une palette de couleurs et un pinceau. Il a attiré l'attention du monde quand il avait montré sa capacité à dessiner des portraits en utilisant un crayon et des caméras. Pour réaliser ses œuvres, *Ai-Da* visualise tout d'abord ce qui se trouve devant elle à l'aide des caméras logées dans ses yeux. Une intelligence artificielle interprète ensuite cette image pour envoyer des ordres à son bras robotisé. La « main » du robot dessine alors l'image au crayon, au stylo ou au pinceau. Il est à noter qu'il est impossible de prédire ce qu'elle va réaliser, et qu'aucune de ses œuvres n'est identique à une autre. Les compétences artistiques d'*Ai-Da* sont toutefois encore limitées. Pour réaliser des sculptures, elle doit, par exemple, être assistée par des humains qui modèlent la matière en suivant ce qu'elle a imaginé.

Ai-Da est-elle une véritable artiste ? Plus précisément, un objet créé par un robot, sans intervention humaine, est-il une œuvre de l'esprit protégée par le droit d'auteur ?

Les productions dont il s'agit ici sont à distinguer des créations assistées par ordinateur dans lesquelles l'homme utilise le robot dans le processus créatif. Réduit au rang de chose, le robot peut être considéré comme un outil. Dans cette hypothèse, l'humain ayant été assisté par le robot pourra être reconnu comme auteur et titulaire des droits d'auteur sur l'œuvre créée si les conditions pour leur octroi sont réunies. Mais qu'en est-il si le robot est le seul à intervenir dans le processus créatif ? Dans le cas des robots humanoïdes, il s'agit en effet de créer la simulation d'un être humain, aussi bien dans sa morphologie corporelle que dans ses compétences motrices et ses ressources. Par ailleurs, ces robots travaillent souvent à partir d'images-sources extraites par leurs yeux-caméras de leur environnement, ce qui leur permet de réaliser des portraits selon la nature de personnes présentes. Enfin, ils sont capables de créer des images matérielles (dessins et peintures) créées grâce à leurs bras et mains robotiques.

Pour autant, la production d'un robot autonome est-elle protégeable par le droit d'auteur, sachant que l'objet du droit d'auteur est une œuvre de l'esprit originale ? L'idée la plus communément admise est que l'objet réalisé par le robot ne peut être assimilé à une « œuvre de l'esprit » qui constitue une création intellectuelle ne pouvant se concevoir sans conscience créative. Si l'on admet que la production artistique d'un robot indépendant est une œuvre de l'esprit, encore faut-il que cette création soit « originale ». La satisfaction de cette condition semble en l'état impossible dans la mesure où l'originalité est assimilée à l'empreinte de la personnalité de l'auteur. Or, l'objet créé par un robot ne peut refléter une personnalité qui n'existe pas et qui ne doit pas être reconnue à une machine. La personnalité au sens du droit d'auteur est donc limitée à la personnalité humaine. Mais à supposer que soit admise l'assimilation de l'objet généré par le robot autonome à une œuvre de l'esprit protégée par le droit d'auteur, qui serait alors en droit de revendiquer la qualité d'auteur ?

Dénué de toute personnalité juridique, le robot ne saurait être titulaire de droits. Comment alors protéger cette œuvre ? Une première piste consisterait à considérer l'auteur comme la personne qui prend l'initiative d'utiliser cette technologie, la paramètre et lui fixe des objectifs à réaliser. N'ayant pas directement participé à la conception de l'œuvre, il paraît difficile d'attribuer la qualité d'auteur au concepteur de l'intelligence artificielle dans la mesure où le robot s'est écarté du programme initialement mis en place. Cette qualité devrait sans doute revenir à l'utilisateur du robot qui a donné des ordres à la machine. C'est d'ailleurs la solution retenue par le droit britannique qui considère l'auteur d'une œuvre créée par ordinateur comme étant « *la personne qui prend les dispositions nécessaires à la création de l'œuvre* », c'est-à-dire dans la plupart des cas l'utilisateur de la machine. Toutefois, sa participation au processus créatif semble limitée puisqu'il n'a concrètement que la possibilité d'activer des variables prédéterminées, ce qui ne permet pas d'identifier un véritable apport personnel de l'utilisateur. En quoi le commanditaire d'une œuvre pourrait-il en être l'auteur ?

Une seconde piste pourrait être d'envisager que l'œuvre tombe automatiquement dans le domaine public plutôt que de chercher à tout prix à accorder la titularité des droits à une personne ayant une légitimité plus ou moins acceptable.

En l'état, les robots étant privés de sensibilité et de conscience, l'accès à une protection de leurs « œuvres » par le droit d'auteur paraît tout de même difficile. Dans tous les cas, ce sera au législateur d'apporter des réponses claires à ces questions.

Une proposition de loi visant à encadrer l'intelligence artificielle par le droit d'auteur, en date du 12 septembre 2023, propose, à ce titre, de considérer que « *lorsque l'œuvre est créée par une intelligence artificielle sans intervention humaine directe, les seuls titulaires des droits sont les auteurs ou ayants droit des œuvres qui ont permis de concevoir ladite œuvre artificielle* ». Autrement dit, la titularité de l'œuvre revient au concepteur de l'intelligence artificielle ayant permis la conception de l'œuvre. La solution ne fait pourtant pas l'unanimité et ne semble pas convaincante.

84 Les IA génératives respectent-elles le droit d'auteur ?

L'intelligence artificielle générative est un type de système d'intelligence artificielle capable de générer du texte, des images, des vidéos et même de la musique en réponse à ce qu'il est convenu d'appeler des « prompts ». ChatGPT, Dall-E, Midjourney, Stable Diffusion, Bard, Gemini ou encore Grok sont parmi les plus connues.

Si la protection des produits de l'IA se pose, la question des droits des auteurs à l'origine des contenus est également au cœur des préoccupations des législateurs. En effet, ces IA doivent être en mesure de pouvoir produire un résultat (texte, son ou image) en fonction d'une immensité de données ; le *machine learning* étant la fonction leur permettant d'entraîner leurs algorithmes sur des bases de données et d'apprendre en autonomie, sans l'aide de programmeurs. Ce processus permet au système de s'améliorer en permanence, prenant des décisions sans nécessiter une programmation étape par étape. Afin de constituer les banques de données dans lesquelles ces IA vont puiser, les acteurs du domaine vont généralement se tourner vers la plus grande banque de contenus ouverte existante, à savoir l'internet. Ils utilisent alors des méthodes dites de *web-scraping*, c'est-à-dire des techniques d'extraction automatisée de données *via* des scripts ou des programmes. Il s'agit d'une collecte, très souvent réalisée sans le consentement des titulaires, qui peut capter de manière indifférenciée un grand nombre de données.

Ces IA constituent une source d'inquiétude pour ces derniers qui craignent de ne pas être associés aux rémunérations générées par l'utilisation de leurs œuvres. Cette captation et « exploitation » sont souvent réalisées sans leur autorisation et à leur insu, ce qui soulève de réelles questions juridiques.

Y a-t-il pour autant une exploitation des œuvres ? Deux conceptions s'opposent. Pour certains, l'exploitation d'œuvres de l'esprit pour alimenter une IA constituerait un acte de reproduction soumis à l'autorisation préalable de l'auteur. Pour d'autres, le droit d'auteur pourrait ne pas avoir à s'appliquer. Dans l'affaire *Pelham* du 29 juillet 2019, la Cour de justice de l'Union européenne a rendu, à ce titre, une décision au sujet de l'échantillonnage musical (*sampling*) dans laquelle elle a affirmé que « *lorsqu'un utilisateur, dans l'exercice de la liberté des arts, prélève un échantillon sonore sur un phonogramme, afin de l'utiliser, sous une forme modifiée et non reconnaissable à l'écoute, dans une nouvelle œuvre, il y a lieu de considérer qu'une telle utilisation ne constitue pas une reproduction* ». Autrement dit, un échantillon non reconnaissable ne nécessite pas l'autorisation du titulaire des droits pour son utilisation. La solution est contestable, car les IA se nourrissent d'une quantité considérable de données rendant difficile l'identification des œuvres utilisées par les systèmes d'IA, faute de transparence quant à la nature des œuvres utilisées pour l'entraînement.

Afin de pouvoir nourrir légalement sans autorisation préalable les systèmes d'IA, le législateur européen a prévu, dans la directive de 2019 sur le droit d'auteur, une exception pour la fouille de textes et de données ou « *text and data mining* », c'est-à-dire « *l'analyse automatisée de textes et données sous forme numérique afin d'en dégager des informations, notamment des constantes, des tendances et des corrélations* ». En pratique, tout concepteur d'intelligences artificielles peut ainsi rassembler des contenus,

même protégés par le droit d'auteur, tant qu'ils sont librement accessibles sur l'internet. Cette exception permet, dès lors que l'œuvre a été licitement divulguée, sa reproduction aux fins de la fouille de textes et de données, valant donc pour tous les usages, y compris commerciaux.

L'auteur peut toutefois écarter l'application de cette exception en s'y opposant, ce qui nécessitera d'obtenir à nouveau son autorisation avant de procéder à l'exploration de ses œuvres. Ce dispositif peut s'appliquer au processus d'apprentissage des algorithmes, mais, en pratique, la mise en œuvre de l'opposition reste théorique, car il est difficile de vérifier si les sociétés créatrices d'IA génératives la respectent. Aussi, cette exception au droit d'auteur ne suffit pas à écarter tout risque de sa violation. Par ailleurs, si, pour l'apprentissage des IA, l'utilisation des contenus est autorisée, qu'en est-il de la production ? Le résultat fourni par cette dernière peut toujours constituer une contrefaçon d'une œuvre existante.

De nombreux auteurs, maisons de disques et organes de presse ont engagé des poursuites judiciaires contre des sociétés telles que *OpenAI*, *Microsoft*, *Stability AI*, *Midjourney*, *Meta* et *Anthropic*. Les plaignants soutiennent que ces sociétés ont enfreint les droits d'auteur en exploitant, à leur insu, leur contenu protégé afin d'entraîner leurs modèles d'intelligence artificielle.

L'une des affaires les plus emblématiques est celle du *New York Times* contre *OpenAI* et *Microsoft*. Le journal américain les accuse d'avoir exploité son contenu journalistique sans aucun consentement, afin d'entraîner leurs agents conversationnels ChatGPT et Copilot. Cela a permis à ces sociétés de bénéficier d'un travail journalistique sans compensation appropriée et en s'exonérant du respect des droits de propriété intellectuelle. Le *New York Times* affirme dans sa plainte que *ChatGPT* a reproduit intégralement certains des articles de sa base, dépassant ce qui est normalement admis sous la doctrine de l'usage loyal ou *fair use* qui constitue un principe juridique du droit d'auteur anglo-saxon autorisant certaines utilisations sans consentement préalable, pour peu que l'œuvre soit transformée de façon substantielle. Les premières décisions rendues, dont notamment l'ordonnance de la Cour fédérale de district du district nord de l'État de Californie du 30 octobre 2023 (District Court, N.D. California, 30 oct. 2023, *Andersen v. Stability AI Ltd.*, n° 3 : 23-cv-00201), semblent opter pour le *statu quo* en faveur des développeurs d'IA génératives.

L'une des principales inquiétudes liées à cette IA générative réside dans son manque de transparence. En effet, *Meta* ou *OpenAI* ne dévoilent pas les textes spécifiques ou les sources utilisées pour entraîner leurs algorithmes. Le secteur culturel en France exprime les mêmes inquiétudes. Le 17 novembre 2023, 80 organisations des secteurs de l'audiovisuel, de l'édition, de la musique, des arts visuels et de la photographie ont soumis au gouvernement français un texte demandant la transparence des données d'entraînement et des contenus générés par les modèles d'intelligence générative, la considérant comme un impératif absolu pour le développement d'une IA éthique.

Leurs revendications ont été, en partie, satisfaites par le règlement européen sur l'IA ou l'AI Act du 13 juin 2024 qui prévoit plusieurs niveaux d'obligations, allant de mesures de transparence et de documentation minimales. Le fournisseur du modèle doit ainsi mettre

en place une politique assurant le respect du droit d'auteur, notamment durant la fouille de textes et de données. En outre, le fournisseur devra faire preuve de transparence en publiant un résumé suffisamment détaillé du contenu ayant permis d'entraîner le modèle.

Malgré cette avancée, la question de la transparence des systèmes d'IA demeurera l'une des préoccupations majeures des auteurs. Il appartiendra aux fournisseurs de justifier de la qualité et de la traçabilité des données d'entraînement de leur système d'IA et, par voie de conséquence, de la transparence de leurs outils. Si cela est possible, une telle transparence serait-elle pour autant suffisante pour promouvoir le développement d'une IA novatrice tout en préservant les droits d'auteur et les autres droits de propriété intellectuelle ? La question mérite d'être posée car la possibilité de divulguer les données d'entraînement demeure incertaine. La seule obligation de mettre à disposition du public un résumé des œuvres protégées, sur lesquelles une IA s'est entraînée, ne suffira pas en effet à garantir une juste compensation pour les auteurs.

85 Que se passe-t-il lorsque l'intelligence artificielle s'inspire ou intègre une œuvre préexistante ?

Les systèmes d'IA fonctionnent grâce à d'importantes bases de données avant de faire usage de ces données pour la composition de nouveaux contenus.

La présence d'œuvres préexistantes dans une base de données a été prévue par l'exception de « *fouilles de textes et de données* » selon laquelle les titulaires de droits d'auteur ne peuvent s'opposer à un tel procédé « *quelle que soit la finalité de la fouille, sauf si l'auteur s'y est opposé de manière appropriée* ». Ainsi, il est donc possible pour les concepteurs d'IA de constituer des bases de données d'œuvres préexistantes, à la seule condition que les ayants droit de celles-ci n'aient pas fait opposition. Le maintien d'une œuvre dans une base après une opposition pourrait alors être considéré comme une contrefaçon.

Il existe une situation fréquente dans le domaine de la création, connu sous le terme d'œuvre dérivée ou d'œuvre composite à travers laquelle un auteur second crée une œuvre en utilisant une ou plusieurs œuvres premières. L'œuvre composite ou dérivée est définie par l'article L. 113-2, alinéa 2, du Code de la propriété intellectuelle comme étant une « *œuvre nouvelle à laquelle est incorporée une œuvre préexistante sans la collaboration de l'auteur de cette dernière* ». Cela signifie qu'il y a alors une pluralité d'auteurs se succédant dans le temps (l'auteur de l'œuvre seconde crée nécessairement après l'auteur de l'œuvre première). L'article L. 113-4 du Code de la propriété intellectuelle tranche la question de la titularité sur l'œuvre composite de la manière suivante : elle est « *la propriété de l'auteur qui l'a réalisée, sous réserve des droits de l'auteur de l'œuvre préexistante* ». L'auteur de l'œuvre seconde doit ainsi veiller à respecter les droits portant sur l'œuvre première. Aussi, si les éléments utilisés pour créer cette œuvre ne sont pas dans le domaine public, une autorisation auprès du titulaire de droits de l'œuvre première est requise.

