Intelligence Artificielle et Propriété Intellectuelle

Pierre Fernandez, Août 2021

Résumé Opérationnel

Dans ce rapport, nous traiterons des liens entre Intelligence Artificielle (IA) et Propriété Intellectuelle (PI) à travers deux aspects. Le premier traitera des données et des algorithmes entraînés avec ces données : quelles sont les droits existants, sont-ils suffisants ? Le second traitera de ce que produisent les algorithmes, notamment, une IA peut-elle générer des œuvres protégeables par droit d'auteur ? Ces deux aspects étant relativement distincts, il est possible de lire chaque partie du rapport indépendamment. Il est toutefois suggéré d'entamer par le préambule pour avoir une idée plus claire des enjeux présentés par la suite.

Table des matières

Introduction

En 1950 dans sa publication *Computing Machinery and Intelligence*, Alan Turing définit un test qui évaluerait la capacité d'une machine à présenter une intelligence similaire à celle d'un être humain. Ce test consiste à mettre en discussion à l'aveugle un humain d'une part avec une machine et d'autre part avec un humain. Si l'humain n'arrive pas à distinguer la machine de l'humain, c'est que la machine est « intelligente ».

Même si cette définition de l'Intelligence Artificielle (IA) est étroite et ne soulève qu'une légère attention de la part des chercheurs, elle illustre bien le fait que les techniques d'aujourd'hui permettent de créer des machines qui possèdent des capacités susceptibles d'être nommées ‘’intelligence‘’. La norme ISO 2382-28 la définit, elle, comme "la capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage". En effet, au cours des dernières décennies, on a vu se développer des programmes informatiques capables de diagnostiquer des maladies, de résoudre des équations sous forme symbolique, d'analyser des circuits électroniques, de comprendre des phrases parlées et écrites, de traduire, de voir et comprendre son environnement ou bien encore d'écrire des programmes informatiques.

C'est une réalité, l'IA fait désormais partie de notre quotidien et est présente dans la majeure partie des secteurs industriels ou académiques. À titre d'exemple, dans le domaine du conseil en propriété industrielle, il existe déjà sur le marché des solutions basées sur l'IA qui aident à la révision et à la rédaction des revendications de brevets, à la recherche et à l'examen de l'art antérieur des brevets, ou encore à la recherche et à la surveillance instantanée de copyrights. L'importance croissante de l'IA, comprise comme John McCarthy l'a un jour (1956) définie comme "la science et l'ingéniosité de la fabrication de machines intelligentes", a fini par transformer (si ce n'est défier) les fondements des droits de propriété intellectuelle, et ce pour plusieurs raisons.

Tout d'abord, l'apprentissage automatique développé durant la dernière décennie nécessite un apport constant et croissant de données. Une fois que la puissance de calcul est satisfaisante, la différence se fait par la quantité et la qualité des données disponibles : plus on a de données et plus la qualité de ces données est bonne, plus on est à même de prédire et d'analyser. Une partie de la recherche en IA est d'ailleurs amenée non pas à créer de nouveaux algorithmes, mais à leur permettre d'être appliqués à plus grande échelle, avec plus de paramètres et plus de données d'apprentissage. Ces données, qui constituent pour certains « l'or noir du XXIe siècle », sont et seront de plus en plus amenées à être collectées, traitées, stockées puis utilisées par des entreprises et des gouvernements. Par ailleurs, le nouveau mode de fonctionnement des algorithmes/logiciels créées par IA, avec ces données, entraîne une difficulté à les protéger et à déterminer leur propriété. Des problématiques nouvelles émergent de cette utilisation massive des données, notamment à travers le concept de mémorisation. (*Droits des données et des algorithmes*)

Jusqu'à présent, nous avons surtout mentionné des applications comme l'imagerie médicale, la conduite autonome, la recommandation, etc. Ce sont les plus connues du grand public et celles qui ont le moins attrait à la créativité. Une question souvent posée à leur sujet est celle de la responsabilité : qui est responsable d'un mauvais diagnostic d'IA en imagerie médicale ou encore d'un accident de voiture autonome ? Est-ce le médecin/le conducteur ? Est-ce celui qui a créé l'application d'Intelligence Artificielle ? Est-ce l'IA elle-même ? Ces questions sont très présentes dans les médias et disposent d'un homologue moins traité dans le domaine de la propriété intellectuelle. En effet, l'IA est encore peu vue comme « inventeur », alors même qu'elle commence à être source de création. À titre d'exemple, en 2018, « Portrait d'Edmond de Belamy », une œuvre peinte par une intelligence artificielle a été vendue aux enchères à plus de 430.000 dollars à New York. L'algorithme utilisé avait au préalable été divulgué en open source par un autre artiste, ce qui a naturellement soulevé la question de la paternité de l'œuvre.

Qui est vraiment son créateur ? Obvious : le collectif qui a utilisé le réseau génératif, Robbie Barrat : le jeune étudiant qui a publié son code et son modèle, ceux qui ont le droit des données utilisées pour l'apprentissage, ou bien l'IA elle-même ? Le cadre légal de la propriété intellectuelle est confronté à de nouveaux enjeux, et les lois actuelles devront évoluer pour être à disposées à les aborder. (*L'IA en tant qu'inventeur ?*)

Portrait d'Edmond de Balamy, Obvious, 2018 : œuvre réalisée par intelligence artificielle et vendue plus de 400.000 euros aux enchères de la maison Christie's

1. Préambule

1.1 Introduction à l'Intelligence Artificielle et à l'apprentissage profond

Les différents types d'apprentissage automatique L'apprentissage machine (très souvent appelé ML en référence au terme anglais *Machine Learning*) consiste essentiellement à automatiser l'identification de tendances dans un jeu de données. Selon le cas, il peut utiliser des connaissances et des techniques issues de la théorie statistique, des algorithmes d'optimisation et du traitement de grands volumes de données. Les systèmes qui intègrent l'apprentissage automatique apprennent à identifier et à prédire des comportements en fonction de données en entrée.

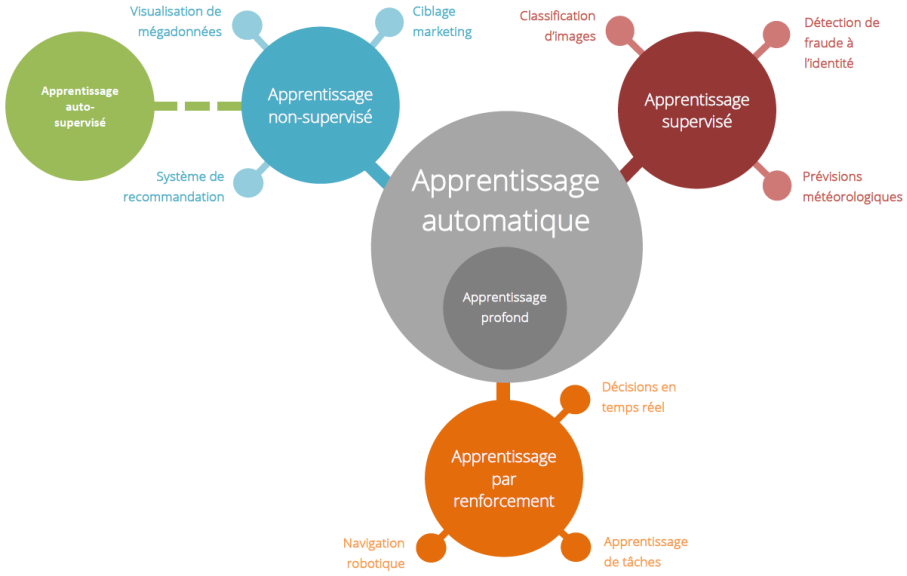

Figure 1: Les différents types d'apprentissage automatique - Source : https://www.coe.int/fr/web/artificial-intelligence/glossary

On distingue usuellement les algorithmes de ML en fonction du type de données dont l'on dispose :

- `Supervisé` : L'apprentissage supervisé utilise des données étiquetées (ou *labeled data* en anglais) pour entraîner les algorithmes. Les données d'entrée sont fournies à l'algorithme qui apprend ensuite à attribuer l'étiquette de sortie (ou *label*). Les applications qui utilisent l'apprentissage supervisé associent des groupes de caractéristiques à des ensembles de données pour déterminer une valeur de sortie. C'est souvent cette stratégie qui est utilisée pour la classification d'images par exemple.
- `Non supervisé` (ou auto-supervisé) : Cela se produit lorsqu'aucune donnée étiquetée n'est disponible pour le pré-entraînement du modèle d'apprentissage automatique. En d'autres termes, seules les données d'entrée sont disponibles. Par conséquent, dans ce schéma, l'algorithme cherche à décrire les données et en à trouver une organisation. L'objectif est d'apprendre à l'algorithme à simplifier l'analyse, à déduire, prédire ou classer sur la base de similitudes ou de différences. Il est très utilisé dans les systèmes de recommandation et dans la segmentation d'images.
  Il est parfois étendu à l'apprentissage semi-supervisé, où peu de données étiquetées (en général les plus couteuses) et beaucoup de données non étiquetées sont utilisées pour entraîner les algorithmes. L'apprentissage semi-supervisé est généralement plus performant que

l'apprentissage supervisé seul grâce au plus grand nombre de données utilisées. C'est souvent ce qui est utilisé dans le domaine médical.