S'agissant des œuvres générées par l'IA, elles reprennent des caractéristiques et des composantes issues de leur base de données, et notamment d'œuvres préexistantes. Ces créations ainsi générées peuvent être *a priori* qualifiées d'œuvres composites. La preuve de la reproduction ou l'imitation par l'IA d'une œuvre préexistante protégée n'est cependant pas aisée, car celle-ci est intégrée à un ensemble complexe composé de nombreuses œuvres antérieures. Cela suppose en effet que l'on puisse reconnaître l'œuvre première dans l'œuvre seconde, ce qui peut poser quelques difficultés en raison d'une dilution des sources utilisées, rendant ainsi la preuve de la titularité des droits antérieures, potentiellement exploités sans autorisation, nettement plus complexe à établir.

De ce fait, la preuve de la reproduction ou l'imitation par l'IA d'une œuvre préexistante est une tâche ardue dans la mesure où elle se retrouve incorporée à un contenu nouveau, complexe et hétérogène. L'intelligence artificielle peut ainsi développer une œuvre si singulière à partir d'une œuvre préexistante, que l'identification de cette dernière sera beaucoup plus difficile à démontrer. L'imitation du style ne saurait d'ailleurs servir de fondement pour établir la preuve de la reproduction.

En effet, selon une formule bien connue « *les idées sont de libre parcours* », ce qui signifie que les idées en tant que telles doivent rester à la disposition de tous pour ne pas entraver le processus créatif. Il en résulte qu'il n'est pas envisageable d'obtenir une protection sur un style. La question mérite d'être posée, car, avec le projet *The Next Rembrandt*, une intelligence artificielle a pu créer une peinture inédite reproduisant fidèlement la technique et le style du peintre Flamand après avoir analysé ses œuvres.

L'exercice d'une action en contrefaçon par un auteur dont les droits auraient été violés semble ainsi particulièrement complexe. Il en résulte qu'une image produite par ChatGPT, par exemple, est susceptible de porter atteinte aux droits d'auteur d'un tiers, sans que cette atteinte puisse être aisément constatée en raison du caractère composite de l'image générée. Dans ce cas, l'action en concurrence déloyale et en parasitisme semble plus adaptée à la sanction des actes déloyaux dont pourraient se rendre coupables les utilisateurs de ChatGPT ou d'Image Creator. Les auteurs pourraient également invoquer le droit moral, et notamment le droit à la paternité si l'auteur n'est pas mentionné lors de l'exploitation de l'œuvre composite, ou le droit au respect de l'œuvre si l'œuvre composite dénature l'œuvre préexistante. Ce droit moral de l'auteur est, à ce titre, prévu dans Code de la propriété intellectuelle à l'article L. 121-1 qui prévoit que : « *L'auteur jouit du droit au respect de son nom, de sa qualité et de son œuvre* ».

En définitive, la protection par le droit d'auteur des productions d'une intelligence artificielle générative soulève des questions inédites. L'absence de protection effective des œuvres servant à entraîner les IA pourrait engendrer une captation de valeur de la propriété intellectuelle par les géants de la Tech. Cette question est ainsi l'une des plus urgentes à résoudre, car un grand nombre d'auteurs sont actuellement démunis face à l'exploitation de leurs œuvres par des IA génératives.

86 Une création issue d'une intelligence artificielle peut-elle être protégée par le *Copyright Act* ?

Le *copyright* est le système de protection de la propriété intellectuelle dans les pays dits de *Common law*, c'est-à-dire les pays anglo-saxons (États-Unis, le Royaume-Uni, l'Australie, etc.). On oppose souvent le droit d'auteur français au *copyright* anglo-saxon. Le droit d'auteur, a en effet mis l'auteur sur un piédestal en lui octroyant des droits exorbitants, notamment par la reconnaissance de droits moraux, alors que le *copyright*, surtout le *copyright* américain, assimile le droit de l'auteur à un monopole accordé pour une période limitée afin d'encourager le développement des arts et de la science. Littéralement, « *copyright* » signifie « *droit de copier* » : il s'agit bien d'un droit d'exploitant, lié à l'œuvre elle-même.

Le droit d'auteur et le *copyright* ont pourtant le même objet, à savoir sanctionner la reproduction d'œuvres faite au mépris des droits de leurs auteurs, et l'opposition entre ces deux systèmes juridiques est en réalité à nuancer. Le *Copyright Act*, qui constitue la loi américaine sur le droit d'auteur, confère aux titulaires des droits d'auteur des droits exclusifs similaires à ceux du droit d'auteur français. Comme le droit d'auteur français, le *copyright* est un droit de propriété intellectuelle qui protège les œuvres originales créées par un auteur. Il peut ainsi protéger les œuvres d'art, les écrits originaux, les films, les photographies, les vidéos, la composition musicale, etc. De même, le *copyright* s'applique automatiquement lors de la création d'une œuvre originale. Il est toutefois recommandé d'effectuer une demande d'enregistrement de l'œuvre auprès du *US Copyright Office* pour faire respecter l'exclusivité du droit d'un auteur, et obtenir des dommages-intérêts en cas de litige.

Après avoir reçu plusieurs demandes d'enregistrement pour des créations générées par IA, le Copyright Office a dû publier un document afin de clarifier ses pratiques. En effet, les lignes directrices du Copyright Office du 10 mars 2023 « *en matière d'examen et d'enregistrement des œuvres contenant du matériel généré par l'utilisation de la technologie de l'intelligence artificielle* » permettent de répondre aux questions relatives à la protection des œuvres générées ou créées au moyen d'outils d'IA.

Dans ce document, le Copyright Office commence par rappeler le principe selon lequel les auteurs sont des êtres humains, principe prévu à la fois par la loi américaine et la jurisprudence. La législation américaine exclut en effet la protection par le droit d'auteur des œuvres qui n'ont pas été générées par une main humaine dont le principe a été rappelé lors de l'affaire « NARUTO », par laquelle l'association Peta tentait d'obtenir la qualité d'auteur pour un singe ayant réalisé des clichés photographiques.

Le Copyright Office précise ensuite la méthodologie à suivre pour déterminer le caractère protégeable ou non d'une œuvre créée à l'aide d'un système d'IA. L'objectif est de déterminer « *si l'œuvre est fondamentalement une œuvre créée par l'homme, l'ordinateur n'étant qu'un instrument d'assistance, ou si les éléments traditionnels de la paternité dans l'œuvre (expression littéraire, artistique ou musicale ou éléments de*

sélection, d'arrangement, etc.) ont en fait été conçus et exécutés non pas par l'homme, mais par une machine ». Le Copyright Office précise ainsi qu'un logiciel ou un algorithme ne peut être un auteur et qu'une création générée par IA n'est donc pas protégeable.

Lorsqu'il s'agit d'œuvres contenant du matériel généré par l'IA, l'Office précise qu'il examinera si les contributions de l'IA sont le résultat d'une « *reproduction mécanique* » ou plutôt de la « *conception mentale originale de l'auteur, à laquelle l'auteur a donné une forme visible* ». La réponse dépendra alors des circonstances, en particulier du fonctionnement de l'outil d'IA et de la manière dont il a été utilisé pour créer l'œuvre finale.

Ce faisant, une simple instruction donnée à l'IA, un prompt, ne serait pas suffisante puisque c'est l'outil qui choisit la forme à donner à la création. Lorsqu'une IA est en revanche utilisée pour assister la création humaine, l'œuvre devient alors protégeable. Par exemple, un être humain peut sélectionner ou arranger une production générée par une IA d'une manière suffisamment créative pour que l'œuvre qui en résulte constitue dans son ensemble une œuvre originale. Le Copyright Office a d'ailleurs accepté l'enregistrement d'une bande dessinée dont les images étaient produites par une IA, mais dont le texte et le scénario avaient été créés par un humain (*U.S. CopyrightOffice, Cancellation Decision re : Zarya of the Dawn*, 21 févr. 2023).

Par cette position, le Copyright Office adopte une solution proche du droit français pour lequel l'auteur doit également être une personne physique. En droit français comme en droit américain, les œuvres créées par des IA ne sont pas protégées. Seuls les éléments originaux d'une œuvre, nécessairement créés par l'homme, sont protégeables. L'application de cette solution est toutefois tributaire des circonstances de chaque espèce, et notamment des conditions de fonctionnement de chaque système d'IA.

L'absence de protection par le droit d'auteur semble certes avantageuse pour les utilisateurs de ces créations qui peuvent diffuser un texte ou une image sans rémunérer un auteur ni devoir respecter ses droits, mais elle a un impact sur l'économie de la création. En effet, l'utilisation d'une IA prive les artistes de revenus en les remplaçant, et les développeurs d'IA ne les rémunèrent pas pour autant lorsqu'ils utilisent leurs œuvres.

La question de la transparence des bases de données d'entraînement des IA génératives est revenue sur le devant de la scène avec une proposition de loi intitulée « *Generative AI Copyright Disclosure Act of 2024* », introduite le 9 avril, devant le Congrès américain. Ce projet de loi exigerait que les entreprises divulguent les données d'entraînement de leurs modèles d'IA générative. Il vise à accroître la transparence dans l'utilisation des données. Il prévoit que toute personne qui crée ou modifie un ensemble de données destinées à entraîner une IA générative devra adresser au Copyright Office le détail des données constituant la base d'entraînement de ce modèle, préalablement à la mise sur le marché de ce dernier.

Si cette proposition est adoptée, elle imposera ainsi des obligations de divulgation des données d'entraînement des modèles d'IA générative pour éviter les procès dans lesquels la question de la preuve de l'entraînement licite de ces modèles ne cesse de se poser.

Cette proposition de loi américaine semble clairement inspirée par les dispositions du règlement européen sur l'intelligence artificielle. Se dessine ainsi une convergence des solutions pour traiter le problème des données d'entraînement.

87 Les inventions réalisées par une IA sont-elles brevetables ?

Le brevet d'invention est un titre de propriété industrielle qui porte sur une nouvelle solution technique à un problème technique. C'est donc le résultat qui importe et non le processus créatif en tant que tel. On dit souvent que les conditions de brevetabilité sont « objectives » en ce que, contrairement au droit d'auteur, elles ne nécessitent pas l'empreinte de la personnalité de l'auteur, notion assez subjective.

Ce titre confère à son titulaire un droit exclusif d'exploitation en échange d'une publication détaillée de l'invention et de ses moyens. Pour que le brevet puisse être enregistré, il faut qu'il porte sur une nouvelle invention brevetable. L'invention doit en outre remplir trois conditions autonomes afin de pouvoir prétendre à un brevet : la nouveauté par rapport à l'état de la technique ; l'activité inventive qui suppose que l'invention ne découle pas de manière évidente de l'état de la technique pour l'homme du métier ; l'application industrielle qui signifie que l'invention doit pouvoir être fabriquée ou utilisée dans une industrie.

La protection du système d'IA peut ainsi passer par le brevet. L'algorithme comme logiciel n'est normalement pas protégeable par le droit des brevets qui exclut les méthodes mathématiques et les programmes d'ordinateur. Il peut être considéré comme un cheminement intellectuel en principe exclu du domaine de la propriété puisque « les idées sont de libre parcours ». L'exclusion des algorithmes se justifie ainsi par la non-protection des idées et des théories. Cette exclusion n'écarte pas totalement les éléments concernés de la protection des brevets dans la mesure où ils sont intégrés à un produit complexe composé d'éléments physiques (processeur, ordinateur) sur lequel l'algorithme est exécuté.

Autrement dit, il est possible d'obtenir un brevet sur un ensemble comportant des caractéristiques abstraites et des entités physiques où l'algorithme est exécuté. Dans la demande de brevet, il suffira d'indiquer que la demande vise une invention qui fait intervenir l'intelligence artificielle, au lieu de viser l'intelligence artificielle elle-même. L'objet de la protection sera alors différent du simple « programme d'ordinateur » ou de la « méthode mathématique ». L'invention mise en œuvre par ordinateur doit également remplir les conditions de brevetabilité que sont la nouveauté, l'activité inventive et la possibilité d'une application industrielle.

La protection de l'outil d'intelligence artificielle par le brevet semble ainsi possible, plus délicate est en revanche la question de la brevetabilité des inventions de l'IA. Pour des domaines déterminés, les machines peuvent être plus rapides que les hommes et peuvent produire un résultat nouveau, notamment dans les domaines de la recherche pharmaceutique. On peut citer, par exemple, le cas d'Eve, une intelligence artificielle développée par l'université de Manchester, qui a découvert à un antibactérien, largement utilisé dans la vie courante, une seconde application brevetable contre le paludisme.

Sur le principe, le droit des brevets n'est pas totalement opposé à la brevetabilité de telles productions de l'intelligence artificielle dès lors que les conditions sont réunies. La nouveauté sera le plus souvent une condition remplie, de même que l'activité inventive, soit la non-évidence pour l'homme du métier. Cette condition est en effet remplie si pour

un homme du métier, la création technique ne découle pas de manière évidente de la technique. Une solution technique ayant été obtenue grâce à un processus automatique peut toutefois ne pas être évidente pour l'homme du métier, mais cela ne conduit pas pour autant à l'absence d'activité inventive. Enfin, l'application industrielle pourra dans de nombreux cas être satisfaite dans des domaines variés (médicaments, inventions génétiques, etc.). Dès lors que ces trois conditions sont réunies, l'invention réalisée par l'intelligence artificielle devrait pouvoir faire l'objet d'un brevet d'invention.

Si les inventions issues des travaux d'une IA peuvent réunir les conditions de fond pour l'obtention d'un brevet, le fait qu'elles n'aient pas été réalisées par une personne physique, soit un humain, les rend inéligibles à la protection par le droit des brevets. La demande de brevet suppose en effet la désignation de l'inventeur. Or, il n'y a point d'inventeur au sens du droit des brevets en présence d'un système intelligent. L'admission de l'IA à titre d'inventeur poserait la question de savoir à qui reconnaître la propriété des brevets sur les réalisations des IA. Même si l'empreinte de la personnalité de l'inventeur est moins intense que celle exigée par le droit d'auteur, le droit des brevets a été aussi conçu au regard de la personne physique, ce qui implique une contribution humaine. Alors que le droit des brevets pouvait sembler plus propice à une intégration des créations de l'IA contrairement au droit d'auteur, la protection par la voie des brevets n'est finalement pas si adaptée. Le personnage de l'inventeur reste malgré tout étroitement lié à l'invention brevetée. Il convient que le législateur s'empare de cette question afin de mettre en place un régime juridique favorisant l'innovation lorsque l'inventeur n'est pas une personne.

88 L'IA peut-elle être désignée comme inventeur dans une demande de brevet ?

La question de la paternité des inventions conçues sans intervention humaine a fait l'objet de nombreux débats au cours de ces dernières années en raison de la célèbre intelligence artificielle baptisée DABUS (un acronyme pour « *Device for the Autonomous Bootstrapping of Unified Sentience* »). DABUS est un système d'intelligence artificielle de type connexionniste qui a notamment procédé à deux inventions : l'une relative à un récipient de boisson, la seconde concernant une balise lumineuse clignotante dans des missions de recherche et de sauvetage. L'équipe *Artificial Inventor Project* (AIP) qui pilotait le projet avait notamment précisé que la machine n'avait reçu qu'une formation sur les connaissances générales dans le domaine et qu'elle a conçu de manière indépendante l'invention.