- `Renforcement` : Dans ce type d'apprentissage, les algorithmes apprennent à partir des stimuli qu'ils reçoivent du monde extérieur en réponse à leurs actions (un peu comme un bébé). De cette façon, il évalue chaque action et détermine laquelle est la meilleure en fonction de l'expérience qu'il a reçue jusqu'ici. C'est ce type d'algorithmes qui a été utilisé pour Alpha Go, la première machine à avoir battu un champion (Lee Sedol) au jeu de Go.

L'apprentissage profond (deep learning) et réseaux de neurones Le *deep learning* est une sous-catégorie de l'apprentissage automatique qui consiste à modéliser des abstractions de haut niveau dans les données, à l'aide d'architectures informatiques complexes et généralement très profondes (d'où le terme *deep*). En effet, une image peut être représentée sous certaines formes (par exemple, un ensemble de pixels), mais certaines représentations sont meilleures que d'autres pour la classification. Par exemple, pour classifier un chat ou un chien, il est beaucoup plus facile d'avoir une représentation de moustaches, de pates, de poils, etc. et de regrouper ses attributs pour déterminer la nature de l'image que de traiter l'image pixel par pixel. Un modèle profond peut alors être vu comme un ensemble de nombreuses fonctions mathématiques qui transforme une image faite de pixels en une représentation plus utile pour la classification ou l'analyse.

Les réseaux de neurones sont des modèles d'apprentissage profonds dont la conception est inspirée du fonctionnement des neurones biologiques (dont on connait à peu près le fonctionnement grâce aux travaux de Hubel et Wiesel, prix Nobel de physiologie). Depuis 2012, ce sont eux qui ont révolutionné l'apprentissage automatique et qui ont fait connaître l'Intelligence Artificielle du grand public : ils sont utilisés pour la robotique, la traduction automatisée, la reconnaissance de visages, etc.

Le neurone dit « formel » est conçu comme un automate doté d'une fonction de transfert qui transforme ses entrées en sortie selon des règles logiques, arithmétiques et symboliques précises, et de paramètres appris lors de l'entraînement. Assemblées en réseau profond, ces neurones formels ont la capacité d'opérer rapidement des classifications et d'apprendre progressivement à les améliorer.

Entraînement ou apprentissage de modèles profonds L'apprentissage, pour les réseaux de neurones formels, consiste à calculer les paramètres de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés lors de l'apprentissage, aussi proches que possible des sorties « désirées », par exemple l'étiquette de l'image que l'on veut classer, ou une fonction représentant la vraisemblance d'une image créée.

Les techniques d'apprentissage des réseaux de neurones formels sont des algorithmes d'optimisation : ils cherchent à minimiser l'écart entre les réponses réelles du réseau et les réponses désirées, en modifiant les paramètres de chaque couche du modèle par étapes successives (appelées « itérations »). Ainsi à chaque itération, le réseau prédit une sortie, puis ses paramètres sont adaptés pour que la sortie s'approche de celle désirée. Par ailleurs, plus le réseau voit de données diverses, plus il sera performant, d'où l'importance du jeu de données.

Réseau discriminatif et génératif L'apprentissage supervisé de réseaux de neurones permet de résoudre efficacement des problèmes de classification, par exemple de reconnaissance d'images. Ce qui est peut-être encore plus surprenant, c'est que ces réseaux de neurones sont également utilisés de façon non-supervisée afin de générer automatiquement des textes ou des images virtuelles, ce que l'on appelle souvent des « deep fakes ». On distingue ces deux types de réseaux par les termes discriminatif et génératif.

Réseau discriminatif *Réseau génératif*

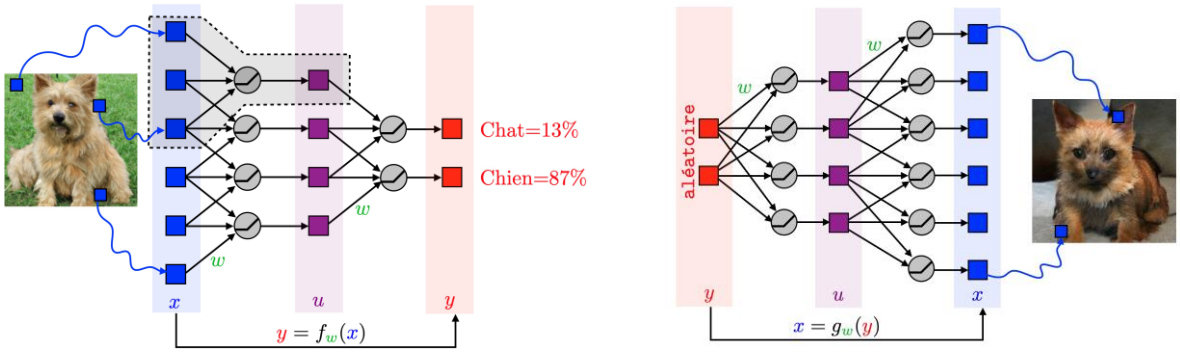

Figure 2 : Réseau discriminatif et génératif - Source : Les mathématiques des réseaux de neurones, Gabriel Peyré

Les réseaux génératifs sont de plusieurs types (GAN, VAE, etc.). Ce sont ces réseaux qui ont par exemple été employés pour générer le portrait présenté en introduction, et qui sont utilisés pour la génération de deep fakes en image, vidéo ou audio que les media ont beaucoup couverts ces dernières années.

1.2 Introduction à la propriété intellectuelle

Propriété Intellectuelle Nous pourrions définir la propriété intellectuelle comme l'ensemble des droits dont disposent les auteurs sur leurs créations, tels que les droits d'auteur, les copyrights, les brevets et les marques. Ils permettent aux créateurs ou aux propriétaires de la propriété intellectuelle de récupérer un bénéfice de leur travail ou de leur l'investissement, en leur donnant le contrôle de l'utilisation de leur propriété. Les droits de propriété intellectuelle sont reconnus depuis longtemps dans divers systèmes juridiques. Certaines initiatives internationales ont vu le jour dès le XIX° siècle pour protéger le droit de la PI, notamment la *Convention de Paris pour la protection de la propriété industrielle (1883)* et la *Convention de Berne pour la protection des œuvres littéraires et artistiques (1886)*. Aujourd'hui, il existe plus de 25 traités internationaux sur la PI administrés par l'Organisation Mondiale de la Propriété Intellectuelle (OMPI). Les droits de propriété intellectuelle sont également protégés par l'article 27.2 de la Déclaration universelle des droits de l'homme (UDHR) :

> *« Chacun a droit à la protection des intérêts moraux et matériels découlant de toute production scientifique, littéraire ou artistique dont il est l'auteur »*

On distingue usuellement le droit d'auteur et des droits connexes (ou droits voisins) qui concernent les créations artistiques, les travaux scientifiques, les œuvres littéraires, etc. de la propriété industrielle qui regroupe marques, brevets d'invention, schémas industriels, etc.

Droit d'auteur et droits connexes / Copyright Tout d'abord, il est important de distinguer le terme copyright de celui du droit d'auteur. Le copyright, si largement utilisé au niveau international, est hérité du droit anglo-saxon. En particulier, le *Statute of Anne* (ou *Copyright Act*, 1709) en Angleterre a été la première loi sur le droit d'auteur au monde, et a servi d'inspiration aux lois nationales des autres pays anglo-saxons. En revanche, le terme de droit d'auteur provient du droit continental et plus précisément du droit français.

Malgré les particularités de chaque pays et les différences d'approche entre le droit anglo-saxon (*Common Law*, plus basé sur la jurisprudence que sur les lois) et le droit continental (plus basé sur les lois), il existe des traités internationaux qui harmonisent les règles de base du droit d'auteur. En particulier, la *Convention de Berne pour la protection des œuvres littéraires et artistiques* (modifiée en 1979) et le *Traité de l'OMPI sur le droit d'auteur* ou *WCT* (1996). Elles assurent notamment que :

- le droit d'auteur naisse dès la création de l'œuvre, sans être tenu d'enregistrer l'œuvre ni d'accomplir d'autres formalités pour en obtenir la protection (bien que certains pays aient mis en place des systèmes d'enregistrement volontaire du droit d'auteur).
- les pays soient tenus de protéger la plupart des œuvres sous droit d'auteur pendant toute la durée de vie du créateur et au moins 50 ans après la mort de ce dernier.

Cependant, une différence nette entre les deux est que le droit d'auteur peut être soit moral soit patrimonial, tandis que le copyright se concentre presque exclusivement sur l'aspect patrimonial ou économique.

Droit d'auteur Il est défini dans le livre I^{er} du *Code de la propriété intellectuelle* (Articles L111-1 à L137-4) qui présente les règles relatives à l'objet du droit d'auteur, aux prérogatives accordées aux auteurs ou créateurs, et à l'exploitation des droits.

L'objet du droit d'auteur, est assez imprécis. Il a par exemple été jugé que la saveur d'un aliment ne peut être qualifiée d'œuvre, faute de pouvoir obtenir de la création en cause une identification précise et objective. Pour accéder à la protection, la création doit être originale, relever d'une activité créatrice de la part de l'auteur et également répondre à certaines exigences de forme. Certaines catégories d'œuvres de l'esprit sont susceptibles d'être protégées par le droit d'auteur (livres, conférences, œuvres cinématographiques, etc. ; CPI L112-2 L112-3). C'est aussi le cas des logiciels, des codes de programmation et des bases de données. En revanche, les idées, concepts et méthodes mathématiques en tant que tels sont exclus par le droit d'auteur (mais pas leur expression).

Les prérogatives accordées aux auteurs peuvent être regroupés en deux grandes catégories : les droits moraux (paternité, intégrité, divulgation...) et les droits patrimoniaux (reproduction, distribution, communication publique, transformation).

- Le droit moral de l'auteur, très protecteur, garantit à celui-ci le respect de son nom, de sa qualité et de son œuvre. Ce droit est perpétuel, inaliénable, imprescriptible et transmissible à cause de mort aux héritiers de l'auteur. L'auteur est le seul à avoir le droit de divulguer son œuvre, c'est-à-dire de la communiquer au public.
- Les droits patrimoniaux de l'auteur confèrent à ce dernier un monopole d'exploitation. À ce titre, l'auteur dispose du droit de représentation (communication de l'œuvre au public par un procédé quelconque) et du droit de reproduction (fixation matérielle de l'œuvre par tous procédés qui permettent de la communiquer au public d'une manière indirecte). Ce monopole est toutefois contraint par quelques exceptions selon lesquelles l'auteur ne peut interdire « l'utilisation » de son œuvre à certains desseins (copie privée, courte citation, revue de presse…). Le droit exclusif d'exploiter une œuvre est limité dans le temps ; au décès de l'auteur, ce droit persiste pendant les soixante-dix années qui suivent.

Droits connexes Les droits connexes protègent quant à eux les droits de certaines personnes ou professions qui participent à l'activité de création mais qui ne remplissent pas les conditions requises pour bénéficier de la protection au titre du droit d'auteur. Ils concernent par exemple les artistes interprètes ou exécutants tels que les chanteurs et les acteurs, les organismes de radiodiffusion et les maisons de disques qui produisent les enregistrements sonores, ou encore les éditeurs de presse et les agences. La protection offerte par ces droits est analogue à celle du droit d'auteur mais la durée d'application est généralement plus courte que celle du droit d'auteur. (Articles L211-1 à L219-4 du CPI)

Brevets Le droit des brevets d'invention constitue l'une des branches du droit de la propriété industrielle. Elle est codifiée dans le livre VI du CPI. Au niveau européen, la Convention de Munich de 1973 pose les règles relatives à la délivrance du brevet européen. Au niveau international, le texte de référence est la Convention de l'Union de Paris de 1883.

Les brevets d'invention sont le moyen le plus répandu de protéger les droits des inventeurs. C'est un droit accordé par l'État pour la protection d'une invention. Il donne à son titulaire le droit exclusif d'empêcher des tiers d'exploiter commercialement l'invention protégée pendant une période limitée, en échange de la divulgation de l'invention au public. Ainsi, le titulaire du brevet peut empêcher d'autres personnes de fabriquer, d'utiliser, d'offrir à la vente, de vendre ou d'importer l'invention brevetée sans autorisation, et peut poursuivre en justice quiconque exploite l'invention brevetée sans autorisation. La théorie économique qui soutient ce système est que les avantages financiers découlant de l'exploitation du brevet et de la divulgation des inventions favoriseront l'innovation et relèveront le niveau technique de l'industrie d'un pays, avec des avantages évidents pour son commerce. En effet, en accordant un droit exclusif, le brevet devient une incitation dans la mesure où il procure à l'inventeur une reconnaissance de son activité créatrice et une rémunération matérielle pour son invention. En contrepartie de l'obtention de droits exclusifs, l'inventeur a l'obligation de divulguer l'invention brevetée au public, afin que des tiers puissent bénéficier des nouvelles connaissances et contribuer ainsi au développement technologique. La divulgation de l'invention est donc un critère essentiel dans les procédures de délivrance des brevets.

Les brevets ne s'appliquent pas qu'aux processus et produits physiques et chimiques et ne sont pas utiles qu'aux grandes entreprises. Ils peuvent généralement être obtenus pour tout domaine technologique, des trombones aux produits pharmaceutiques complexes. Il existe des milliers de brevets pour des produits de tous les jours tels que des filtres, des bouteilles en verre, des tissus ou des bicyclettes. L'article L611-10 du CPI précise le champ d'application du brevet :

> *« Sont brevetables, dans tous les domaines technologiques, les inventions nouvelles impliquant une activité inventive et susceptibles d'application industrielle. »*

En revanche, les découvertes ainsi que les théories scientifiques et les méthodes mathématiques ou bien les créations esthétiques ne peuvent être considérées comme des inventions.

Ce droit de brevet exclusif est accordé pour une durée limitée, 20 ans à compter de la date de dépôt de la demande, à condition que le titulaire paie les taxes annuelles de maintien en vigueur, et n'est valable que dans le pays où la protection est demandée (principe de territorialité). Une fois cette période passée, l'invention n'est plus sous brevet et chacun est libre de la fabriquer, la commercialiser ou l'utiliser.

Marques, modèles industriels, etc. Il existe d'autres moyens de protection de la propriété individuelle comme les licences, les marques, les dessins, etc. Ils sont moins pertinents dans le cadre et ne seront donc pas abordés.

2. Droits des données et des algorithmes

Avec le développement fulgurant des applications d'IA (et plus précisément de ML) dans de nombreux domaines, les données sont devenues l'un des enjeux industriels majeur de ces dernières années. Face à cet enjeu, la loi s'est adaptée pour mieux protéger les données des utilisateurs, notamment avec la mise en application du RGPD (Règlement Général pour la Protection des Données). Sur un autre plan, les programmes et algorithmes d'intelligence artificielle utilisant ces données disposent eux aussi de protection relevant à la fois du domaine de la propriété individuelle et industrielle. Toutefois l'interaction nouvelle et croissante entre données, qui renouvellent sans cesse les algorithmes d'IA à chaque réentrainement de modèles, et ces derniers qui utilisent et mémorisent ces données, pousse la législation à se réinventer.

2.1 Protection des données et RGPD

Histoire du RGPD Avec les progrès de la technologie et l'invention d'Internet, l'UE a reconnu la nécessité de protections modernes. C'est ainsi qu'en 1995, elle a adopté la directive européenne sur la protection des données, qui établit des normes minimales en matière de confidentialité et de sécurité des données, sur lesquelles chaque État membre s'appuie pour mettre en œuvre sa propre législation. Cette directive devenant caduque face à l'arrivée de nouveaux acteurs (notamment Facebook et Google) l'autorité européenne de protection des données a déclaré que l'UE avait besoin d'une approche globale de la protection des données personnelles et les travaux ont commencé pour mettre à jour la directive de 1995.

Le règlement général sur la protection des données (RGPD) est la loi sur la confidentialité et la sécurité la plus stricte au monde. Bien qu'elle ait été rédigée et adoptée par l'Union européenne (UE), elle impose des obligations aux organisations où qu'elles soient, tant qu'elles ciblent ou collectent des données liées à des personnes dans l'UE. Le règlement est entré en vigueur le 25 mai 2018. Avec le RGPD, l'Europe signale sa position ferme sur la confidentialité et la sécurité des données, à une époque où de plus en plus de personnes confient leurs données personnelles à des services de cloud et où les violations sont quotidiennes. Le règlement lui-même est vaste, d'une grande portée et assez peu précis.

Objet du RGPD L'article 4 de la réglementation définit de nombreux termes relatifs aux données personnelles, notamment :

- (1) «données à caractère personnel», toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée «personne concernée») ; est réputée être une «personne physique identifiable» une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale;
- (2) «traitement», toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction;
- (4) «profilage», toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique;
- (7) «responsable du traitement», la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement; lorsque les finalités et les moyens de ce traitement sont déterminés par le droit de l'Union ou le droit d'un État membre, le responsable du traitement peut être désigné ou les critères spécifiques applicables à sa désignation peuvent être prévus par le droit de l'Union ou par le droit d'un État membre;
- (8) «sous-traitant», la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui traite des données à caractère personnel pour le compte du responsable du traitement;

Principes de protection des données Le RGPD impose pour les entreprises, si elles traitent des données, de le faire conformément à sept principes de protection et de responsabilité décrits à l'article 5.1-2 :

- Licéité, loyauté et transparence - Le traitement doit être licite, loyal et transparent pour la personne concernée.
- Limitation de la finalité – Le traitement des données doit se faire pour les finalités légitimes spécifiées explicitement à la personne concernée lorsqu'elles ont été collectées.
- Minimisation des données - Ne collecter et traiter que les données absolument nécessaires aux fins spécifiées.
- Exactitude - Les données personnelles doivent être conservées exactes et à jour.
- Limitation du stockage –Les données d'identification personnelle ne peuvent être stockées que pendant la durée nécessaire à la réalisation de l'objectif spécifié.
- Intégrité et confidentialité - Le traitement doit être effectué de manière à garantir une sécurité, une intégrité et une confidentialité appropriées (par exemple, en utilisant le cryptage).
- Responsabilité - Le responsable du traitement des données doit être en mesure de démontrer la conformité au RGPD de tous ces principes.