Des demandes de brevet sur les productions de DABUS ont été présentées devant plusieurs offices de propriété intellectuelle dans le monde. La particularité de ces demandes de brevet était qu'avait été désigné comme inventeur le système d'intelligence artificielle DABUS. Du fait de cette mention, les offices de plusieurs pays ont été contraints de se prononcer sur la possibilité de reconnaître une IA comme inventeur et de la mentionner, à ce titre, dans la demande de brevet. Plus précisément, les demandes de brevet précisaient que le déposant était Stephen Thaler et que l'inventeur était DABUS. Pour rejeter la demande de brevet, les différents offices ont plus ou moins invoqué le même argument.

L'Office européen des brevets (OEB) a ainsi rejeté deux demandes de brevet, le 27 janvier 2020, et sa position fut clairement défavorable à l'admission de la qualité d'inventeur pour l'IA DABUS au motif qu'aucun être humain n'avait été désigné comme inventeur. Au Royaume-Uni, l'Office britannique de la propriété intellectuelle (UKIPO) a, quant à lui, adopté la même solution concernant ces demandes de brevet. Il a admis que DABUS avait créé les inventions, mais il a déclaré qu'une machine ne peut pas être désignée comme inventeur car il ne s'agit pas d'une personne physique.

Aux États-Unis, l'USPTO (*United States Patent and Trademark Office*) est également parvenu à une solution similaire à celles de l'OEB et de l'UKIPO dans une décision du 6 août 2020. En se fondant sur la loi américaine en droit des brevets qui se réfère à l'inventeur pris en tant qu'individu, l'office américain a retenu que l'inventeur ne peut être qu'une personne physique. En outre, le déposant doit être capable de prêter serment lors du dépôt, ce qu'une IA ne peut faire. Pour écarter cet argument, le déposant a indiqué en vain que ces textes de loi encadrant le dépôt de brevet ont été rédigés au début des années 1960 et ne pouvaient donc pas envisager le cas de l'IA. Par ailleurs, il précise que le terme « inventeur » n'est pas clairement défini dans le droit américain ni anglais, et peut donc faire l'objet d'une interprétation tant que n'est pas expressément exclue la reconnaissance d'une IA en tant qu'inventeur.

La même solution a été adoptée à Taïwan. Dans une décision du 19 août 2021, le tribunal du commerce et de la propriété intellectuelle a confirmé la décision de l'Office taïwanais de la propriété intellectuelle de rejet d'une demande de brevet dans laquelle l'inventeur

désigné était DABUS, au motif qu'un inventeur devait être un humain soit, en langage juridique, une personne physique. Le tribunal a également rappelé l'intention du législateur, qui, par le système des brevets, est d'encourager et de protéger les fruits de l'activité mentale humaine. Or, une intelligence artificielle n'est pas un être humain mais une « chose » qui ne peut détenir un droit mais seulement en être l'objet.

Le refus de reconnaître la qualité d'inventeur à une intelligence artificielle semble être une solution suivie par les principaux offices de propriété intellectuelle, mais elle ne fait pas l'unanimité. Par une décision du 28 juillet 2021, l'office sud-africain des brevets a délivré pour la première fois au monde un brevet sur une invention réalisée par une intelligence artificielle, le brevet désignant Dabus comme étant l'inventeur. Cette solution a été également suivie par la *Federal Court* d'Australie, dans une décision du 30 juillet 2021, par laquelle elle s'est prononcée en faveur d'une telle position après le rejet par l'office australien des demandes de brevets en validant les demandes de brevets désignant l'IA DABUS comme inventeur. Le juge australien précise que rien en droit des brevets ne conduit à interpréter la loi comme excluant les inventeurs non-humains, à l'inverse du droit d'auteur qui requiert qu'un auteur soit une personne.

Il admet ainsi qu'un inventeur puisse être un système d'intelligence artificielle, mais il ne pourrait pas être le propriétaire, l'utilisateur ou le titulaire du brevet. Il indique aussi que le fait de refuser le dépôt d'un brevet au prétexte que l'inventeur n'est pas doté de personnalité juridique serait contraire à l'esprit de la loi australienne (*Patents Regulations* 1991) qui a pour objectif de favoriser l'innovation technologique et sa diffusion auprès du public.

Ces décisions ont connu un important retentissement dans le monde même si leur portée est à relativiser. Il s'agit en effet avant tout de décisions symboliques, car la qualité d'inventeur est reconnue à une IA pour la première fois dans le monde conférant ainsi à un système intelligent les mêmes qualités inventives qu'un humain.

Le mérite de l'affaire DABUS est d'avoir contribué à confirmer l'état du droit sur cette question par certaines instances les plus influentes en matière de propriété intellectuelle, mais elle illustre l'importance cruciale de définir qui peut prétendre à la propriété intellectuelle lorsque des IA sont impliquées. Il est vrai que l'IA bouscule le droit de la propriété intellectuelle ayant été conçu à partir de la personne physique en considérant que la créativité et l'innovation sont le fait de l'être humain.

En raison du progrès de l'intelligence artificielle, il serait utile de clarifier cette notion d'inventeur, pour indiquer qu'il s'agit d'une personne physique, ou à défaut pour exclure de cette qualité les inventeurs artificiels relevant de la catégorie des choses. Une clarification semble nécessaire, car la frontière entre les personnes et les choses tend à se brouiller.

89 Comment protéger techniquement les œuvres de l'esprit contre les utilisations abusives ?

Les fournisseurs de systèmes d'IA se voient régulièrement reprocher des violations des droits de propriété intellectuelle et des manquements en matière de données à caractère personnel, en raison de leurs pratiques de collecte de données en ligne (*web scraping*). Pour protéger les artistes, dont les œuvres sont en grande majorité utilisées par les IA sans autorisation ni rémunération de l'auteur ou de ses ayants droit, des outils ont été développés afin de leur permettre de savoir si leurs œuvres protégées ont été utilisées dans le cadre de l'entraînement des intelligences artificielles. Pour faire face à cette exploitation, des artistes piègent volontairement leurs créations pour les rendre inutilisables grâce à des outils spécifiques.

Une équipe de chercheurs de l'Université de Chicago a développé, par exemple, un logiciel baptisé « Glaze » permettant d'insérer dans une œuvre des pixels invisibles à l'œil nu qui rendent l'image floue et brouillée dès qu'une IA tente de l'utiliser. « Glaze », pouvant être traduit par vernis, lustre, glaçage en français, est un logiciel permettant ainsi de générer une version masquée d'une œuvre originale (dessin, peinture, etc.) afin de la protéger contre les utilisations faites par l'intelligence artificielle. Pour parvenir à ce résultat, le logiciel s'appuie sur des techniques d'apprentissage automatique avancées. Il scrute l'œuvre d'art et évalue les meilleures façons de la transformer sans altérer son apparence visuelle pour les humains.

L'auteur peut ensuite publier en ligne l'image modifiée. Si un modèle d'IA générative tente de s'en servir, les données informatiques ajoutées empêcheront la machine d'analyser correctement le style et de le reproduire. Le logiciel Glaze constitue ainsi une autre façon de lutter contre les exploitations abusives des œuvres par les IA dans le domaine artistique. Il s'agit concrètement de recouvrir numériquement l'œuvre d'un artiste en modifiant certains éléments quasiment imperceptibles à l'œil nu afin de tromper une IA et de l'empêcher de copier son style.

Si cette protection technique semble prometteuse, elle restera cependant temporaire car les IA apprendront probablement à décoder les images protégées par les artistes. Pour autant, l'idée de « piéger » l'œuvre se développe et d'autres outils apparaissent. La même équipe de chercheurs de l'Université de Chicago a, par exemple, développé un outil complémentaire de Glaze. Plus offensif que ce dernier, Nightshade ne se limite pas à se défendre contre une imitation de style, il est conçu pour transformer les images en échantillons « empoisonnés », inadaptés à la formation de modèles d'IA. Il vise à faire dérailler l'algorithme, qui proposera ensuite, par exemple, une image de chat alors qu'un chien a été demandé. Il s'agit de manipuler les données d'entraînement pour introduire un comportement inattendu dans le modèle au moment de l'entraînement. Lorsque ces images altérées sont en effet intégrées dans des bases de données utilisées par des IA, elles perturbent leur capacité à interpréter correctement les données visuelles. En exploitant cette vulnérabilité, il est possible d'introduire des résultats de mauvaise classification qui conduiront à la production de contenus erronés ou déformés lorsqu'une IA tentera de créer de nouvelles images basées sur ces données corrompues.

D’autres outils permettent aux auteurs de savoir si certaines images protégées ont été utilisées dans le cadre de l’entraînement des intelligences artificielles. C’est, par exemple, le cas du site web « *Have I Been Trained?* », créé par la start-up Spawning, qui permet grâce à une simple requête aux artistes de vérifier si leurs œuvres ont été utilisées dans le cadre des données d’entraînement pour les IA génératives. La même start-up a aussi mis au point Kudurru, logiciel qui détecte les tentatives de collecte massive sur des plateformes d’images. L’artiste peut alors décider soit de bloquer l’accès à son œuvre, soit d’envoyer une autre image que celle demandée afin d’empoisonner le modèle d’IA en affectant sa fiabilité.

Dans le même esprit de Glaze, AntiFake, développé par l’université de Washington à St. Louis (Missouri), permet, quant à lui, de piéger un fichier audio afin d’empêcher les IA de réutiliser une voix pour un *deepfake*. Ce logiciel enrichit un fichier son de bruits supplémentaires, imperceptibles à l’oreille humaine, rendant impossible l’imitation crédible d’une voix humaine.

La multiplication de ces outils répond à l’inquiétude des auteurs face à l’exploitation abusive de leurs œuvres et constitue une réponse afin de forcer les concepteurs d’IA à respecter leurs droits d’auteur. C’est également un moyen de montrer que les IA peuvent aussi être utilisées pour en contrer d’autres, il ne serait donc pas étonnant que de tels outils se multiplient à l’avenir.

90 Quels sont les problèmes posés par la création de *deepfakes* ?

Le terme *deepfake* est une combinaison de *deep* qui fait référence à l'apprentissage profond (*deep learning*) et de *fake* qui signifie faux. Cette technologie met à profit l'intelligence artificielle pour animer un visage ou retranscrire une voix à partir de simples images ou d'échantillons audio. Les *deepfakes* ont été popularisés en 2017 grâce à des logiciels tels que *DeepFaceLab* ou *FakeApp*, ayant permis de créer des *deepfakes* de manière relativement simple. De nombreuses autres applications se sont développées par la suite rendant la création de *deepfakes* encore plus accessible. Elles ont permis de créer, par exemple, une fausse image du pape François en doudoune blanche ou de Donald Trump en état d'arrestation.

Il s'agit essentiellement de superposer des traits humains sur le corps d'une autre personne ou de manipuler les sons pour générer une expérience humaine réaliste. L'algorithme apprend ainsi à analyser les expressions faciales et les traits du visage de la personne cible grâce à l'apprentissage profond. Grâce à cet apprentissage, il est possible de remplacer une vidéo dans laquelle une autre personne apparaît par une vidéo de la personne cible. Si autrefois, ces images ou ces vidéos étaient de faible qualité et facilement identifiées, il devient possible grâce au progrès de l'IA de créer des *deepfakes* qui pourraient être confondus avec des vidéos de la personne réelle. Il devient surtout possible de faire faire ou faire dire n'importe quoi à n'importe qui à des fins de divertissement ou de malveillance.

Les *deepfakes* se répandent sur la toile à des fins pornographiques, parodiques, ou encore de désinformation. Dans ces cas, la victime d'un *deepfake* a la possibilité d'utiliser plusieurs fondements pour faire valoir ses droits. Elle peut invoquer tout d'abord le non-respect de sa vie privée dans la mesure où le *deepfake* est susceptible de porter atteinte au droit à l'image et à la voix. N'ayant pas autorisé le *deepfake* et étant identifiable, la victime peut agir sur le fondement de l'article 9 du Code civil qui consacre le droit au respect de la vie privée afin d'obtenir des dommages-intérêts.

Elle peut également se prévaloir du droit d'auteur si le *deepfake* a été réalisé en violation de ses droits. En effet, les concepteurs d'un *deepfake* se doivent d'obtenir l'accord des titulaires de droits s'ils utilisent des œuvres protégées par le droit d'auteur ou les droits voisins. Aussi, si le concepteur d'un *deepfake* superpose une image avec une musique dont il ne possède pas les droits, il réalise alors un acte de contrefaçon. Les concepteurs de *deepfakes* pourraient toutefois se prévaloir de certaines exceptions, soumises à des conditions, au droit d'auteur prévues dans le Code de la propriété intellectuelle, telles que les exceptions de courte citation ou de parodie.

Enfin, la victime peut aussi invoquer la protection de ses données à caractère personnel. En effet, la voix et l'image bénéficient de la protection prévue par RGPD qui impose des obligations spécifiques au responsable du traitement, telles que le recueil du consentement de la personne concernée.

Si les utilisations répréhensibles semblent être l'aspect le plus inquiétant de cette technologie, elle peut aussi être utilisée pour des usages créatifs. Par exemple, les *deepfakes* peuvent être utilisés pour créer des scènes de films ou de séries télévisées de

manière plus réaliste et moins coûteuse. Par exemple, ils ont été utilisés dans le film *The Irishman* pour rajeunir le visage de certains acteurs afin de les faire paraître plus jeunes. On peut aussi citer le vidéoclip du rappeur américain Kendrick Lamar, *The Heart Part 5*, dans lequel l'artiste arbore successivement les visages de différentes personnalités afro-américaines disparues ou controversées.

Bien que l'utilisation des *deepfakes* puisse avoir des aspects positifs, il est important de demeurer vigilant face à leur utilisation abusive. À ce titre, la loi du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique, dite loi « SREN », prévoit des dispositions afin de lutter contre les *deepfakes* préjudiciables. En effet, la loi SREN punit d'un an d'emprisonnement et de 15 000 euros d'amende le fait de porter à la connaissance du public ou d'un tiers le « *contenu visuel ou sonore généré par un traitement algorithmique et représentant l'image ou les paroles d'une personne* » sans l'autorisation de la personne concernée et qui n'apparaît pas à l'évidence comme un contenu généré par une IA ou s'il n'en est pas expressément fait mention (art. 226-8 du Code pénal). En outre, la loi SREN a créé un nouveau délit dans le Code pénal, à l'article 226-8-1, dans le but de sanctionner la publication des *deepfakes* à caractère sexuel. Est ainsi puni de 60 000 € d'amende le fait de porter à la connaissance du public ou d'un tiers, par quelque voie que ce soit, « *un contenu visuel ou sonore à caractère sexuel généré par un traitement algorithmique et reproduisant l'image ou les paroles d'une personne, sans son consentement* ».

Si ces dispositions constituent une avancée au regard de l'objectif visé, il est à craindre qu'elles soient de portée limitée, car elles visent essentiellement les *deepfakes* qui n'apparaissent pas à l'évidence comme des contenus générés à l'aide d'un algorithme, ou, le cas échéant, si la mention de l'utilisation d'un tel procédé n'est pas expressément mentionnée. L'information du public sur le caractère factice du contenu permet ainsi de contourner le texte pour échapper aux sanctions. Par ailleurs, le texte ne s'applique pas si l'usage de l'IA est évident pour le public, créant ainsi une incertitude juridique génératrice de contentieux. Ces dispositions devront sans doute être complétées pour prendre en compte tous les aspects de cette pratique.