Sécurité des données Les données doivent être traitées en toute sécurité en mettant en œuvre des "mesures techniques et organisationnelles appropriées". Les mesures techniques peuvent aller de l'obligation pour les employés d'utiliser une authentification à deux facteurs sur les comptes où sont stockées des données personnelles, à la conclusion d'un contrat avec des fournisseurs de services de cloud qui utilisent un cryptage de bout en bout. Les mesures organisationnelles peuvent être des formations du personnel, l'ajout d'une politique de confidentialité des données dans le manuel de l'employé, ou la limitation de l'accès aux données personnelles aux seuls employés de votre organisation qui en ont besoin. (Remarque : tout ceci est bel et bien appliqué chez Facebook même pour un stagiaire qui ne reste que 4 mois !)

En cas de violation des données, le responsable du traitement a 72 heures pour informer les personnes concernées, sous peine de sanctions. (Cette obligation de notification peut être levée si sont utilisées des garanties technologiques, telles que le cryptage, pour rendre les données inutiles à un attaquant).

Protection des données dès la conception et par défaut Désormais, tout ce qui est fait dans une organisation doit, "par conception et par défaut", prendre en compte la protection des données. En pratique, cela signifie que l'on doit tenir compte des principes de protection des données lors de la conception de tout nouveau produit ou activité. Le RGPD couvre ce principe à l'article 25. Supposons, par exemple, que vous lanciez une nouvelle application pour votre entreprise. Vous devez réfléchir, dès le début de la conception de l'application, aux données personnelles que l'application pourrait éventuellement collecter auprès des utilisateurs, puis envisager des moyens de minimiser la quantité de données et la manière dont vous les sécuriserez avec les dernières technologies (cryptage, tatouage, etc.).

Contexte de traitement L'article 6 énumère les cas dans lesquels il est légal de traiter des données personnelles. Le traitement (et la collecte, le stockage, la vente à des publicitaires, etc.) doit pouvoir se justifier par l'un des éléments suivants :

- La personne concernée a donné un consentement spécifique et sans ambiguïté pour traiter les données. (Par exemple, elle s'est inscrite sur une liste d'e-mails de marketing).
- Le traitement est nécessaire pour exécuter ou préparer la conclusion d'un contrat auquel la personne concernée est partie. (Par exemple, effectuer une vérification des antécédents avant de louer une propriété à un locataire potentiel).
- Pour se conformer à une des obligations légales. (Par exemple, recevoir une ordonnance d'un tribunal de votre juridiction).
- Pour sauver la vie de quelqu'un.
- Pour exécuter une tâche dans l'intérêt public ou pour exercer une fonction officielle. (Par exemple, une entreprise privée de collecte d'ordures).

- Il y a un intérêt légitime à traiter les données personnelles de ladite personne. Il s'agit de la base légale la plus souple, bien que les "droits et libertés fondamentaux de la personne concernée" l'emportent toujours sur les intérêts du traitant.

Une fois que la base légale du traitement des données a été déterminée, elle doit être documentée et la personne concernée doit en être informée (c'est le principe de transparence). Si l'entreprise décide plus tard de modifier cette justification, elle doit fournir une raison valable, documenter cette raison et en informer la personne concernée.

Consentement Il existe de nouvelles règles strictes sur ce qui constitue le consentement d'une personne concernée pour traiter ses informations. Il doit être "librement donné, spécifique, éclairé et sans ambiguïté". Les demandes de consentement doivent être "clairement distinguées des autres questions" et présentées dans "un langage clair et simple". Les personnes concernées peuvent retirer un consentement donné précédemment quand elles le souhaitent, et leur décision doit être respectée. Les enfants de moins de 13 ans ne peuvent donner leur consentement qu'avec l'autorisation de leurs parents. Des preuves documentaires du consentement doivent également être conservées.

Délégués à la protection des données Le délégué à la protection des données (Data Protection Officer en anglais, DPO) est une fonction née du RGPD dont le but est de conseiller et accompagner les organismes qui le désignent vis-à-vis de leur conformité. Tous les responsables du traitement des données ne sont pas tenus de désigner un délégué à la protection des données (DPD). Il existe trois conditions dans lesquelles ils en sont tenus :

- Être une autorité publique autre qu'un tribunal agissant à titre judiciaire.
- Leurs activités principales les obligent à contrôler des personnes de manière systématique et régulière à grande échelle. (Par exemple, Google ou Facebook).
- Leurs activités principales consistent à traiter à grande échelle des catégories particulières de données énumérées à l'article 9 du RGPD ou des données relatives aux condamnations pénales et aux infractions mentionnées à l'article 10. (Par exemple, un cabinet médical.)

Il est également possible de désigner un DPD même sans y être tenu.

Ses tâches de base consistent à comprendre le RGPD et la manière dont il s'applique à l'organisation, à conseiller les personnes de l'organisation sur leurs responsabilités, à organiser des formations sur la protection des données, à mener des audits et à contrôler la conformité au RGPD, et à servir de liaison avec les régulateurs.

Le droit à la vie privée des personnes Le RGPD reconnaît de nouveaux droits à la vie privée pour les personnes concernées, qui visent à donner aux individus plus de contrôle sur les données qu'ils prêtent aux organisations : le droit d'être informé, le droit d'accès, le droit de rectification, le droit à l'effacement, le droit de restreindre le traitement, le droit à la portabilité des données, le droit d'opposition et les droits relatifs à la prise de décision automatisée et au profilage.

2.2 Protection des algorithmes et programmes informatiques

Définition d'un algorithme d'IA dans le cadre légal Les chercheurs ne sont pas d'accord sur les termes de définition de ce qui constitue un algorithme. De plus, définir un algorithme à partir d'une seule expression, telle qu'une machine à états abstraits (structures de données) ou un récurseur ne suffit pas à en cerner toute la portée. D'une façon simple, les algorithmes peuvent donc être considérés comme une méthode constituant une séquence d'étapes utilisées pour calculer différentes variables de données afin de produire des résultats. De cette façon, les algorithmes peuvent "dire" aux ordinateurs comment accomplir des tâches lorsqu'ils sont incorporés dans des programmes informatiques. Ce processus est

rendu possible par le code, qui est la méthode d'implémentation réelle des algorithmes, obtenue en fournissant les instructions algorithmiques aux ordinateurs dans certains langages de programmation. Les programmes informatiques comprennent donc des séquences codifiées qui expriment des algorithmes (instructions).

Les algorithmes ont diverses structures de contrôle et sont donc divisés en sous-catégories algorithmiques, chacune justifiant une nouvelle définition en fonction de sa fonction de contrôle, par exemple les algorithmes de "tri" et de "recherche". Les algorithmes de ML revêtent ici une importance particulière puisqu'ils ne sont pas des processus fixes, mais essayent de prédire des résultats futurs sur la base de données historiques. Ils parcourent de grands ensembles de données à la recherche de modèles ou de corrélations entre différentes variables et le résultat final pour engendrer des prédictions. L'apprentissage automatique, une méthode par laquelle les systèmes informatiques apprennent à partir de données et agissent ensuite de manière autonome avec peu ou pas d'intervention humaine, facilite généralement ce processus. (Cf. 1.1 Introduction à l'Intelligence Artificielle et à l'apprentissage profond pour plus de détails)

Ces algorithmes sont de plus en plus souvent employés pour les systèmes d'IA introduits dans la société. En effet, leur influence est très étendue. Ils ont un impact direct sur notre vie de tous les jours, en déterminant des condamnations pénales, l'octroi de prêts ou l'admission à l'université, nos "fils d'actualité" et bien d'autres choses encore. Parfois, ces systèmes sont open source, mais souvent, les titulaires de droits cherchent à obtenir une protection complète de la propriété intellectuelle pour protéger leurs modèles commerciaux.

Une situation à cheval Plus précisément, les algorithmes sont définis dans le droit français comme « l'étude de la résolution de problèmes par la mise en œuvre de suites d'opérations élémentaires selon un processus défini aboutissant à une solution ». Ils ressortent à la fois de la propriété intellectuelle et de la propriété industrielle. En effet, un droit d'auteur apparaît sur le code source lors de l'écriture, mais ils peuvent aussi être considérés comme des inventions à des fins industrielles. La question du caractère brevetable ou non tient en partie à leur situation à cheval entre propriété intellectuelle (protection par simple droit d'auteur) et propriété industrielle (protection par brevet, etc.).

Tout d'abord ils constituent au sens du droit français une méthode mathématique, relevant des idées dites de « libre parcours », qui échappe à la protection au titre du droit d'auteur, et est exclue par le droit des brevets (Article L611-10 du CPI). Comme en pratique les algorithmes sont intégrés dans les codes sources d'un logiciel, ils sont alors potentiellement protégeables par un droit d'auteur dédié (Article L112-2 du CPI), à condition toutefois de satisfaire au critère d'originalité.

D'une autre manière, les algorithmes peuvent faire l'objet d'un brevet s'ils constituent un programme, incorporé à une invention brevetable, capable de produire des effets techniques supplémentaires. Il peut s'agir d'avancées notables dans les technologies d'IA ou de nouveaux procédés faisant appel à l'IA.

Ces deux formes de protection, par leur manque d'adaptabilité, sont de plus en plus caduques dans le cadre de l'application à l'IA. En effet, les algorithmes peuvent être amenés à évoluer en fonction des données qui leur sont fournies. Une troisième solution tend à être privilégiée afin d'éviter de révéler la configuration de l'algorithme aux concurrents : une protection par la directive « secret des affaires ».

La Directive sur la protection des savoir-faire et des informations commerciales non divulgués (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites (ou directive « secret des affaires »), est entrée en vigueur le 5 juillet 2016. Son objet, le « secret d'affaires » doit correspondre aux critères suivants :

- a) elles sont secrètes en ce sens que, dans leur globalité ou dans la configuration et l'assemblage exacts de leurs éléments, elles ne sont pas généralement connues des personnes appartenant aux

milieux qui s'occupent normalement du genre d'informations en question, ou ne leur sont pas aisément accessibles ;
- b) elles ont une valeur commerciale parce qu'elles sont secrètes ;
- c) elles ont fait l'objet, de la part de la personne qui en a le contrôle de façon licite, de dispositions raisonnables, compte tenu des circonstances, destinées à les garder secrètes

Cette définition est à même d'englober un algorithme tenu secret par l'entreprise. Pour la suite nous nous fixerons dans le cadre du droit européen.

2.2 Protection par droit d'auteur

Exigences du droit d'auteur Pour constituer une "œuvre", les algorithmes doivent

- *Être originaux en étant la création intellectuelle originale de l'auteur*. Dans sa jurisprudence, la CJUE (Cour de Justice de l'Union Européenne) a précisé que cela peut se manifester par des reflets de la personnalité de l'auteur, des choix créatifs ou des séquences et combinaisons dans lesquelles les auteurs s'expriment de manière originale. De plus, la CJUE a jugé qu'un objet n'est pas original lorsque sa réalisation a été dictée par des considérations techniques, des règles ou d'autres contraintes qui n'ont laissé aucune place à la liberté créatrice. En revanche, les œuvres qui sont en partie nécessaire à l'obtention d'un résultat technique peuvent bénéficier d'un droit d'auteur, à condition que l'œuvre soit originale et qu'elle soit l'expression d'une création.
- *Constituer une expression de cette création*. Il n'est pas nécessaire que l'expression soit permanente, mais elle doit être identifiable, objective et ne pas être dictée par une fonctionnalité technique.

L'article 2 de la Convention de Berne est plus ouvert et accorde une protection aux œuvres, qu'elles soient scientifiques ou artistiques, « quel que soit le mode ou la forme de leur expression ». Cet article pourrait donc ouvrir une porte pour englober les algorithmes, bien que la CJUE ait apparemment fermé cette porte dans ses décisions.

Respect des exigences pour les algorithmes Le respect de ces deux exigences dépend du type d'algorithme. Il est peu probable que les algorithmes et modèles d'apprentissage non supervisé ou semi-supervisé soient considérés comme une création intellectuelle propre à l'auteur, car le développeur joue un rôle mineur dans le fonctionnement des algorithmes. En effet, une œuvre ne peut pas être considérée comme la propre création intellectuelle de l'auteur si celui-ci ne la crée pas réellement. En outre, comme ils sont développés sur des données non étiquetées, le développeur n'assemble pas les données d'une manière originale reflétant un choix créatif. Les algorithmes d'apprentissage supervisé ont également du mal à satisfaire à ces deux exigences, car ils sont souvent basés sur de l'open source, et donc non originaux, avec des expressions essentiellement techniques. Par exemple, les algorithmes de régression linéaire et les algorithmes de régression logistique relient des ensembles d'entrées à une seule variable de sortie pour révéler des relations linéaires sous forme de graphique. Étant donné les aspects techniques de ces expressions et leur manque d'originalité, ces algorithmes ne relèvent pas du droit d'auteur.

En ce qui concerne les modèles algorithmiques, les plus courants sont : les modèles basés sur la régression ; les modèles basés sur la classification ; les arbres de décision ; les réseaux neuronaux ; les machines à vecteurs de support et les systèmes experts. Comme les réseaux de neurones contiennent des couches cachées au développeur, exécutent des tâches sans programmation préalable et limitent le contrôle humain sur ses corrélations, ils entrent en conflit avec le critère de création intellectuelle propre à l'auteur. En revanche, les modèles basés sur la régression et la classification (pouvant utiliser en sous-unité des réseaux de neurones), ainsi que les arbres de décision, comprennent des combinaisons

d'algorithmes développés sur des données étiquetées. Derrière les données étiquetées se cachent des choix d'agencement spécifiques. L'empilement d'algorithmes et l'assemblage de données pour développer un modèle complet adapté à un objectif spécifique reflètent également la liberté créative dans laquelle le développeur produit une création intellectuelle originale, satisfaisant ainsi sans doute à la première condition. Surmonter l'exigence d'expression présente un plus grand obstacle.

Les modèles de régression et de classification, ainsi que les arbres de décision, peuvent tous être exprimés de manière identifiable, comme par du texte brut ou des représentations graphiques. Leur expression est cependant dictée par une fonction technique. Mais, comme précisé auparavant l'obstacle de la fonction technique peut être surmonté si la dictée technique n'empêche pas l'auteur de refléter sa personnalité dans l'œuvre par des choix libres et créatifs. Cela peut, à première vue, représenter une ouverture pour ces modèles algorithmiques supervisés.

Ce que les tribunaux et les législateurs ne reconnaissent pas, c'est qu'un programmeur fait des choix libres et créatifs lorsqu'il écrit et superpose des algorithmes, et qu'il entraîne ces algorithmes sur des données choisies avec soin afin de créer une composition. Les choix que l'on fait à cet égard changent entièrement le modèle, et comme le programmeur les décide librement ils reflètent sa personnalité. C'est l'originalité du programmeur qui conduit à cette composition, et aucune dictée technique ne l'empêche de le faire en soi. Ces aspects devraient conférer la part de créativité nécessaire à l'application du droit d'auteur. Mais comme le droit applique strictement le critère de la fonction technique, et que les algorithmes contiennent manifestement de telles fonctions, ils ont moins de chances de bénéficier de la protection du droit d'auteur si l'on suit les décisions antérieures de la CJUE. En l'état actuel, le droit d'auteur ne protège que le code source, et non la configuration sur laquelle le code source est basé, c'est-à-dire les algorithmes, les modèles algorithmiques, leurs données et leur formation, ce qui aboutit au résultat particulier de protéger la couverture du livre, mais pas son contenu.

Protection des bases d'entraînement Les choix de données ne bénéficient pas non plus d'une protection individuelle par le droit d'auteur en vertu de la directive sur les bases de données. En effet, le droit d'auteur peut s'appliquer aux structures de données, mais si elles sont dictées par des considérations, des règles ou des contraintes, elles perdent leur éligibilité et ne satisfont donc probablement pas au seuil d'originalité.

La question de savoir si les données bénéficient d'une protection en vertu de la directive sur les bases de données, qui est comparable mais distincte du droit d'auteur, est discutable. L'exigence d'un investissement substantiel derrière l'obtention, et non la création, des données aurait amené les spécialistes à s'accorder sur le fait que les big data et les données liées à l'IA n'entrent pas dans son champ d'application. Mais l'article 7 accorde également une protection à la vérification ou à la présentation des données, à condition qu'il y ait eu un investissement substantiel, que ce soit sur le plan quantitatif ou qualitatif. La CJUE n'a pas encore statué sur ces deux options, mais la Commission européenne a récemment noté que la possibilité de protéger les big data et les données relatives à la génération de machines et à l'internet of things en tant que tels sont ouverte. Les données d'apprentissage, du moment qu'elles sont présentées ou vérifiées, et ont été choisies avec soin et analysées, peuvent donc plausiblement bénéficier d'une protection. Si tel est le cas, il est interdit aux tiers d'extraire des portions substantielles de données des ensembles de données des propriétaires, et donc d'entraîner des modèles avec ces données.

2.3 Protection par brevet

Exigences du brevet d'invention Les systèmes d'IA ont du mal à bénéficier d'une protection par brevet. Toutefois, comme nous le verrons, il est possible d'obtenir des brevets pour ces systèmes s'ils sont déposés en tant qu'inventions mises en œuvre par ordinateur (IMO).

Pour obtenir une protection par brevet, une invention doit satisfaire trois critères principaux (rapidement exposées dans 1.2 Introduction à la propriété intellectuelle). Premièrement, elle doit être nouvelle, ce qui implique que l'invention ne fasse pas partie de l'état de l'art, c'est-à-dire qu'elle ne soit pas disponible dans le monde entier avant le dépôt d'une demande de brevet. Deuxièmement, elle doit relever d'une activité inventive, ce qui signifie qu'elle n'est pas évidente pour un homme du métier. Troisièmement, elle doit avoir un caractère technique, démontré soit par la création d'un effet qui sert un objectif technique spécifique, soit par son adaptation à une mise en œuvre technique spécifique.