X.
LA RÉGLEMENTATION DE L'INTELLIGENCE ARTIFICIELLE

91 Existe-t-il une réglementation internationale de l'IA ?

L'intelligence artificielle n'est pas une zone de non-droit. Le droit existant lui est déjà applicable. Cependant, il peut sembler utile d'adopter une réglementation internationale spécifique afin d'encadrer les conséquences transfrontières du développement de l'intelligence artificielle. Le cadre réglementaire international de l'IA est influencé par l'orientation stratégique respective des grands acteurs internationaux, notamment la Chine, les États-Unis et l'Union européenne qui voient dans l'IA un champ d'action géopolitique de première importance. Il existe toutefois de nombreuses organisations qui discutent de l'application des règles existantes, ou de l'adoption de nouvelles règles visant à réguler l'IA dans son ensemble.

Parmi celles-ci, nous pouvons citer le Conseil de l'Europe qui est une organisation internationale fondée en 1949 chargée de préserver les droits humains, la démocratie et l'État de droit en Europe. Il compte actuellement 46 États membres (dont 27 sont aussi membres de l'Union européenne) et a son siège à Strasbourg.

Ayant pris conscience de la nécessité pour les États de régir le développement et l'utilisation des systèmes d'IA, le Conseil d'État a créé en 2019, le Comité *ad hoc* sur l'intelligence artificielle du Conseil de l'Europe (CAHAI) chargé d'examiner la possibilité d'adopter un cadre juridique pour le développement, la conception et l'application de l'IA. Ce dernier a décidé, fin 2021, que la meilleure solution serait la négociation d'une convention qui devrait être juridiquement contraignante en matière de développement, de conception et d'utilisation de l'IA. Il constate en effet que les diverses questions juridiques soulevées par les systèmes d'IA ne sont pas spécifiques aux États membres du Conseil de l'Europe, mais transnationales par nature, en raison de l'implication de grand nombre d'acteurs mondiaux et des effets globaux qu'ils engendrent. Il recommande donc qu'un instrument juridiquement contraignant transversal du Conseil de l'Europe soit rédigé de manière à faciliter l'adhésion d'États qui n'en sont pas membres, mais qui en partagent les valeurs tels que le Canada, le Japon, les États-Unis et le Mexique.

Une convention internationale ou une convention-cadre a été finalement élaborée à l'issue de ces travaux ayant rassemblé les 46 États membres du Conseil de l'Europe, l'Union européenne et 11 États non-membres ainsi que des représentants issus du secteur privé, de la société civile et du monde universitaire, intervenant en qualité d'observateurs. La Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit a d'ailleurs déjà été adoptée par le Comité des ministres, le 17 mai 2024. Elle sera ensuite ouverte à la signature des États le 5 septembre 2024.

La Convention vise les activités du cycle de vie des systèmes d'IA susceptibles d'interférer avec les droits de l'homme, la démocratie et l'État de droit, mais uniquement lorsqu'elles sont menées par des organismes publics ou des acteurs privés agissant en leur nom. Pour ce qui est du secteur privé, une formulation plus flexible a été adoptée offrant aux États deux modalités pour se conformer à ses principes et obligations pour la régulation de ce secteur : choisir d'appliquer directement les obligations de la Convention-cadre, ou d'adopter d'autres mesures pour s'y conformer.

La Convention-cadre comprend également une série d'obligations pour les États qui doivent garantir que les droits humains, la démocratie et l'État de droit sont respectés lors du développement et de l'utilisation des systèmes d'IA. Elle laisse toutefois une grande marge d'interprétation et de mise en œuvre à de nombreux endroits et se réfère plutôt à des principes généraux tels que la transparence, la fiabilité, la non-discrimination, le respect de la vie privée et de la protection des données à caractère personnel, etc.

Ce texte adopte une approche fondée sur les risques concernant la conception, le développement, l'utilisation et la mise hors service des systèmes d'IA, imposant de ce fait un examen attentif de toutes les potentielles conséquences négatives liées à l'utilisation de systèmes d'IA. Les États doivent notamment prendre des mesures afin que les responsables des systèmes d'IA anticipent et atténuent les risques. Ils doivent également évaluer la nécessité de moratoires ou d'interdictions pour certaines utilisations de systèmes d'IA. Ils doivent enfin prévoir des voies de recours pour les victimes en cas de violation des droits de l'homme et instaurer un organisme indépendant chargé de surveiller l'application des règles établies par la Convention.

Même s'il s'agit du tout premier texte d'ampleur internationale dans le domaine de l'IA qui soit juridiquement contraignant, sa portée reste limitée. En effet, la limitation du champ d'application de la Convention aux seuls organismes publics réduit considérablement l'objectif affiché du traité qui est d'être transversal. Aussi, il est donc important que les États appliquent directement la Convention au secteur privé, car une technologie aussi importante ne peut être laissée à des intérêts privés, mais doit être étroitement encadrée.

Ce champ d'application est encore plus limité en ce qui concerne la sécurité nationale, car les États ne sont pas obligés d'appliquer la Convention aux systèmes d'IA liés à la protection de leurs intérêts en matière de sécurité nationale. La seule exigence est que ces activités soient menées dans les limites du droit international et dans le respect des processus démocratiques. De même, la Convention-cadre ne s'applique pas aux questions de défense nationale ni aux activités de recherche et de développement, sauf lorsque les essais de systèmes d'intelligence artificielle risquent de porter atteinte aux droits de l'homme, à la démocratie ou à l'État de droit.

Ainsi, ce qui aurait dû être le premier traité contraignant sur l'intelligence artificielle s'est finalement transformé, au fil des négociations entre États, en une déclaration commune par laquelle ils s'engagent à respecter certaines obligations de la manière dont ils veulent le faire, ce qui conduira inévitablement à des divergences en ce qui concerne la mise en œuvre du texte. Malgré ce résultat mitigé, cette initiative pourra encourager des efforts similaires à l'avenir.

92 Faut-il élaborer une réglementation spécifique à l'IA ?

L'intelligence artificielle fait d'ores et déjà partie de notre quotidien. Ses nombreuses applications soulèvent parfois des questions juridiques inédites. Jusqu'à récemment, il n'existait pas véritablement de cadre juridique spécifique à l'IA. Or, l'on entend parfois dire à tort dans les médias qu'il n'y aurait pas de réglementation applicable à l'IA et qu'il y aurait un vide juridique. Le droit existant lui est pourtant bien applicable. En effet, cette dernière est principalement régie par la réglementation existante, que ce soit le RGPD pour la protection des données à caractère personnel ou les réglementations sectorielles dans le domaine de la finance, la santé, l'automobile, etc.

Le développement d'une IA de confiance est essentiel. Se pose alors la question de savoir comment établir un cadre juridique approprié pour régir cette technologie. Faut-il se limiter à une approche éthique ou doit-on mettre en place une réglementation ? Si l'on choisit l'approche éthique, plusieurs principes ont déjà été identifiés : respect de la personne humaine qui doit être en mesure de conserver son autodétermination totale et effective ; prévention de toute atteinte aux êtres humains ; principe d'équité permettant à chaque personne d'être en mesure de contester les décisions prises par ces systèmes d'IA ; principe de l'explicabilité impliquant la transparence des caractéristiques et des finalités des systèmes d'IA.

En raison des risques potentiels liés au développement de l'IA, l'approche éthique peut paraître insuffisante. Dès lors, une réglementation de l'IA semble s'imposer, mais parle-t-on alors de réglementer la technologie elle-même, ses objectifs ou ses applications spécifiques ? Vouloir encadrer une technologie en constante évolution apparaît comme une approche réductionniste pouvant conduire à une obsolescence rapide de la réglementation envisagée. L'évolution rapide d'une technologie complique en effet la tâche du législateur qui doit veiller à l'élaboration d'une réglementation qui ne serait pas rapidement obsolète tout en ne freinant pas l'innovation. Dans son rapport remis au gouvernement français en 2018, Cédric Villani rappelait, à ce titre, que : « *La loi ne peut pas tout [...] car le temps du droit est bien plus long que celui du code* ».

Par ailleurs, l'apprentissage automatique est souvent considéré comme une « boîte noire » en raison de l'opacité du processus mis en œuvre, ce qui rend incompréhensible le fonctionnement technique de l'IA. Cette difficulté se répercute sur la réglementation et sur l'acceptation sociale de cette réglementation. En effet, le droit peut difficilement régir une technologie dont il ne comprend pas le fonctionnement. Le risque de l'obsolescence de la loi est donc bien réel en raison de la difficulté de la règle de droit à appréhender une technique peu transparente. En dépit de cet obstacle, il est important que le débat relatif à la réglementation de l'IA ne soit pas abandonné entre les mains des entreprises privées du numérique. L'autorégulation est en effet traditionnellement considérée comme un cheval de Troie des entreprises américaines, majoritaires sur le marché de l'intelligence artificielle et parmi les acteurs qui développent ou utilisent des applications d'IA.

La nécessité d'une intervention législative peut se justifier par divers arguments dont celui de la sécurité que procure le cadre légal aux acteurs de l'IA dans leurs stratégies de développement économique et, également, celui de la confiance apportée aux utilisateurs. Elle permet aussi de renforcer la souveraineté numérique et d'assurer l'imposition d'un certain modèle de valeurs au développement de l'IA.

Il est vrai qu'il existe une certaine concurrence au niveau international pour définir un modèle de standards applicables à l'IA. La Chine, l'Union européenne et les États-Unis ont d'ailleurs tous lancé des initiatives pour réglementer cette technologie afin d'atténuer ses risques. Cette compétition pour réglementer l'intelligence artificielle a bien commencé avant ChatGPT, mais l'IA générative a considérablement accru cette tendance.

La Chine fait partie des premiers pays à avoir réglementé l'IA. Ainsi, en 2017, le gouvernement chinois a publié le « Plan de développement de nouvelle génération d'IA » ayant pour ambition de faire de la Chine le leader mondial en la matière, d'ici à 2030. Avec l'apparition des systèmes d'IA générative, l'administration chinoise du cyberespace (CAC) a publié, le 11 avril 2023, un projet de « Mesures administratives pour les services d'intelligence artificielle générative », qui aborde notamment les enjeux relatifs à la gouvernance des données, tels que les biais et la qualité des données d'entraînement. Les mesures finales sur l'IA générative sont entrées en vigueur le 15 août 2023. L'approche globale de la Chine demeure axée sur l'atténuation des dommages aux individus, le maintien de la stabilité sociale et du contrôle de l'État, tout en encourageant l'innovation et le développement de l'intelligence artificielle et en influençant le débat sur sa réglementation.

Les États-Unis, quant à eux, adoptent une approche de la réglementation de l'IA plus fragmentée voire restreinte. Il existe des projets de loi proposés aux niveaux fédéral et étatique, mais rien de similaire à l'IA Act européen pour le moment. Autrement dit, une réglementation globale juridiquement contraignante n'est pas le modèle adopté par ce pays. L'administration du président Joe Biden s'est cependant montrée très intéressée par la régulation de l'IA. Le 30 octobre 2023, a ainsi été signé un décret présidentiel intitulé *Safe, Secure, and Trustworthy Artificial Intelligence*. Ce nouveau décret vise notamment à établir de nouvelles normes en matière de sûreté et de sécurité de l'IA, à protéger la vie privée des Américains, à faire progresser l'équité et les droits civils, à défendre les consommateurs et les travailleurs, etc. Le décret ne crée toutefois pas de nouvelles obligations législatives. Il introduit plutôt un certain nombre d'orientations pour les agences gouvernementales.

La régulation adoptée par les États-Unis tend ainsi à se rapprocher de celle de l'Union européenne sur le fond puisqu'émergent des approches conceptuelles communes en dépit de divergences. Le règlement sur l'IA et le décret présidentiel contiennent, par exemple, des définitions proches qui s'inspirent des définitions proposées par l'OCDE. L'Union européenne et les États-Unis partagent ainsi une compréhension commune des systèmes d'intelligence artificielle « dignes de confiance » ou encore « centrés sur l'humain ». L'approche européenne s'avère toutefois plus ambitieuse que l'approche américaine, car elle repose sur un cadre réglementaire harmonisé basé sur des principes directeurs.

93 Qu'est-ce que le règlement sur l'intelligence artificielle ou IA Act ?

Pour garantir de meilleures conditions de développement et d'utilisation de cette technologie innovante, l'Union européenne a souhaité réglementer l'intelligence artificielle dans le cadre de sa stratégie sur le numérique. Le 21 avril 2021, la Commission européenne a rendu public l'IA Act (*Artificial Intelligence Act*) ou « loi sur l'IA », son projet de règlement sur l'intelligence artificielle. Cette proposition a été le résultat de nombreuses études, livres blancs et analyses démarrés en 2018 ayant souligné les lacunes de la législation en vigueur.

Répondant à plusieurs noms : « Règlement sur l'IA », « Loi européenne sur l'IA », ou encore « IA Act », le règlement a pour objectif de promouvoir le développement d'une IA « digne de confiance » et l'innovation en matière d'IA dans l'Union européenne, tout en respectant les droits et valeurs fondamentaux de celle-ci. Le règlement sur l'IA fait d'ailleurs partie d'un ensemble plus large de mesures politiques comprenant également le train de mesures sur l'innovation dans le domaine de l'IA pour soutenir les PME et start-up européennes et le plan coordonné sur l'IA pour soutenir les investissements en matière d'IA.

Le règlement sur l'IA a fait l'objet de longues négociations. Après la proposition de règlement présentée en 2021, un accord du Parlement européen et du Conseil le 8 décembre 2023, et une adoption officielle par le Conseil le 21 mai 2024, la première législation au monde sur l'intelligence artificielle a été publiée au *Journal officiel* de l'Union européenne du 12 juillet 2024. Il est entré en vigueur vingt jours après, soit le 1er août 2024. Son application sera toutefois progressive afin de permettre aux acteurs concernés de s'adapter graduellement aux nouvelles obligations et de mettre en place les mesures nécessaires pour se conformer aux dispositions du règlement. Il sera ainsi pleinement applicable 24 mois après son entrée en vigueur, à l'exception de l'interdiction des pratiques interdites, qui s'appliquera 6 mois après la date d'entrée en vigueur, des règles concernant l'IA à usage général, notamment la gouvernance (12 mois après l'entrée en vigueur), et des obligations pour les systèmes à haut risque (36 mois). Le règlement sera donc pleinement appliqué le 2 août 2026, soit deux ans après son entrée en vigueur.

Les États membres ont jusqu'au 2 août 2025 pour désigner les autorités nationales chargées de superviser les nouvelles règles. Le Bureau européen de l'intelligence artificielle, créé en février 2024 au sein de la Commission européenne, est chargé de superviser l'application et la mise en œuvre du règlement avec les États membres. Cet organisme aidera les entreprises à préparer, dès à présent, leur mise en conformité.