Avant d'aborder les algorithmes, il est important de noter que la question de savoir si les programmes d'ordinateur et les algorithmes sont brevetables reste un sujet controversé. L'article 27 de L'Accord de l'OMC sur les aspects des droits de propriété intellectuelle liés au commerce (Accord sur les ADPIC - l'accord multilatéral le plus complet concernant la propriété intellectuelle) autorise les brevets pour toutes les inventions, qu'il s'agisse de produits ou de procédés, dans tous les domaines technologiques, tandis que l'article 52 de la Convention sur la délivrance de brevet européen (CBE – ou convention de Munich), le cadre juridique des brevets en Europe, exclut de la brevetabilité les méthodes mathématiques et les programmes. Pourtant, l'Office européen des brevets (OEB) a délivré plus de 30 000 brevets de logiciels depuis 1978.

En ce qui concerne les algorithmes, les lignes directrices de l'office européen du brevet (OEB) réaffirment que les modèles de calcul et les algorithmes permettant l'IA et l'apprentissage automatique ne sont généralement pas brevetables. Cependant, si les revendications de brevet consistent en une méthode « impliquant l'utilisation de moyens techniques » - comme un ordinateur - un caractère technique est conféré à l'objet dans son ensemble, ce qui permet l'éligibilité au brevet. En raison des difficultés à breveter les programmes d'ordinateur et de la nécessité des moyens techniques matériels pour conférer un caractère technique aux modèles mathématiques, les IMO sont apparues comme un compromis. Elles couvrent les revendications impliquant « l'utilisation d'un ordinateur, d'un réseau informatique ou d'un autre appareil programmable et dont une ou plusieurs caractéristiques sont réalisées totalement ou en partie par un programme d'ordinateur ».

En tant que partie d'un programme d'ordinateur Malgré la plausibilité du brevetage des programmes d'ordinateur, les algorithmes ont peu de chances d'obtenir une protection par brevet en tant que partie du programme d'ordinateur auquel ils appartiennent. Par exemple, dans une affaire IBM de 1998, la Chambre de recours technique (CRT) de l'OEB a rejeté intégralement une demande de brevet car elle avait pour objet un programme d'ordinateur.

La CRT a traité peu d'affaires relatives aux programmes d'ordinateur, mais celles qui ont été jugées démontrent que la probabilité de protéger un programme d'ordinateur par un brevet dépend de son caractère technique. En conséquence, le type de programme et ses algorithmes sous-jacents déterminent la brevetabilité. Les algorithmes de ML pertinents pour cette étude sont construits pour améliorer la performance des opérations commerciales internes, ainsi que pour faciliter les applications sur Internet, et non pour améliorer la fonctionnalité interne d'un ordinateur ou produire un effet technique supplémentaire entre le matériel et le logiciel. Ces algorithmes ne satisfont donc pas à la norme technique envisagée par le CRT.

Les décisions de l'OEB ne sont pas strictement contraignantes pour les États contractants de la CBE, et les tribunaux nationaux français sont libres d'interpréter les dispositions de la CBE, sauf lorsqu'il s'agit du refus ou de la révocation d'un brevet européen. Certains Etats refusent cependant les brevets appliqués aux programmes d'ordinateur ce qui explique qu'ils soient assez souvent protégés en tant qu'IMO.

En tant qu'inventions mises en œuvre par ordinateur

Les conditions fondamentales de brevetabilité sont (d'après l'article 52 des Directives relatives à l'examen pratiqué à l'Office européen des brevets) :

- il doit y avoir "invention" dans un quelconque domaine technologique ;
- l'invention doit être "susceptible d'application industrielle" ;
- l'invention doit être "nouvelle" ;
- l'invention doit impliquer une "activité inventive"

Les algorithmes ont plus de chances d'être brevetés en tant qu'IMO, à condition que leurs revendications de méthode contiennent des étapes exécutables par ordinateur, ou réalisent une certaine fonctionnalité lorsqu'ils sont déployés. Ainsi, l'inclusion de modèles mathématiques dans les revendications de brevet ne scelle pas automatiquement leur inéligibilité au brevet.

Les algorithmes non supervisés ou par renforcement dans lesquels les humains jouent un rôle moins important, comme c'est souvent le cas avec les réseaux de neurones, peuvent être plus difficiles à protéger. En effet ils peuvent évoluer sans apport humain continu, ce qui les transforme en un système qui n'est pas entièrement créé par une personne physique et qui échappe pour le moment au cadre établi (Cf. L'IA en tant qu'inventeur ?). Cela dit, l'humain est responsable de la construction du réseau original et peut améliorer ses performances tout au long du processus de développement et après son déploiement. Sans cette participation, il n'y aurait pas d'invention, de sorte que le critère de l'invention n'exclut pas la brevetabilité des réseaux de neurones. Il en va de même pour les algorithmes supervisés, car ils nécessitent une participation humaine encore plus importante.

Néanmoins, même dans ce cas, le fait d'employer un algorithme dans un ordinateur pour accomplir des tâches n'est surement pas assez technique. Et même si l'algorithme remplit le critère rigoureux de considération technique, il lui manque souvent une activité inventive, et la nouveauté, car de nombreux algorithmes sont créés à partir d'open source. Par ailleurs les modèles algorithmiques peuvent tomber sous l'exception de modèle mathématique de l'article 52, ce qui les rend en soi non techniques et non brevetables. S'ils sont nouveaux et font partie de revendications de brevet comprenant des caractéristiques techniques qui contribuent à un effet technique sur l'invention, comme l'utilisation de structures de données à des fins fonctionnelles, ils sont brevetables. On voit donc toute la complexité du caractère brevetable d'un algorithme, même en tant qu'IMO, puisqu'il faut à la fois des revendications techniques et non techniques, pour correspondre au critère d'inventivité sans être uniquement concept mathématique. Finalement, la réussite d'une demande implique donc une incertitude juridique et des coûts importants, ce qui signifie qu'il y a peu d'incitation à essayer.

2.4 Des interactions nouvelles entre algorithmes et données

Mémorisation dans GPT-2 Il est devenu courant d'utiliser et de publier des modèles de langage de grande taille (plus d'un milliard de paramètres) qui ont été formés sur des ensembles de données privés. Dans une publication récente (Extracting Training Data from Large Language Models), des chercheurs de l'université de Berkeley (de Berkeley Artificial Intelligence Research plus précisément), ont montré qu'il était possible d'extraire des données d'entraînement en interrogeant le modèle de langage de la bonne façon.

La recherche porte sur GPT-2, un modèle de language publié par OpenAI. Pour donner un peu de contexte, OpenAI est une entreprise privée, connue pour avoir en partie été fondée par Elon Musk. L'objectif de cette société est de promouvoir et développer une intelligence artificielle à visage humain qui bénéficiera à toute l'humanité, et elle était à but non-lucratif jusqu'à Mars 2019. L'un de leur projet se nomme GPT-2. Il s'agit d'un modèle de traitement automatique du langage naturel, basé sur des méthodes de Machine Learning d'apprentissage non supervisé. Il a été entrainé sur des données d'internet, on peut donc le voir comme une IA apprenant d'elle-même en passant de site internet en site internet et qui essaye de représenter et donner un sens à ce qu'elle voit. Ce modèle est capable de compléter et générer des paragraphes entiers de texte ayant une cohérence syntaxique, grammaticale et

informative. Le modèle peut lire et comprendre un texte, le retranscrire, le résumer et est même capable de répondre à des questions concernant sa structure ou les informations qu'il contient. Tout ceci, et c'est bien là que réside l'exploit, sans entrainement préalable spécifique à chacune de ces tâches en particulier et sur n'importe quel sujet imaginable.

Revenons désormais à la recherche de BAIR. Les chercheurs ont constaté qu'au moins 0,1% de ses générations de texte (une estimation très prudente) contiennent de longues chaînes de texte qui sont "copiées-collées" d'un document de son ensemble d'apprentissage. Une telle mémorisation serait un problème évident pour les modèles de langage qui sont formés sur des données privées, par exemple, sur les courriels des utilisateurs, car le modèle pourrait produire par inadvertance les conversations sensibles d'un utilisateur. Cependant, même pour les modèles qui sont formés sur des données publiques du Web (par exemple, GPT-2, GPT-3, RoBERTa – les plus connus), la mémorisation des données de formation soulève de multiples questions réglementaires difficiles, allant de l'utilisation abusive d'informations personnellement identifiables à la violation des droits d'auteur.

Mémorisation de données à caractère privé et RGPD Par exemple, parmi les données mémorisées, en donnant le début du texte d'une adresse, le modèle peut autocompléter des données personnelles permettant d'identifier une personne. Environ 13 % des exemples mémorisés contiennent des noms ou des coordonnées (courriels, pseudos Twitter, numéros de téléphone, etc.) de personnes et d'entreprises. Bien qu'aucune de ces informations personnelles ne soit secrète (n'importe qui peut les trouver en ligne), leur inclusion dans un modèle linguistique pose tout de même de nombreux problèmes de confidentialité. En particulier, elle pourrait violer les législations relatives à la protection de la vie privée des utilisateurs, telles que le RGPD, décrit auparavant.

En effet, cette mémorisation s'oppose à presque tous les principes du cadre de collecte et d'utilisation des données décrits dans l'article 5.1-2 (cf. 1.2 Introduction à la propriété intellectuelle).