L'IA Act définit largement l'intelligence artificielle. Le « *système d'intelligence artificielle* » couvre tout « *système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels* ». Outre la mention de l'IA générative, cette définition fait reposer la définition

de l'IA sur deux approches algorithmiques : l'apprentissage automatique ou « *machine learning* » qui relève de ce que l'on nomme « l'IA connexionniste », d'une part, et les approches basées sur la logique et la connaissance, aux rangs desquelles les systèmes experts, souvent désignées comme relevant de l'« IA symbolique », d'autre part. Ce qui signifie que l'ensemble des systèmes automatisés n'ont pas vocation à être saisis par ce règlement, mais seulement des systèmes dotés d'une importante complexité découlant des deux approches algorithmiques employées pour les concevoir. En effet, sont principalement saisis par ce règlement les systèmes d'IA à risque inacceptable, à haut risque, et à risque systémique, laissant, sauf au titre des obligations de transparence à l'égard de l'humain interagissant avec des systèmes qui s'imposent à l'ensemble des systèmes d'IA, les systèmes d'IA à risque faible ou modéré en dehors de toute réglementation. Les systèmes automatisés exclus du règlement pourront toujours être régis par d'autres instruments européens consacrés à la sécurité des produits.

Par ailleurs, une nouvelle catégorie de modèles dits à usage général, notamment dans le domaine de l'IA générative a été prévue par le règlement. Ce « *système d'IA à usage général* » est défini comme étant un système d'IA « *qui a la capacité de répondre à diverses finalités, tant pour une utilisation directe que pour une intégration dans d'autres systèmes d'IA* ». Ces systèmes sont entraînés avec une grande quantité de données et sont capables d'effectuer un large éventail de tâches, tels que ChatGPT, ce qui les rend difficiles à classer dans les catégories précédentes. Les fournisseurs de ces modèles doivent notamment mettre à la disposition du Bureau de l'IA et aux autorités nationales compétentes une documentation technique et des instructions d'utilisation, se conformer à la directive sur les droits d'auteur et publier un résumé du contenu utilisé pour l'entraînement de leurs algorithmes.

Le règlement s'appliquera aux fournisseurs de systèmes et de modèles d'IA, à ceux qui les mettent en œuvre ou qui les importent et distribuent, dès lors que ces systèmes sont mis sur le marché dans l'Union européenne ou que leur utilisation a une incidence sur des personnes situées sur ce territoire. Son champ d'application est donc très large avec une forte portée extraterritoriale, puisqu'il s'appliquerait à tout système d'IA ayant un impact dans l'Union, quel que soit le lieu d'établissement du fournisseur. À l'instar du RGPD, l'IA Act pourrait ainsi devenir une norme mondiale et inspirer les États qui souhaiteraient réglementer les systèmes d'IA.

94 Que prévoit le règlement ? Que signifie l'approche fondée sur les risques ?

Le règlement tente de concilier la volonté de faire de l'Union européenne un acteur de premier plan dans le domaine de l'IA en permettant le développement d'une IA « digne de confiance » et de garantir un niveau élevé de protection au niveau de la santé, de la sécurité et des droits fondamentaux. La conciliation de ces deux intérêts apparaît dans la méthode adoptée par ce texte. En effet, les autorités européennes ont préféré une approche par les risques plutôt que de mettre en place un cadre juridique commun à tous les systèmes d'IA.

L'approche fondée sur les risques qui a été choisie par le législateur européen dans l'élaboration de l'IA Act, constitue « *la base d'un ensemble proportionné et efficace de règles contraignantes* » (cons. 27). Le législateur européen explique sa méthode en indiquant que : « *cette approche devrait adapter le type et le contenu de ces règles à l'intensité et à la portée des risques que les systèmes d'IA peuvent générer. Il est donc nécessaire d'interdire certaines pratiques inacceptables en matière d'IA, de fixer des exigences pour les systèmes d'IA à haut risque et des obligations pour les opérateurs concernés, ainsi que de fixer des obligations de transparence pour certains systèmes d'IA* » (cons. 26).

À cet effet, le règlement opère une répartition des systèmes d'IA en fonction du niveau de risques pour la santé, la sécurité et les droits fondamentaux. Autrement dit, il s'agit d'une approche graduée selon les risques : les applications d'IA sont classées selon leur niveau de risque et les règles juridiques varient selon le niveau de risque identifié.

L'évaluation du risque doit permettre de classer les systèmes d'IA en quatre niveaux :

- Les systèmes d'IA interdits en raison du risque inacceptable : systèmes d'IA utilisés pour la manipulation inconsciente, l'exploitation des vulnérabilités des personnes, la notation sociale, etc. ;
- Les systèmes d'IA à haut risque (les plus nombreux) : les systèmes biométriques, les systèmes utilisés dans le recrutement, ou pour des usages répressifs ;
- Les systèmes d'IA à risque faible : obligations de transparence en ce qui concerne l'utilisation des chatbots pour lesquels les utilisateurs devront être informés qu'ils interagissent avec une machine ;
- Les systèmes d'IA à risque minimal : par exemple, les filtres anti-spam ou les jeux vidéo.

Ces catégories ont pour vocation de contrôler au mieux les IA et d'empêcher les dérives, notamment avec les applications qui auraient tendance à collecter des données sensibles ou à participer à la manipulation du comportement humain. L'essentiel des dispositions du règlement vise toutefois à encadrer les IA à haut risque. Sans être inacceptables, certains usages présentent en effet un haut risque. Ils font ainsi l'objet d'exigences strictes notamment de documentation, de qualité des données, de surveillance humaine et de gestion des risques. Par ailleurs, des règles spécifiques s'appliqueront aux IA génératives, comme ChatGPT d'*Open AI*, pour s'assurer de la qualité des données utilisées dans la mise au point des algorithmes et du respect des droits d'auteur.

L’enjeu de cette approche par le risque est d’atteindre la meilleure conciliation entre le respect des droits fondamentaux et de la sécurité de l’IA, d’une part, et la limitation des coûts de mise en conformité, d’autre part. Le règlement engage donc les entreprises à de nouvelles conformités aux fins de protéger la société des effets potentiellement néfastes de l’intelligence artificielle. Ce cadre juridique s’appliquera aussi bien aux acteurs publics que privés. Si le système d’IA est disponible sur le marché européen, l’entreprise devra alors se conformer aux obligations prévues par ce texte.

Il est certain que la qualification du système d’IA sera un enjeu fondamental tant les obligations varient. Le rattachement à une catégorie risque d’être une source de contentieux en raison du caractère contraignant du cadre applicable aux systèmes d’IA à haut risque qui contraste avec celui prévu pour les systèmes à risque faible ou minimal.

Certains considèrent d’ailleurs que le règlement risquerait de brider l’innovation dans un domaine en constante évolution et dans lequel certains concurrents internationaux auraient moins de contraintes à respecter. L’objectif est pourtant d’intégrer l’Union européenne dans un processus de compétitivité internationale économique et réglementaire. En effet, les principaux acteurs, notamment américains et chinois, ont déjà investi le marché de l’IA et négliger le sujet de la réglementation pourrait conduire à l’imposition de leur modèle à l’Europe. C’est la raison pour laquelle l’Union européenne a décidé de s’emparer de ce sujet et de proposer une réglementation horizontale correspondant à ses valeurs.

Plusieurs options normatives ont été envisagées allant d’une régulation souple, fondée sur l’autorégulation, à un encadrement législatif strict et obligatoire, pour tous les systèmes d’IA, en passant par une approche sectorielle. C’est la voie médiane qui a finalement été choisie consistant à adopter une approche normative horizontale fondée sur le risque généré par le système d’IA, associé au cadre souple des codes de conduite volontaires pour les systèmes présentant les risques les plus faibles pour les droits fondamentaux. Le règlement reconnaît ainsi la nécessité de sortir d’une simple approche éthique, au profit de règles de droit ayant force obligatoire sachant que l’éthique et le droit sont complémentaires. En effet, le droit peut s’inspirer des valeurs éthiques qui, une fois intégrées dans la loi, prennent force obligatoire et exécutoire. En faisant le choix du droit, l’Union européenne ramène dans le giron du droit certains principes éthiques.

95 Qu'est-ce qu'un système d'IA présentant des risques inacceptables au sens du règlement ?

Au sommet de la pyramide des risques figure les systèmes d'IA impliquant des risques considérés comme inacceptables au regard des valeurs européennes, notamment des droits fondamentaux. Ces systèmes d'IA sont prohibés car ils constituent une menace pour les personnes physiques. Il s'agit des systèmes d'IA considérés comme présentant une menace pour la sécurité, les moyens de subsistance et les droits des personnes, allant de la notation sociale par les gouvernements aux jouets utilisant une assistance vocale encourageant les comportements dangereux.

L'article 5 § 1 du règlement énumère, à ce titre, certains usages :

- Les systèmes d'IA ayant recours à des techniques subliminales, manipulatrices ou trompeuses pour fausser le comportement et altérer le libre arbitre d'une personne ;
- L'exploitation des vulnérabilités liées à l'âge, au handicap ou à la situation socio-économique pour fausser le comportement et causer des dommages importants (par exemple, des jouets activés par la voix qui peuvent induire des comportements dangereux pour les enfants) ;
- Les systèmes qui utilisent des données biométriques pour classer les personnes selon des catégories spécifiques telles que la race, la religion ou l'orientation sexuelle ;
- La notation sociale, soit l'évaluation ou la classification d'individus ou de groupes sur la base de leur comportement social ou de leurs traits personnels ;
- La prédiction d'infraction pénale concernant une personne ;
- La reconnaissance des émotions sur le lieu de travail et dans les établissements d'enseignement ;
- La constitution de bases de données de reconnaissance faciale par l'extraction non ciblée d'images faciales sur l'internet ou d'images de vidéosurveillance ;
- L'identification biométrique à distance en temps réel dans les espaces accessibles au public pour des finalités répressives, avec les exceptions suivantes :
- La recherche de personnes disparues, de victimes d'enlèvement et de personnes victimes de la traite des êtres humains ou de l'exploitation sexuelle ;
- La prévention d'une menace grave et imminente pour la vie ou d'un attentat terroriste prévisible ;
- Identifier les suspects de crimes graves (meurtre, viol, vol à main armée, trafic de stupéfiants et d'armes illégales, criminalité organisée, crimes contre l'environnement, etc.).

À noter que l'utilisation d'un système identification biométrique en temps réel basé sur l'IA n'est autorisée que lorsque la non-utilisation de l'outil causerait un préjudice considérable et doit tenir compte des droits et libertés des personnes concernées. Avant son déploiement, la police doit réaliser une évaluation de l'impact sur les droits

fondamentaux et enregistrer le système dans la base de données de l'UE. Dans des cas d'urgence dûment justifiés, le déploiement peut cependant commencer sans enregistrement, à condition qu'il soit enregistré ultérieurement sans retard injustifié.

Avant ce déploiement, doit également être obtenue l'autorisation d'une autorité judiciaire ou d'une autorité administrative indépendante, bien que, dans des cas d'urgence dûment justifiés, le déploiement puisse commencer sans autorisation, à condition que l'autorisation soit demandée dans les 24 heures. Si cette autorisation est rejetée, il doit être mis fin à son utilisation avec effet immédiat, et toutes les données, ainsi que les résultats et sorties de cette utilisation doivent être supprimés. Malgré la mise en place de garde-fous, l'on peut constater que les nombreuses dérogations finissent par vider le principe d'interdiction d'une partie de sa substance.

Les pratiques interdites en matière d'IA entreront en vigueur à partir du 2 février 2025, soit 6 mois après l'entrée en vigueur du règlement. Cette application anticipée par rapport aux autres dispositions se justifie par la nécessité d'interdire rapidement les systèmes d'IA présentant des risques inacceptables afin de prévenir tout dommage potentiel. Ils doivent donc être supprimés dès que possible pour assurer la sécurité et la protection des droits fondamentaux sans période de mise en conformité.

96 Qu’est-ce qu’un système d’IA à haut risque ?

Les applications d’IA à haut risque sont celles qui présentent des risques importants pour la santé, la sécurité ou les droits fondamentaux des personnes. La classification d’un système d’IA comme étant à haut risque repose sur la finalité du système d’IA. Il existe, à ce titre, deux grandes catégories de systèmes d’IA à haut risque.

Dans la première catégorie, un système d’IA est considéré comme étant à haut risque lorsque deux conditions cumulatives sont remplies. Les systèmes d’IA doivent tout d’abord être destinés à être utilisés en tant que composants de sécurité de produits (par exemple, application de l’IA en chirurgie assistée par robot) ou constituent eux-mêmes des produits de sécurité relevant de la législation européenne comme l’aviation, les voitures, les jouets. Ils doivent ensuite faire l’objet d’une évaluation de la conformité par un tiers, en vue de la mise sur le marché du produit.

Dans la deuxième catégorie, les systèmes d’IA à haut risque sont ceux qui touchent à des domaines sensibles relatifs au respect des droits fondamentaux énumérés dans une liste prévue à l’annexe III du règlement mentionnant l’identification biométrique et la catégorisation des personnes physiques ; l’emploi, la gestion des salariés et l’accès au travail indépendant (logiciel de tri des CV pour les procédures de recrutement, etc.) ; l’accès et la jouissance des services privés essentiels, ainsi que les services et avantages publics (notation de crédit empêchant les citoyens d’obtenir un prêt, etc.) ; les services répressifs susceptibles d’interférer avec les droits fondamentaux des personnes (par exemple, l’évaluation de la fiabilité des éléments de preuve) ; la gestion de la migration, de l’asile, du contrôle des frontières (examen automatisé des demandes de visa) ; l’administration de la justice et des processus démocratiques (solutions d’IA pour rechercher des décisions de justice) ; les infrastructures critiques (transports, électricité, eau, etc.), qui pourraient mettre en danger la vie et la santé des citoyens.

Cette liste n’est évidemment pas le fait du hasard et s’appuie sur de nombreuses études ayant montré les risques de discrimination envers les individus générés par des systèmes d’IA déployés dans ces domaines. Elle présente l’intérêt particulier de ne pas prendre en considération le statut juridique du produit, mais le secteur du système, c’est-à-dire de sa dangerosité présumée. De manière générale, les systèmes d’IA sont toujours considérés comme présentant un risque élevé s’ils établissent des profils de personnes, soit un traitement automatisé de données personnelles pour évaluer divers aspects de la vie d’une personne, tels que ses performances professionnelles, sa situation économique, sa santé, sa fiabilité, son comportement, sa localisation ou ses déplacements.

Il existe ainsi deux grandes catégories de systèmes d’IA à haut risque : d’une part, les systèmes d’IA utilisés en tant que composants de sécurité de produits relevant de la législation européenne et les autres systèmes d’IA autonomes soulevant essentiellement des questions quant au respect des droits fondamentaux, d’autre part. Tous ces systèmes d’IA à haut risque doivent ainsi être évalués non seulement avant la mise sur le marché, mais aussi tout au long de leur cycle de vie.

Il s’agit de systèmes réglementés parce qu’à haut risque. En d’autres termes, il s’agit de systèmes autorisés dès lors que les prescriptions légales sont satisfaites. Il s’agit de la plus importante catégorie du règlement pouvant avoir un impact important sur la vie des personnes concernées par son utilisation.

C’est la raison pour laquelle ces systèmes d’IA, tels que ceux utilisés dans les domaines de la santé ou de la justice, sont soumis à des diverses contraintes. Ils doivent être conformes à des normes strictes de gestion des risques, de gouvernance des données et de contrôle humain. Il s’agit bien évidemment de garantir leur fiabilité et leur sécurité en raison des risques d’atteintes aux droits fondamentaux.