Par exemple :

- Exactitude –Dans un cas exposé par les chercheurs, le modèle a généré un reportage sur le meurtre de M. R. (un événement réel). Cependant, GPT-2 attribue à tort le meurtre à A. D., qui était en fait la victime d'un crime sans rapport avec celui-ci.
- Limitation du stockage – Les données utilisés en entraînement subsistent longtemps après leur utilisation, puisqu'elles sont extractables du modèle.
- Intégrité et confidentialité – Il devient très dur de garantir une sécurité, une intégrité et une confidentialité appropriées, puisque même des données à caractère personnel permettant l'identification d'une personne peuvent être extraites.

De plus, lorsque l'utilisateur a mis ses coordonnées en ligne, elles avaient un contexte d'utilisation prévu. Malheureusement, les applications construites au-dessus de GPT-2 ne connaissent pas ce contexte et peuvent donc partager involontairement les données de l'utilisateur d'une manière qu'il n'avait pas prévue. Par exemple, ses informations de contact peuvent être transmises par inadvertance par un chatbot du service clientèle. La mémorisation de données personnelles ne constitue probablement pas une "sécurité appropriée, et l'on peut faire valoir que l'inclusion implicite des données dans les sorties des systèmes en aval n'est pas compatible avec l'objectif initial de la collecte des données, à savoir la modélisation du langage générique.

Outre les violations de l'utilisation abusive des données, la présentation erronée des informations personnelles des personnes dans des contextes inappropriés touche également aux réglementations existantes en matière de protection de la vie privée contre la diffamation ou les délits de fausse publicité. De même, la déformation de noms d'entreprises ou de produits peut constituer une violation des lois sur les marques.

Finalement, les problèmes présentés ci-dessous pourraient contraindre les personnes à demander que leurs données soient supprimées du modèle. Elles pourraient le faire en invoquant les lois sur le « droit à l'effacement » du RGPD. Il existe une zone d'ombre juridique quant à la manière dont ces réglementations devraient s'appliquer aux modèles d'apprentissage automatique. Par exemple, les utilisateurs peuvent-ils demander que leurs données soient supprimées des données d'entraînement d'un modèle, sachant que le modèle a déjà été entraîné ? Cela voudrait dire réentraîner un nouveau modèle de zéro (procédé couteux et qui prend du temps …). Le fait que les modèles puissent mémoriser et utiliser à mauvais escient les informations personnelles d'un individu rend certainement plus convaincants les arguments en faveur de la suppression des données et du recyclage.

Mémorisation de données protégées par droit d'auteur Un autre type de contenu que le modèle mémorise sont les œuvres protégées par le droit d'auteur (pas uniquement textuelles, les images et musiques peuvent aussi être affectés lorsque des modèles de compréhension sur ces media seront au même niveau).

Un premier exemple provient non plus de GPT-2 mais de GPT-3, un modèle 100 fois plus grand que GPT-2 (il a été montré que les modèles de langage plus grands mémorisent davantage). Les chercheurs ont demandé à GPT-3 de compléter le chapitre 3 de Harry Potter et la pierre philosophale, à partir de la première phrase du chapitre. Le modèle reproduit correctement environ une page complète du livre (environ 240 mots) avant de commettre sa première erreur.

De la même façon, il est possible de récupérer du code source de projets sous licence, voir des fichiers entiers. L'outil GitHub copilot basé sur l'OpenAI Codex, est un outil permet de traduire du texte en code. C'est-à-dire, en décrivant la fonction que doit réaliser un code, l'outil génère le code permettant d'effectuer cette fonction. Ayant été entraîné sur des milliards de lignes de code public, l'une des premières questions qui a été soulevée concernant Copilot a porté sur les problèmes de droits d'auteur, en pointant spécifiquement l'idée de la licence GPL, qui exige que toutes les œuvres dérivées portent cette même licence. L'une des critiques les plus vives étaient que le projet, en quelque sorte, blanchissait des travaux open source en projet commercial (alors que OpenAI codex n'est pas open source). Est-ce que le fait d'entraîner sur les données font du modèle un dérivé susceptible d'être soumise à la licence GPL ?

Étant donné que les modèles de langage (et bientôt les modèles de vision !) mémorisent et régurgitent du contenu protégé par le droit d'auteur, cela signifie-t-il qu'ils constituent une violation du droit d'auteur ? La légalité de la formation de modèles sur des données protégées par le droit d'auteur est encore très débattue, avec des arguments en faveur et contre la caractérisation de l'apprentissage automatique comme « utilisation loyale » (fair use) du droit américain et exceptions légales du droit français (notion autrement plus étroite).

La question de la mémorisation des données a certainement un rôle à jouer dans ce débat. En effet, en réponse à une demande de commentaires de l'Office américain des brevets (les Américains sont les plus avancés sur les sujets liés à l'IA), de nombreuses parties plaident en faveur de la caractérisation de l'apprentissage automatique comme usage loyal, en partie parce que les modèles d'apprentissage automatique sont supposés ne pas émettre de données mémorisées. C'est par exemple une ligne de défense soutenue par OpenAI et l'Electronic Frontier Foundation. De la même manière, la rubrique Protecting Originality du site de GitHub copilot se protège par :

> *« Est-ce-que GitHub Copilot récite du code des données d'entraînement ? - GitHub Copilot est un synthétiseur de code, pas un moteur de recherche : la grande majorité du code qu'il suggère est généré de manière unique et n'a jamais été vu auparavant. Nous avons constaté qu'environ 0,1% du temps, la suggestion peut contenir des extraits qui sont textuellement issus de l'ensemble d'entraînement. […] La plupart de ces cas se produisent lorsque vous ne fournissez pas un contexte suffisant (en particulier, lors de l'édition d'un fichier vide), ou lorsqu'il existe une solution commune, voire universelle, au problème. Nous construisons un traqueur d'origine pour aider à détecter les rares cas de code répété à partir de*

l'ensemble d'entrainement, afin de vous aider à prendre de bonnes décisions en temps réel sur les suggestions de GitHub Copilot. »

Bien sûr, la défense de l'usage loyal par les parties précédemment mentionnées ne repose pas uniquement sur l'hypothèse que les modèles ne mémorisent pas leurs données d'apprentissage, mais cette dernière affaiblit la ligne d'argumentation. En fin de compte, la réponse à cette question pourrait dépendre de la manière dont les sorties d'un modèle de langage sont utilisées. Par exemple, la régurgitation d'une page de Harry Potter dans un autre livre constitue un cas beaucoup plus clair de violation du droit d'auteur que l'affichage fallacieux du même contenu par un système de traduction.

3. L'IA en tant qu'inventeur ?

La question épineuse de la mémorisation est étroitement corrélée à celle de savoir si une IA peut être considérée comme inventeur. En effet, l'IA est capable de générer des textes, des images, des programmes témoignant d'une créativité que l'on jugeait propre à l'Homme il y a moins d'une dizaine d'année. Mais est-ce vraiment de la créativité si elle n'est en fait que copieuse ? En s'écartant désormais du débat philosophique qui entoure la question, il reste du point de vue cadre légal de nombreuses zones d'ombres quant aux créations, et aux IA qui les génèrent.

3.1 IA créative

La créativité est l'une des notions majeures de la propriété intellectuelle. Sans créativité, il n'y a pas de droit d'auteur ni de brevet. Cette notion était déjà assez floue dans certains cas (e.g. distinction entre inspiration et plagiat) avant que l'IA n'apparaisse. L'apparition d'une « IA créative » redistribue à nouveau les cartes.

L'IA créative est une branche de l'intelligence artificielle dans laquelle l'IA peut créer des peintures, des films, de la musique, des jeux et bien d'autres arts créatifs comme le fait un humain. Tout comme les êtres humains apprennent à explorer et à créer, l'IA apprend progressivement à partir des données fournies au réseau de neurones. Voici quelques exemples.

Design DALL·E (mélange entre Dali, peintre espagnol du surréalisme et WALL·E robot du film d'animation Pixar du même nom) est une version de GPT-3 d'Open AI, sortie en 2021 et entraînée à générer des images à partir de descriptions textuelles, en utilisant un ensemble de paires texte-image. Elle possède un ensemble diversifié d'aptitudes, y compris la création de versions anthropomorphisées d'animaux et d'objets, la combinaison de concepts non liés de manière plausible, le rendu de texte et l'application de transformations à des images existantes. Par exemple, en saisissant en entrée du modèle le texte « une chaise en forme d'avocat », le modèle est capable de générer des images ultra-réalistes avec des relations créatives entre les idées entrantes.

Figure 3 : Génération par DALL·E d'un fauteuil en forme d'avocat - Source : https://openai.com/blog/dall-e/

Bande-annonce de film Des chercheurs d'IBM ont collaboré en 2016 avec la 20th Century Fox pour créer la toute première bande-annonce cognitive du film *Morgan*.

Art pictural Nous avons déjà vu le *Portrait d'Edmond de Belamy*, un tableau AI qui a été vendu pour 432 000 dollars. Son créateur est un réseau génératif adversatif (GAN), qui a été alimenté par un ensemble de données de 15 000 portraits couvrant six siècles pour alimenter son résultat. Une communauté d'artistes a depuis grandi autour de ces thèmes, utilisant la créativité de l'IA à des fins artistiques (voir https://aiartists.org/).