L’approche basée sur les risques apparaît judicieuse même s’il existe toujours des questions relatives à l’étendue et aux critères de ce qu’il faut appeler un « haut risque ». Les différents acteurs doivent ainsi d’ores et déjà réaliser une analyse approfondie de chaque système d’IA afin de déterminer s’il appartient à cette catégorie. Cette classification sera utile pour anticiper les exigences en matière de mise en conformité. En principe, les dispositions relatives aux systèmes d’IA à haut risque sont applicables à partir du 2 août 2026, soit 24 mois après l’entrée en vigueur du règlement.

97 Quelles seront les obligations qui pèseront sur le fournisseur et le déployeur d'un système d'IA à haut risque ?

Dans la mesure où les IA à risque inacceptable sont proscrites, les exigences de conformité visent essentiellement l'utilisation et la mise à disposition des systèmes ou modèles d'IA à haut risque qui sont assorties de très nombreuses obligations de diligence à la charge des opérateurs économiques les mettant sur le marché européen ou en service dans l'Union.

Dans une logique de mise en conformité, le règlement prévoit des obligations préalables et postérieures à la commercialisation des systèmes d'IA à haut risque. Pour les systèmes d'IA qui ne sont pas considérés à haut risque, le texte encourage l'adoption volontaire de codes de conduite reprenant ces obligations. Le règlement sur l'IA énumère à ce titre différents acteurs, également appelés « opérateurs », qui jouent des rôles spécifiques dans la mise sur le marché et l'utilisation des systèmes d'IA. Ces acteurs comprennent le fournisseur, le fabricant de produits, le déployeur, le représentant autorisé, l'importateur ou le distributeur. Le règlement accorde une attention particulière au fournisseur et au déployeur.

Le fournisseur est défini comme étant « *une personne physique ou morale, une autorité publique, une agence ou un autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, que ce soit à titre onéreux ou gratuit* ». Quant au déployeur, il s'agit d'une « *personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant un système d'IA sous son autorité* », à l'exclusion de toute utilisation à des fins personnelles.

Le déployeur, également appelé utilisateur, partage une responsabilité importante dans l'application et le bon fonctionnement du système d'IA, même si le fournisseur demeure le principal responsable. Les autres intervenants dans la chaîne d'approvisionnement et de distribution sont également soumis à certaines obligations mais moins contraignantes. Dans le cas d'un système d'IA pour la prédiction des crises cardiaques, le fournisseur pourrait être, par exemple, une entreprise spécialisée dans la recherche cardiovasculaire qui utilise les systèmes d'IA pour créer des algorithmes analysant les données de santé des patients afin de détecter les risques de crise cardiaque. Le déployeur pourrait être, quant à lui, une société de services de santé numériques spécialisée dans l'intégration des systèmes d'IA dans les systèmes de gestion des dossiers médicaux électroniques. Son rôle consisterait à déployer le système d'IA dans les établissements de santé. Lorsqu'il existe une collaboration entre plusieurs opérateurs, il est ainsi nécessaire de définir au préalable le rôle de l'opérateur afin de déterminer les obligations qui sont à sa charge. La désignation d'un seul responsable en la personne du fournisseur permet d'éviter une dispersion de la responsabilité. Il est vrai toutefois que les obligations de mise en conformité peuvent être contraignantes, voire difficiles à mettre en œuvre.

S'agissant des obligations du fournisseur, il doit tout d'abord procéder à une évaluation de la conformité des systèmes d'IA avant leur mise sur le marché. Ce dernier pourra ainsi démontrer que ces systèmes sont conformes aux normes harmonisées ou aux

spécifications définies dans le règlement (par exemple, la qualité des données, la documentation et la traçabilité, la transparence, le contrôle humain, l'exactitude, la cybersécurité et la robustesse). Cette évaluation doit être répétée si le système ou sa finalité sont substantiellement modifiés.

Le fournisseur doit également mettre en place un système de gestion des risques, documenté et tenu à jour, pour identifier les risques connus et raisonnablement prévisibles pour la santé, la sécurité ou les droits fondamentaux lorsque le système est utilisé conformément à sa destination ou dans des conditions de mauvaise utilisation raisonnablement prévisibles, ainsi que les autres risques susceptibles d'apparaître. Il doit adopter des mesures appropriées et ciblées afin que le risque résiduel soit jugé acceptable. En d'autres termes, il doit réaliser une analyse de risques afin d'identifier et limiter autant que possible ces risques.

Par ailleurs, il doit mettre en place une politique de gouvernance des données pour surveiller, détecter et corriger les biais dans les systèmes d'IA. Cela suppose de sélectionner des ensembles de données de formation, de validation et de test qui soient pertinents et représentatifs et d'adopter des mesures appropriées pour atténuer les risques de biais. Il devra en outre établir une documentation technique pour démontrer la conformité et fournir aux autorités les informations nécessaires à l'évaluation de cette conformité. Enfin, le fournisseur devra assurer une journalisation des événements pertinents pour l'identification des risques, fournir des informations claires et adéquates au déployeur, garantir une surveillance humaine appropriée pour réduire au minimum les risques et coopérer avec les autorités compétentes en cas d'incident grave.

Quant aux déployeurs de systèmes d'IA à haut risque, ils ont la responsabilité de garantir une utilisation sûre et conforme de ces technologies, et d'assurer une surveillance continue de leur fonctionnement. Par ailleurs, ils doivent mettre en place des mesures techniques et organisationnelles appropriées pour s'assurer que l'utilisation des systèmes est conforme aux instructions fournies avec ces systèmes. En outre, ils doivent assurer un contrôle humain adéquat, veiller à ce que les données d'entrée des systèmes soient pertinentes et représentatives pour assurer le bon fonctionnement du système lorsqu'ils en ont le contrôle, et enfin informer les fournisseurs, les importateurs ou distributeurs, ainsi que les autorités de surveillance compétentes en cas de risque ou d'incident grave.

Pour assurer l'effectivité de ces dispositions, les États membres devront instaurer des sanctions effectives, proportionnées et dissuasives en cas de violation des règles applicables aux systèmes d'IA. Les sanctions peuvent être civiles mais aussi administratives. L'IA Act prévoit précisément des amendes administratives dont le montant varie selon la nature de la non-conformité (non-respect des usages interdits, manquement aux exigences pour les applications à haut risque, etc.) et la catégorie de l'entreprise. Le montant des amendes administratives va de 1 % à 7 % du chiffre d'affaires annuel mondial de l'entreprise réalisé au cours de l'exercice précédent, ou déterminées à partir de montants seuils allant de 7,5 à 35 millions d'euros, le montant le plus élevé étant retenu.

Le règlement sur l'IA établit ainsi un cadre juridique pouvant être adapté aux évolutions futures puisqu'il définit des exigences et des obligations axées sur les résultats sans prôner des solutions techniques concrètes, laissant ainsi le soin aux opérateurs la

possibilité d’adapter ces règles aux nouvelles solutions technologiques.

98 Quelle est la réglementation applicable aux données à caractère personnel traitées par l'IA ?

L'entraînement des algorithmes consomme une quantité importante de données, notamment de données à caractère personnel dont l'usage est encadré pour protéger la vie privée des personnes. Le règlement général sur la protection des données (RGPD), entré en application le 25 mai 2018, encadre en effet toutes les formes de technologies qui traitent des données à caractère personnel. On peut donc en déduire qu'il s'applique aussi à l'IA et à l'apprentissage automatique.

Les notions d'IA et d'apprentissage automatique n'apparaissent d'ailleurs pas dans aucune disposition du RGPD. L'absence de référence à de telles technologies s'explique par le principe de « neutralité technologique » de ce règlement. En conséquence, les principes et règles définis par ce dernier sont susceptibles de s'appliquer quelle que soit la technologie considérée dès lors que des données à caractère personnel sont traitées. Concrètement, le RGPD s'appliquera à tous les traitements de données personnelles aussi bien lors de la phase de développement que lors celle d'utilisation (déploiement) d'un système d'IA.

Toute structure privée ou publique, quels que soient sa taille, son pays d'implantation et son activité, peut être concernée. En effet, le RGPD s'applique à toute organisation qui collecte et traite des données personnelles pour son compte ou non, dès lors qu'elle est établie sur le territoire de l'Union européenne, ou que son activité cible directement des résidents européens. Le responsable du traitement est ainsi tenu de respecter ces dispositions, sous peine de sanctions qui peuvent être très lourdes.

Le règlement introduit dans le corpus de règles applicables en matière de protection des données la notion d'« *Accountability* », souvent traduite en français par « principe de responsabilité ». Ce principe découle de l'approche par les risques du RGPD que l'on retrouve à nouveau dans l'IA Act. L'article 24 du RGPD insiste d'ailleurs sur le fait que « *le responsable du traitement met en œuvre des mesures techniques et organisationnelles appropriées pour s'assurer et être en mesure de démontrer que le traitement est effectué conformément au présent règlement. Ces mesures sont réexaminées et actualisées si nécessaire* ». Responsabiliser les acteurs concernés est l'un des enjeux majeurs de cette réglementation. Ils doivent pouvoir démontrer qu'ils ont identifié, évalué, et encadré leurs risques en matière de protection des données personnelles et prouver ainsi leur conformité au RGPD.

La mise en œuvre de cette politique de gestion de la conformité devra être adoptée en fonction du niveau des risques : plus les traitements seront considérés comme sensibles en termes de protection de la vie privée et des données, plus le responsable du traitement devra justifier, dans sa documentation interne, de mesures élevées de protection.

Le RGPD a introduit, à ce titre, deux nouveaux concepts importants : d'une part, la protection des données dès la conception (*Privacy by design*), et d'autre part, la protection des données par défaut (*Privacy by default*). Le responsable du traitement doit donc dès la conception, dans le cadre de la phase projet, prendre en compte les principes de protection des données afin qu'ils soient initialement inclus dans l'application (par

exemple : la possibilité d'extraire des données de manière anonymisée pour faire des statistiques, d'archiver des données à l'issue d'une période déterminée, etc.). La protection par défaut signifie, quant à elle, que le responsable du traitement doit mettre en œuvre par défaut les principes les plus protecteurs en matière de protection des données (par exemple : ne pas stocker un login sans avoir informé et recueilli le consentement de l'intéressé).

Il prévoit également les cas dans lesquels le traitement des données est conforme au droit. Un traitement de données à caractère personnel en effet obéir à un certain nombre de principes directeurs. Le premier principe est celui de la licéité, de la loyauté et de la transparence du traitement des données. Cela signifie que, d'une part, le traitement doit être fait sur une base licite (consentement, exécution d'un contrat, etc.) et de façon loyale, conformément à ce qui a été annoncé par le responsable du traitement à la personne concernée. D'autre part, la collecte et le traitement doivent être faits en toute transparence, c'est-à-dire que les données ne peuvent pas être collectées, ni traitées d'une manière déloyale vis-à-vis de la personne dont les données sont collectées, et de façon opaque, soit sans qu'elle n'ait été informée des traitements qui seront réellement faits de ses données personnelles.

Le second principe, quant à lui, est celui de la limitation des finalités qui signifie que les données personnelles doivent être collectées pour des finalités déterminées, explicites et légitimes. Le traitement de données doit ainsi répondre à un but ou à un besoin défini. En d'autres termes, les données collectées ne peuvent pas être traitées ultérieurement de manière incompatible avec les finalités initiales, c'est-à-dire pour une finalité différente que celle pour laquelle le consentement initial a été donné.

Concernant le troisième principe est celui de la minimisation des données qui signifie que les données collectées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées. Ce principe conduit à éviter de collecter et traiter des données à caractère personnel à moins que l'objectif recherché rende cette collecte indispensable. Par exemple : est-il nécessaire de collecter le numéro de téléphone de quelqu'un qui s'inscrit à une lettre d'information ? Les données doivent en outre respecter le principe d'exactitude, c'est-à-dire être exactes et, si nécessaire, tenues à jour.

S'agissant du quatrième principe, il concerne la limitation de la conservation qui suppose que les données soient conservées uniquement le temps nécessaire à l'accomplissement de la finalité poursuivie lors de la collecte. Une fois l'objectif atteint, les données doivent être en principe supprimées. Enfin, le dernier principe est celui de l'intégrité et de la confidentialité des données à caractère personnel, ce qui signifie que le responsable du traitement doit garantir la sécurité des données qu'il détient pour empêcher leur perte, leur destruction ou que des tiers non autorisés y aient accès.

Il conviendra ainsi d'articuler ces principes avec ceux prévus par le règlement sur l'IA, ce qui peut poser des difficultés en raison de l'effet boite noire de certains algorithmes. Le RGPD impose, par exemple, la collecte du consentement pour tout traitement des données à caractère personnel. Comment donner son consentement à un traitement dont la finalité n'est même pas explicitée, ni même comprise avant les résultats ?

L’articulation de ces textes ne semble pas si évidente mais elle revêt une importance fondamentale, car la donnée est au cœur des systèmes d’IA, tant au stade de leur élaboration que de leur utilisation.

99 Comment articuler le RGPD et l'IA Act ?

L'IA Act et le RGPD ne réglementent pas les mêmes objets et n'adoptent pas la même approche sur tous les sujets. Il est vrai toutefois que la conformité à l'IA Act facilite celle au second texte. La conformité du système d'IA au RGPD est, par exemple, incluse dans la déclaration UE de conformité exigée par le règlement sur l'IA. Autrement, ces deux textes présentent une complémentarité, même si leurs objets et approches diffèrent.

Afin d'aider les professionnels à concilier innovation et respect des droits des personnes, la CNIL a publié ses premières recommandations sur le développement des systèmes d'IA, le 8 avril 2024. Ces recommandations se présentent sous la forme de sept fiches pratiques dont l'objectif est d'apporter des réponses concrètes aux enjeux juridiques et techniques liés à l'application du RGPD à l'IA. Elles concernent le développement des systèmes d'IA et non le déploiement, lorsque celui-ci implique le traitement de données à caractère personnel. Il est précisé que le développement comprend la conception du système, la constitution de la base de données, l'apprentissage et parfois l'intégration, alors que le déploiement comprend le calibrage et l'utilisation.

Sont ainsi concernés par ces recommandations : les systèmes fondés sur l'apprentissage automatique (*machine learning*) ; les systèmes dont l'usage opérationnel est défini dès la phase de développement et les systèmes à usage général qui pourront être utilisés pour nourrir différentes applications (« *general purpose AI* ») ; les systèmes dont l'apprentissage est réalisé « une fois pour toutes » ou de façon continue, par exemple en utilisant des données d'utilisation pour son amélioration.

La CNIL élabore un plan qui se décompose en sept étapes : définir la finalité du système d'IA ; qualifier les acteurs pour identifier leur responsabilité ; trouver la base légale du traitement ; s'assurer de la possibilité de réutilisation des données personnelles ; minimiser les données ; définir la durée de conservation des données et réaliser une analyse d'impact. Cette liste reprend les principes applicables à tout traitement de données personnelles afin de les adapter au contexte des systèmes d'IA.

La définition de la finalité du système d'IA est fondamentale. Un système d'IA reposant sur l'exploitation de données personnelles doit en effet être développé avec une « finalité », c'est-à-dire un objectif bien défini. Cette étape est importante pour identifier les données pertinentes afin de limiter la quantité de données stockées. Les traitements en phase de développement sont normalement soumis au RGPD. Il semble cependant possible de donner une finalité relativement générale (par exemple : développement d'un modèle de reconnaissance vocale capable d'identifier un locuteur, sa langue, son âge, son genre, etc.) et de la préciser au fur et à mesure.

Concernant la définition des responsabilités, plusieurs acteurs peuvent intervenir dans le développement d'un système d'IA, avec divers degrés d'implication sur les traitements de données personnelles (fournisseur, importateurs, distributeurs, et les utilisateurs ou déployeurs). Dans la mesure où le règlement sur l'IA utilise d'autres notions que celles du RGPD, la CNIL fournit des exemples afin d'identifier l'acteur qui pourra être qualifié de responsable de traitement ou de sous-traitants au sens de ce texte.

S’agissant de la base légale, il est rappelé que le RGPD en donne une liste exhaustive qui en comprend six possibilités : le consentement, le respect d’une obligation légale, l’exécution d’un contrat, l’exécution d’une mission d’intérêt public, la sauvegarde des intérêts vitaux, la poursuite d’un intérêt légitime. Sur ce point, la CNIL indique que le consentement constitue, le plus souvent, la base légale la plus appropriée, mais elle reconnaît les difficultés inhérentes à l’utilisation de cette base légale et envisage de manière plus pragmatique la possibilité de recourir à l’intérêt légitime. Il faut démontrer que les données personnelles sont vraiment nécessaires à l’entraînement du système, parce qu’il n’est pas possible de n’utiliser que des données ne se rapportant pas à des personnes physiques ou des données anonymisées. Elle insiste toutefois sur la nécessité de ne pas porter une « atteinte disproportionnée » à la vie privée des personnes concernées.

Si le développement du système d’IA implique la réutilisation d’une base de données à caractère personnel, ce qui sera souvent le cas en pratique, cette réutilisation doit être licite. Le responsable de traitement doit ainsi déterminer si ce traitement ultérieur est compatible avec la finalité pour laquelle les données ont été initialement collectées.

Par ailleurs, la CNIL reconnaît que le principe de minimisation n’interdit pas d’entraîner un algorithme avec des volumes très importants de données. Cependant les données personnelles collectées et utilisées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard de l’objectif défini. Ce principe de minimisation doit être appliqué de manière rigoureuse lorsque les données traitées sont sensibles (données de santé, etc.). Elle recommande d’établir une documentation des données utilisées pour assurer leur traçabilité.

À propos de la durée de conservation des données, la CNIL indique qu’il est possible de retenir une durée longue dans la mesure où cela est nécessaire pour les audits et la mesure de certains biais.

Enfin, concernant l’analyse d’impact qui permet de cartographier et évaluer les risques d’un traitement sur la protection des données personnelles et d’établir un plan d’action pour les réduire à un niveau acceptable, celle-ci est recommandée pour réduire les risques. Elle est même considérée comme nécessaire pour le développement des systèmes d’IA à haut risque.

Les recommandations de la CNIL sur le développement des systèmes d’IA constituent un outil utile pour faciliter l’articulation des deux textes, même si elles sont loin de répondre à toutes les questions. La CNIL poursuit ce travail d’accompagnement des opérateurs et prévoit de publier de nouvelles fiches sur la récupération de données sur l’internet, l’intérêt légitime comme base légale, l’exercice des droits d’accès, de rectification et d’effacement, le recours ou non à des licences ouvertes.

100 Existe-t-il d'autres réformes en cours en vue de réglementer l'IA ?

Le 28 septembre 2022, deux propositions de directives du Parlement européen et du Conseil ont été déposées : l'une envisageant une réforme de la responsabilité du fait des produits défectueux, et l'autre relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle. L'objectif de ces propositions est d'harmoniser la question de la responsabilité au sein des États membres afin d'éviter de brider l'innovation technologique tout en définissant un cadre de responsabilité qui inciterait par avance les fabricants, fournisseurs, utilisateurs à le respecter. La législation européenne aborde donc à la fois la question de la prévention des risques dans le cadre de l'IA Act, mais aussi la réparation du préjudice subi si ces risques se réalisent, liés à la mise en œuvre de solutions d'intelligence artificielle. Les adaptations du droit existant sont en effet nécessaires pour prendre en compte tous les aspects résultant du recours à un système d'IA, notamment en matière de preuve dans le cas d'un préjudice causé par un dysfonctionnement ou un produit défectueux.

Adoptée il y a près de quarante ans, la directive du 25 juillet 1985 sur la responsabilité du fait des produits défectueux a introduit des règles protectrices des consommateurs, mettant notamment en place un régime de responsabilité sans faute du fait des produits défectueux imputable au fabricant. Cela signifie que si un produit défectueux cause un dommage, la responsabilité du fabricant peut être engagée même sans faute ni négligence. La Commission rappelle, à ce titre, que le régime de responsabilité des produits défectueux était difficilement adaptable aux notions modernes de l'économie digitale. Il était donc nécessaire de réviser cette directive à la lumière des évolutions liées aux nouvelles technologies, notamment à l'intelligence artificielle.

La proposition de directive étend le champ d'application du régime de responsabilité sans faute à de nouveaux types de produits. En effet, si la directive du 25 juillet 1985 ne vise que les biens « meubles », la notion de produit englobe désormais les produits numériques et leurs mises à jour, et notamment les logiciels, les systèmes d'IA ou encore les services digitaux (drones, robots, systèmes domestiques intelligents etc.). À ce titre, l'article 4 de la directive précise que les fichiers de fabrication numériques sont considérés comme des produits. La directive ne s'appliquera toutefois pas aux logiciels libres développés ou fournis en dehors d'une activité commerciale.

À l'instar de tout autre produit ou composant, le système d'IA dommageable ne déclenchera une responsabilité civile que s'il est défectueux sachant qu'un produit est défectueux lorsqu'il n'offre pas la sécurité à laquelle le grand public peut légitimement s'attendre. Les critères d'appréciation de la défectuosité du produit ont été révisés afin de prendre en considération le caractère évolutif des produits mis sur le marché. C'est le cas de la « *capacité à poursuivre son apprentissage ou à acquérir de nouvelles caractéristiques après sa mise sur le marché ou sa mise en service* » ou des « *exigences de cybersécurité pertinentes pour la sécurité* ». L'un des éléments pris en compte lors de l'évaluation de la défectuosité est ainsi la capacité d'un produit à continuer à apprendre,

d'acquérir de nouvelles caractéristiques ou connaissances. Cet ajout est destiné à couvrir la technologie de l'intelligence artificielle basée sur des techniques d'apprentissage automatique.

Quant au dommage, outre une réparation des dommages causés par la mort ou les lésions corporelles et les dommages aux biens, est prévue une indemnisation des dommages causés par une perte ou une corruption de données.

Par ailleurs, la nouvelle de directive allège la charge de la preuve pour faciliter l'accès à l'indemnisation des victimes et tenir compte des difficultés probatoires rencontrées par les victimes de produits défectueux. Si la charge de la preuve continue à reposer sur le demandeur, sa demande sera cependant facilitée grâce à deux mécanismes. La victime du dommage causé par un produit défectueux qui a « *présenté des faits et des éléments de preuve suffisants pour étayer la plausibilité de sa demande en réparation* » pourra adresser une injonction au défendeur de divulguer les éléments de preuve pertinents dont il dispose. Le non-respect de cette injonction entraînera l'application d'une présomption de défectuosité du produit et il incombera au défendeur de la réfuter. La défectuosité est également présumée dans les cas où le plaignant rencontre des difficultés excessives dues à la complexité technique ou scientifique pour prouver la défectuosité du produit ou le lien de causalité, telles que celles impliquant l'IA ou des produits technologiques complexes. Ce mécanisme assimilable à une preuve « *par vraisemblance* » peut être utile lorsqu'un système d'IA a pu causer un dommage, sous réserve que le juge estime que la victime a suffisamment prouvé la contribution du produit au dommage et la probabilité de la défectuosité dommageable du produit.

La directive responsabilise tous les « opérateurs économiques », incluant les fabricants/producteurs, mais également tous les intervenants de la chaîne d'approvisionnement dont le fournisseur d'un service connexe, le représentant autorisé ou encore le distributeur, et même les plateformes en ligne si elles ne sont pas qualifiées d'hébergeurs.

Cette nouvelle directive du 23 octobre 2024 relative à la responsabilité du fait des produits défectueux vient d'être publiée le 18 novembre au *Journal officiel* de l'UE. Elle entrera en vigueur et abrogera l'ancienne le 9 décembre 2026. Elle sera alors applicable aux produits mis sur le marché ou mis en service après cette date.

Cette réforme intervient dans un contexte de développement de l'intelligence artificielle qui rend nécessaire l'adaptation du cadre législatif pour assurer une protection efficace du consommateur. En effet, la complexité de cette technologie ne permet pas toujours d'identifier aisément le responsable du dommage. C'est la raison pour laquelle une seconde proposition de directive relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle a été publiée le même jour que la proposition de directive sur la responsabilité du fait des produits défectueux. Cette proposition de directive sur la responsabilité en matière d'IA vient ainsi compléter le règlement européen sur l'intelligence artificielle.

Elle poursuit trois objectifs principaux : établir des règles uniformes pour l'accès à l'information et l'allègement de la charge de la preuve en ce qui concerne les dommages causés par des systèmes d'IA, instaurer une protection plus large pour les victimes

(particuliers ou entreprises) et promouvoir le secteur de l'IA en renforçant les garanties. Elle s'appliquera aux dommages causés par tout type de système d'IA, qu'ils soient ou non à haut risque. Ces nouvelles règles visent les actions en responsabilité intentées au niveau national pour faute ou omission, quelle que soit la personne ayant commis celle-ci, afin d'obtenir la réparation pour tout type de dommage et pour tout type de victime (particuliers, entreprises, organisations, etc.). Les conditions d'engagement de la responsabilité sont inchangées, mais elles prévoient deux principales garanties pour les victimes en ce qui concerne la charge de la preuve et l'établissement du lien de causalité.

À l'avantage des victimes par rapport au droit commun de la responsabilité, la proposition établit tout d'abord une présomption réfragable de lien de causalité entre la faute du défendeur et le résultat produit par le système d'IA, lorsque la victime parvient à démontrer que le défendeur a commis une faute en ne respectant pas une obligation donnée pertinente pour le dommage et que l'existence d'un lien de causalité avec la performance de l'IA est raisonnablement probable. Le défendeur devra alors renverser cette présomption (en prouvant, par exemple, qu'une autre cause a entraîné le dommage)

Ensuite, les victimes disposeront d'un plus grand nombre d'outils pour demander réparation en justice, grâce à l'introduction d'un droit d'accès aux éléments de preuve auprès des entreprises et des fournisseurs, lorsque des systèmes d'IA à haut risque sont utilisés. Ces dernières pourront demander à la juridiction d'ordonner la divulgation d'informations concernant des systèmes d'IA à haut risque, ce qui leur permettra d'identifier la personne qui pourrait être tenue pour responsable et de découvrir la cause du problème.

Les caractéristiques spécifiques des systèmes d'IA, telles que l'opacité, l'autonomie, la complexité, l'adaptation continue et le manque de prévisibilité exigent ainsi une adaptation de la législation existante afin de lever les obstacles pour les victimes souhaitant obtenir la réparation d'un préjudice résultant de l'utilisation de tels systèmes. Ces nouvelles directives servent également les intérêts des entreprises qui pourront davantage anticiper la manière dont les règles de responsabilité en vigueur seront appliquées et évaluer ainsi les risques qu'elles encourent.

Conclusion

Nous arrivons au bout de cette exploration sans prétendre avoir couvert tous les aspects du sujet. Il était cependant nécessaire de démystifier cette technologie qui nourrit les fantasmes les plus extravagants. Cet examen a pour objectif d'anticiper et de comprendre les évolutions à venir.

La démocratisation de cette discipline est fondamentale afin de permettre à chacun de comprendre non seulement les aspects techniques de l'intelligence artificielle, mais aussi ses risques et ses enjeux éthiques, juridiques et économiques. Aucun secteur d'activité, privé ou public, ne sera épargné par cette quatrième révolution industrielle. Aucune révolution industrielle n'a d'ailleurs eu un impact aussi global sur notre mode de vie, notre manière de travailler ainsi que nos rapports sociaux. Il est donc aisé de comprendre pourquoi elle suscite autant de craintes que d'espoirs.

Prenant conscience de ce bouleversement, il nous appartient d'anticiper les effets qu'auront les usages de l'intelligence artificielle, afin d'en maîtriser les risques et en garder le contrôle. Le besoin d'une éthique de l'intelligence artificielle ne fait guère de doute, elle est même indispensable. En soi, les technologies sont neutres tant sur le plan moral qu'éthique. En l'état, l'intérêt premier de l'éthique serait de résister aux logiques productivistes et managériales voulant toujours contrôler plus étroitement les personnes au détriment de leurs libertés fondamentales. Il s'agit donc avant tout de prévenir les risques d'asservissement de l'humain à la technique. Nous devons donc nous assurer que l'intelligence artificielle soit entourée de garanties suffisantes afin d'éviter qu'elle ne porte atteinte aux droits fondamentaux. La question de l'IA soulève donc inévitablement la question de la nécessité d'adapter le droit pour encadrer le développement de ces nouveaux outils.

Le cadre juridique mis en place par le règlement sur l'intelligence artificielle ouvre la voie à cette réflexion juridique qui devra se poursuivre afin d'aboutir à une réglementation de l'IA susceptible d'apporter confiance et sécurité aux utilisateurs, sans brider la recherche et les innovations. Le défi est de taille, car il s'agit de trouver un équilibre entre la course à l'innovation technologique et le développement de l'utilisation de l'intelligence artificielle au sein de l'Union européenne afin de créer un marché unique numérique, tout en respectant les droits fondamentaux et les valeurs européennes. L'autorégulation actuelle devra donc laisser la place à une réglementation fondée sur le niveau de risque engendré par le système d'IA.

Quoi qu'il en soit, le cadre juridique proposé par l'Union européenne pose les bases d'une approche basée sur l'éthique et la prévention des risques inhérents aux nouvelles technologies. Sans remettre en cause la domination technologique sino-américaine dans ce domaine, l'Europe espère assurément atteindre ses objectifs en termes de souveraineté numérique et défendre sa vision de l'intelligence artificielle. La question de l'intelligence artificielle se trouve ainsi au cœur d'une véritable bataille géopolitique et économique.

L’Europe n’est pourtant pas la seule à s’inquiéter des risques du modèle promu par les États-Unis et la Chine, en témoigne le scandale *Cambridge Analytica* ayant montré qu’il existait une conscience, bien au-delà du vieux continent, qu’une technologie sans régulation présentait de nombreux risques. Cette bataille de l’IA n’est d’ailleurs pas seulement une course aux moyens, elle est aussi une lutte des modèles et des valeurs. La gouvernance et la réglementation jouent ainsi un rôle fondamental dans la détermination des orientations que prendra l’intelligence artificielle dans les années à venir.

Par conséquent, les autorités publiques doivent s’emparer de toutes ces questions et ne pas laisser se développer de manière anarchique des pratiques profitant aux entreprises dominant le marché de cette technologie.

Bibliographie

A. Beelen, P. Dambly, *(R)évolution de l'intelligence artificielle : vers un cadre juridique et technique de l'IA*, éd. Anthemis, 2023.

A. Bensamoun (dir.) et G. Loiseau (dir.), *Droit de l'intelligence artificielle*, coll. « Les Intégrales », volume 15, LGDJ, 2e éd., 2022.

D. Brenet, *L'intelligence artificielle expliquée – Des concepts de base aux applications avancées de l'IA*, éd. Eni, 2024.

A. A. Casilli, *En attendant les robots – Enquête sur le travail du clic*, coll. « La couleur des idées », Seuil, 2019.

F. Cazals, C. Cazals, *Intelligence artificielle – L'intelligence amplifiée par la technologie*, préf. A. Bouverot, éd. De Boeck Supérieur, 2019.

J.-L. Dessalles, *Des intelligences très artificielles*, éd. Odile Jacob, 2019.

J.-G. Ganascia, *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle ?*, coll. « Science ouverte », Seuil, 2017.

J.-G. Ganascia, *Intelligence Artificielle : vers une domination programmée ?*, coll. « Idées reçues », éd. Le Cavalier Bleu, 2017.

L. Julia, *L'intelligence artificielle n'existe pas*, First Édition, 2019.

Y. Le Cun, *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*, éd. Odile Jacob, 2019.

Y. Meneceur, *L'intelligence artificielle en procès. Plaidoyer pour une réglementation internationale et européenne*, préf. A. Garapon, postface J. Kleijssen, Bruylant, 2020.

E. Sadin, *L'intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*, éd. L'échappée, 2018.

Glossaire

Algorithme : ensemble d'instructions ou de règles qu'un ordinateur suit pour effectuer une tâche particulière. Dans le domaine de l'intelligence artificielle, l'algorithme s'appuie sur des modèles mathématiques complexes. Il est auto-apprenant, c'est-à-dire que le traitement et l'analyse de quantités importantes de données lui permettent de s'adapter, d'évoluer et de se reconfigurer pour fournir des résultats précis. Les algorithmes sont à l'œuvre dans tous les domaines, des requêtes sur les moteurs de recherche à la sélection d'informations recommandés aux internautes.

Analyse prédictive : un domaine de l'analyse statistique qui extrait l'information à partir des données pour prédire les tendances futures et les motifs de comportement.

Apprentissage automatique ou apprentissage machine (*machine learning*) : un sous-domaine de l'IA dans lequel les systèmes informatiques apprennent et s'améliorent automatiquement sans être explicitement programmés. Après une phase d'entraînement préliminaire sur un large corpus de données, le programme est capable de résoudre des problèmes pour lesquels il n'a pas été développé. Ils utilisent des données et des algorithmes pour faire des prédictions et prendre des décisions. L'apprentissage automatique est fréquemment utilisé pour le traitement du langage naturel et la vision par ordinateur, ou pour effectuer des diagnostics et des prévisions.

Apprentissage non supervisé : un type d'apprentissage automatique dans lequel l'algorithme utilise des données non étiquetées, c'est-à-dire sans connaître leur résultat attendu. L'algorithme est capable de découvrir des patterns et des structures sans connaissances ou cibles préalables. L'apprentissage non-supervisé peut être, par exemple, utilisé par les algorithmes de recommandations susceptibles de prédire un comportement ou besoin en fonction des habitudes de navigation.

Apprentissage par renforcement : un type d'apprentissage automatique dans lequel un programme fonctionne en autonomie en le confrontant à des situations dont il tire des leçons. À chaque action, il reçoit des récompenses ou des pénalités, l'amenant ainsi à affiner ses stratégies pour maximiser ses gains. L'efficacité de l'apprentissage par renforcement a été attestée dans certains jeux stratégiques comme le jeu de go.

Apprentissage profond (*deep learning*) : un sous-domaine de l'apprentissage automatique qui utilise des réseaux neuronaux multicouches, appelés réseaux neuronaux profonds, pour simuler le pouvoir de décision complexe du cerveau humain. Il alimente de nombreux produits et services du quotidien, tels que les voitures autonomes ou encore l'IA générative.

Apprentissage semi-supervisé : une combinaison de l'apprentissage supervisé et non supervisé, dans lequel l'algorithme est entraîné avec une petite quantité de données étiquetées et une grande quantité de données non étiquetées. L'algorithme peut classer des données inédites à l'aide des modèles appris.

Apprentissage supervisé : un type d'apprentissage automatique dans lequel l'algorithme est entraîné avec des données étiquetées fournissant la réponse souhaitée pour chaque entrée. L'algorithme est capable de prédire les résultats pour de nouvelles données en

se basant sur ce qu'il a appris pendant l'entraînement. L'apprentissage supervisé est utile pour classifier et détecter des anomalies ou établir des probabilités dans un contexte donné. Il est utilisé, par exemple, pour la reconnaissance d'images et la traduction automatique.

Biais : ce sont des raccourcis pris par le cerveau menant à des conclusions pouvant être erronées. Le terme biais dans le contexte de l'IA fait référence au phénomène selon lequel les algorithmes présentent des biais indésirables résultant des données d'apprentissage. Il peut en résulter une attribution erronée de certaines caractéristiques à certains groupes sur la base de stéréotypes (couleur de peau, sexe, etc.), ce qui peut conduire à des discriminations.

Big data : ensemble de données massives et complexes, de source hétérogène (*open data*, données propriétaires, données achetées commercialement), ne pouvant être facilement gérés, traités ou analysés avec les méthodes traditionnelles de traitement des données. De nouveaux algorithmes ont été développés afin de pouvoir les stocker, les classer et les analyser.

Blockchain : une technologie de stockage et de transmission d'informations, sécurisée et transparente, fonctionnant sans organe central de contrôle. Elle est connue pour être la technologie à l'origine des cryptomonnaies, comme le Bitcoin.

Chatbot ou agent conversationnel : un programme informatique conçu pour simuler une conversation avec des utilisateurs humains, surtout sur l'internet. Il peut répondre à des requêtes en fonction de scénarios prédéfinis ou en autonomie grâce à l'apprentissage automatique. Les progrès de l'IA en matière de traitement de langage naturel (*Natural Language Processing* ou NLP) lui permettent d'analyser et de comprendre les messages des utilisateurs.

Couche de neurones : dans les réseaux de neurones artificiels, les signaux peuvent traverser successivement plusieurs couches de neurones. Les couches d'entrée reçoivent les signaux de l'extérieur du réseau alors que les couches de sortie les transmettent vers l'extérieur. Il peut y avoir plusieurs couches intermédiaires entre les deux couches. Ces couches sont alors dites cachées, car elles demeurent invisibles de l'extérieur du réseau.

***Data Mining* ou l'exploration de données** : il s'agit d'une technique visant à découvrir des patterns et des connaissances à partir de grands ensembles de données en utilisant des techniques statistiques, des algorithmes de *machine learning* et des méthodes d'analyse. Elle a pour but d'extraire de l'information utile et des *insights* à partir de données brutes.

Data science : La *data science* ou « science des données » est un domaine regroupant un ensemble de disciplines relatives à la collecte, la gestion et l'analyse des données.

Data scientist : un expert ayant pour mission la collecte, l'organisation, l'analyse et l'interprétation de vastes ensembles de données afin de fournir des *insights*, c'est-à-dire une donnée informative, une tendance ou bien une recommandation qui provient de l'analyse de différentes informations collectées. Il s'agit de faire remonter à la surface des informations pouvant aider les organismes à prendre des décisions plus intelligentes.

Donnée à caractère personnel : toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Parmi celles-ci, les données sensibles concernent les données à caractère personnel relatives à l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale ainsi que les données génétiques, les données biométriques, les données concernant la santé, la vie sexuelle ou encore l'orientation sexuelle.

Explicabilité : notion faisant référence à la capacité de comprendre comment une décision a été prise par un système d'IA. C'est donc la capacité à comprendre les raisons qui ont conduit à une décision donnée prise par une machine.

GAN ou *Generative Adversarial Network* (réseau antagoniste génératif) : une technique de *machine learning*. Elle repose sur deux réseaux neuronaux distincts, le générateur et le discriminateur, qui sont entraînés simultanément par un processus de compétition mutuelle. Le générateur crée de nouvelles données, tandis que le discriminateur évalue la qualité de ces données. Les deux réseaux s'entraînent en boucle, améliorant ainsi leurs performances respectives.

GPT (*Generative Pre-trained Transformer*) : une architecture de modèle de langage développée par *OpenAI*. GPT est largement utilisé pour diverses tâches de traitement du langage naturel, telles que la génération de texte, la traduction automatique et la réponse aux questions.

Hallucination : la situation lorsqu'une IA génère des informations fausses ou non fondées, souvent en réponse à un prompt ou une question. C'est notamment le cas lorsqu'une IA imagine des faits ou des détails qui ne sont pas fondés sur ses données d'entraînement.

Intelligence artificielle : ensemble de théories et de techniques ayant pour objet la compréhension des mécanismes de la cognition en vue de concevoir des machines capables de simuler l'intelligence humaine.

IA faible ou restreinte : une IA faible est capable d'exécuter une (ou plusieurs) tâches de façon autonome, mais dans un cadre défini par l'homme et à la suite de décisions prises par lui seul. Elle n'a pas de conscience ou de sensibilité à l'inverse de l'IA forte.

IA forte ou générale : une forme d'intelligence artificielle similaire à l'intelligence humaine, dotée de conscience et de sensibilité, et capable de résoudre tout type de problème comme le ferait un être humain. Elle est considérée comme l'un des défis les plus difficiles à réaliser dans le domaine de l'IA, car elle exige une compréhension complète de l'environnement de l'homme et de son fonctionnement.

IA générative ou *Generative AI* : une branche spécialisée de l'intelligence artificielle qui crée de nouveaux contenus en apprenant à partir de vastes ensembles de données. Alors que les systèmes d'IA traditionnels peuvent reconnaître les modèles ou classifier le contenu existant, l'IA générative peut créer du nouveau contenu sous plusieurs formes, comme du texte, une image, un fichier audio ou du code logiciel. Les technologies d'IA générative les plus connues incluent les GAN (*Generative Adversarial Networks*) et les modèles de langage comme GPT (*Generative Pretrained Transformer*).

***Internet of Things* (IoT) ou l'internet des objets** : objets du quotidien connectés à l'internet pouvant recevoir et transmettre des informations en ligne, tels que des montres connectées ou la domotique. Un objet connecté récolte, grâce à des capteurs (de température, de vitesse, d'humidité…), des données et les envoie *via* l'internet, afin qu'elles soient analysées par des ordinateurs. L'intelligence artificielle joue un rôle crucial dans l'analyse et l'utilisation de ces données.

Justice prédictive : une justice prédite par des algorithmes grâce à l'analyse de grandes masses de données de la justice afin de repérer des récurrences permettant de prévoir autant qu'il est possible l'issue d'un litige.

Médecine prédictive : partie de la médecine qui vise à étudier les maladies et les pathologies au travers de marqueurs génétiques et biologiques afin de déterminer les prédispositions à ces maladies avant même l'apparition des premiers symptômes dans le but de faciliter leur prise en charge voire d'éviter leur survenue.

***Mind uploading* ou téléchargement de l'esprit** : une idée transhumaniste selon laquelle le contenu du cerveau humain pourrait être traduit dans un code binaire informatique, et donc téléchargé (*upload*) dans un ordinateur.

Open data : un mouvement d'ouverture et de mise à disposition des données produites et collectées par les services publics (administrations, collectivités locales…). Ces données peuvent être réutilisées de manière non-onéreuse dans les conditions d'une licence spécifique pouvant notamment préciser ou prohiber certaines finalités de réemploi.

Profilage : un traitement de données à caractère personnel réalisé afin d'évaluer certains aspects de la vie d'une personne physique (situation économique, santé, préférences personnelles, etc.).

Prompt : les requêtes textuelles adressées par les utilisateurs à des systèmes d'IA générative tels que ChatGPT, DALL-E ou Midjourney. En fonction de l'affinage et de la contextualisation du prompt, les réponses apportées seront plus ou moins exhaustives.

Réalité augmentée : une technologie qui permet d'intégrer en temps réel des éléments virtuels, généralement en 3D, au sein de l'environnement réel de l'utilisateur pour lui permettre d'analyser et d'interagir à la fois avec les mondes physique et digital.

Réalité virtuelle : une technologie qui permet de créer numériquement un nouvel environnement entièrement artificiel. Il peut s'agir d'une reproduction du monde réel ou bien d'un univers totalement imaginaire. Elle ne doit pas être confondue avec la réalité augmentée.

Réseau de neurones artificiels : algorithme informatique inspiré du fonctionnement du cerveau humain. Il se compose de plusieurs couches de neurones artificiels qui transmettent et traitent des informations pour comprendre des modèles complexes et faire des prédictions. Les réseaux neuronaux peuvent aider les ordinateurs à prendre des décisions intelligentes avec une assistance humaine limitée. Les réseaux de neurones sont beaucoup utilisés pour le traitement du langage naturel, la reconnaissance vocale et d'image.

Robotique : ensemble des disciplines et des techniques qui permettent de concevoir, de construire et de programmer des robots capables d'exécuter de manière autonome une ou plusieurs tâches dans des environnements spécifiques. La robotique fait appel à

l'intelligence artificielle pour permettre aux robots de prendre des décisions et de s'adapter à leur environnement.

Superintelligence artificielle : un type d'IA forte capable de surpasser les capacités cognitives et l'intelligence humaines. À ce jour, il s'agit d'une simple théorie.

Système d'intelligence artificielle : un système conçu pour simuler le fonctionnement de l'intelligence humaine afin d'exécuter des fonctions relevant normalement de celle-ci.

Système expert : un système informatique développé à partir du savoir de l'homme du métier afin de résoudre des problèmes relevant de son domaine de compétence.

Test de Turing : nommé d'après le célèbre mathématicien et informaticien Alan Turing, il teste la capacité d'une machine à se comporter comme un humain. La machine réussit le test si un humain ne peut pas distinguer la réponse de la machine de celle d'un autre humain.

***Text Mining* ou fouille de textes** : un processus visant à extraire des informations utiles et des connaissances à partir de grandes quantités de données textuelles. Il repose sur des techniques de traitement du langage naturel, d'apprentissage automatique et d'analyse statistique pour découvrir des tendances et des modèles dans les textes.

Traitement du langage naturel (*Naturel Language Processing*, NLP) : un domaine impliquant la linguistique, l'informatique et l'intelligence artificielle. Il désigne la capacité d'un programme informatique à comprendre le langage humain tel qu'il est parlé et écrit. Il englobe une variété de sous-tâches, dont la traduction automatique, la reconnaissance vocale, l'analyse des sentiments et la génération de texte. Cette technologie est notamment utilisée par les assistants intelligents tels que Siri d'*Apple* ou Alexa d'*Amazon*.

Vision par ordinateur : une branche de l'intelligence artificielle dont l'objectif est de permettre aux ordinateurs de percevoir, d'analyser et de comprendre les images et les vidéos. Les applications de la vision par ordinateur incluent la reconnaissance faciale, la détection d'objets et la navigation autonome des véhicules.

Voiture autonome : également connue sous le nom de véhicule autonome ou véhicule à conduite automatique, est un véhicule capable de se déplacer et de naviguer dans un environnement sans intervention humaine en faisant usage de plusieurs technologies avancées, dont la vision par ordinateur, le radar, le lidar, la localisation et cartographie simultanés, et l'intelligence artificielle, pour percevoir un environnement et prendre des décisions de navigation.

Index alphabétique

Les numéros renvoient aux questions.

W