Musique La créativité de l'IA a également trouvé sa place dans l'industrie musicale. Plusieurs projets exploratoires ont été menés sur l'utilisation de l'IA pour composer de la musique. Jusqu'à présent, les progrès ont été remarquables, mais dans la plupart des cas, il reste un compositeur humain guidant ou supervisant l'IA.

L'un des projets les plus impressionnants dans ce domaine est *Flow Machine*, dirigé par François Pachet, directeur du Spotify Creator Technology Research Lab et développé en partie par Sony CSL. En 2016, Pachet a utilisé la Flow Machine pour créer "Daddy's Car", une chanson inspirée des œuvres des Beatles. L'IA a composé la mélodie de base. Les compositeurs humains ont ensuite complété l'œuvre avec des harmonies, une instrumentation et des paroles...

On peut aussi citer le projet AIVA capable de générer des musiques de fonds et d'ambiance pour des films par exemple.

Littérature Dans un autre domaine, des applications de NLP (Natural Language Processing) permettent de créer une écriture narrative à partir de données. Par exemple, l'agence américaine *Associated Press* utilise actuellement l'IA pour rédiger des milliers de rapports sportifs. De nouveaux sites web ont également vu le jour, générant un texte unique à partir d'un seul titre pour créer une ébauche complète d'un nouvel article.

En fournissant un outil créatif supplémentaire et très diversifié dans un certain nombre de secteurs, l'IA remet en question les possibilités de création et la créativité elle-même. Cependant, elle reste aujourd'hui « outil » et non indépendante : elle peut uniquement guider le créateur, lui laisser explorer. Bien qu'elle commence à développer des caractéristiques créatives, il lui manque le lien complexe entre imagination, pensée abstraite et mémoire qui constitue la créativité. (Pourtant, alors que les créateurs humains sont loués pour leur originalité, bon nombre de leurs œuvres les plus célèbres ont été influencées et imitées - utilisées pour susciter de nouveaux mouvements et de nouveaux modes de pensée.)

Jeu de go Au jeu de go, deux adversaires tentent d'occuper le plus d'espace sur un plateau quadrillé en plaçant alternativement des pions (pierres) noirs et blancs. La taille du tablier - 19 lignes sur 19 - offre un nombre incalculable de configurations possibles - davantage qu'il n'y a d'atomes dans l'univers. Ce qui signifie que l'intuition et la créativité sont essentielles pour gagner à très haut niveau. L'un des exemples les plus cités pour traiter de la créativité de l'IA est celui du 37e coup de la deuxième partie opposant Lee Sedol à AlphaGo (une IA développée par DeepMind). L'IA choisit de poser une pierre sur un endroit complètement inattendu. Tellement d'ailleurs, que les commentateurs de la partie, retransmise dans le monde entier, pensent l'erreur si grande que la partie est gagnée d'avance et que la machine (qui a dominé le champion lors de la première partie) n'est en fait pas encore au point. Lee Sedol, lui aussi, a du mal à croire ce qu'il voit et sa réaction est assez frappante. Ce coup aura une grande influence sur la fin de la partie, remportée par l'IA. (Un documentaire publié sur la chaine Youtube de DeepMind est disponible pour lecteur intéressé.)

La question qui émerge est de savoir si une création d'IA dispose de l'originalité nécessaire pour être l'objet d'un droit d'auteur, de savoir qui disposerait du droit d'auteur des générations d'IA.

3.2 Droits d'auteur sur les œuvres générées par IA

Attribuer le droit à l'IA ? Le droit d'auteur français et européen est-il applicable ou non aux œuvres générées par des systèmes d'IA autonomes ? Comme vu dans la section précédente, les créations de l'IA deviennent de plus en plus autonomes. Elles prennent leurs propres décisions et, dans certains cas, créent même indépendamment de toute intervention humaine directe, avec ce qui ressemble très fortement à de la créativité.

La loi sur le droit d'auteur est construite pour n'être accordée qu'aux êtres humains. Une histoire illustrant cela : en juillet 2011, un photographe britannique David Slater laisse son appareil photo sur un trépied, à la suite de quoi un singe s'octroie l'appareil et prend une série de selfies, ensuite publiés par le photographe. En 2015, une association de protections des droits des animaux (PETA) réclame que le droit d'auteur soit attribué au singe et intente un procès au photographe. En janvier 2016, le juge de première instance déboute PETA au motif que même si le singe avait pris les photos, il était impossible de donner suite au procès sachant que les animaux n'ont pas qualité pour agir en justice et ne peuvent donc pas engager des poursuites pour atteinte au droit d'auteur.

En outre, les œuvres sont principalement considérées comme n'atteignant l'originalité que si elles sont créées par un être humain (le selfie du singe fait-il preuve de créativité ?). Il y a donc des questions qui ne peuvent être négligées concernant la propriété, la paternité et l'originalité des œuvres créées par IA.

L'octroi de la protection du droit d'auteur aux œuvres générées par l'IA n'a jamais été interdit. Cependant, les législations de nombreux pays ne sont pas adaptées à ce scénario, car seule une œuvre créée par un humain peut être protégée par le droit d'auteur. Et, même si L'IA est capable de manière autonome de générer des idées et de produire de nouvelles formes d'expression, selon l'article 2.6 de la Convention de Berne, la protection doit être au profit de l'auteur. Cela signifie généralement que l'auteur doit être une personne physique et qu'une IA ne peut pas y prétendre.

Attribuer le droit à l'humain ? Comme précisé dans 2.2 Protection des algorithmes et programmes informatiques, une œuvre créée par un programme issu de l'apport créatif de son créateur devrait bénéficier de la protection du droit d'auteur, ce qui réglerait le problème. Cette proposition pose un problème : l'IA a évolué de telle sorte qu'elle peut, par un processus d'auto-supervision, produire des œuvres de façon autonome sans aucune intervention d'un être humain. Le créateur du programme informatique original n'est pas nécessaire à l'IA dans le processus de création. De plus, l'IA se développe elle-même et crée des œuvres d'une manière telle que lesdites œuvres et le programme informatique d'origine ne peuvent pas être considérés comme originaux en raison de leur lien très faible avec le créateur original.

Par conséquent, la condition d'originalité peut être la condition la plus problématique lorsque l'on souhaite accorder la protection du droit d'auteur à des créations d'IA. En effet, il est très peu probable qu'une IA puisse être considérée comme l'auteur d'une œuvre, car le processus créatif doit venir d'un être humain. De plus, l'œuvre doit refléter la personnalité de l'auteur, par des choix libres et créatifs, afin d'être considérée comme originale. Ceci est difficile à prouver étant donné que le lien entre le programmeur humain et l'IA n'est pas assez fort pour admettre que le programmeur dicte l'expression finale de l'œuvre. Le résultat généré par l'IA et le lien avec les choix libres et créatifs des créateurs humains sont donc, selon la législation européenne, trop distants et l'œuvre ne peut donc pas non plus être attribuée au programmeur humain.

L'IA comme outil de création À ce jour, l'IA ne peut être créée que par l'esprit humain, même si elle peut évoluer d'elle-même par des processus d'auto-apprentissage. C'est pourquoi l'IA peut être considérée comme un outil plutôt que comme un créateur autonome. En effet, il y a toujours un objectif économique ou artistique derrière la création de la machine.

L'IA est utilisée par un programmeur et/ou une entreprise, afin d'atteindre un certain produit final et/ou objectif prédit. Considérer l'IA comme un outil signifie que la paternité est attribuée au programmeur/utilisateur/propriétaire humain. En fin de compte, cela résout les problèmes mentionnés précédemment concernant le manque d'originalité et l'établissement de la relation entre l'humain et la création finale générée par l'IA. Dans ce cas particulier, les droits d'auteur peuvent être attribués au propriétaire, au programmeur original ou à l'utilisateur final de l'IA, si le produit final varie selon l'utilisateur du programme et s'il y a démonstration du travail et de la créativité de l'utilisateur

Cependant, on peut y faire deux objections. Premièrement celle que le droit d'auteur ne reviendrait qu'à l'utilisateur de l'IA, alors que le caractère créatif du travail provient en fait du créateur de l'IA et des données d'entraînement. Cette solution est donc entièrement arbitraire et ne fait finalement que contourner le problème. Deuxièmement, certains types d'IA - entièrement ou suffisamment – non supervisées ne peuvent plus être considérées comme des outils, car l'humain n'est pas suffisamment impliqué dans le processus créatif de la machine. En effet, même s'il guide l'outil, le procédé créatif est le fruit du hasard (par exemple l'entrée du GAN ayant servi à produire le portrait de Belamy est un bruit aléatoire). Ce caractère aléatoire est quelque chose qui ne peut être attribué au programmeur humain ou à l'artiste.

Conséquences du flou quant à la protection Les ambiguïtés concernant les œuvres générées par l'IA font qu'il est très difficile pour les créateurs d'évaluer la valeur de leur œuvre. Cette situation est évidemment problématique, car elle peut dissuader les artistes à utiliser ce type d'outil. Un nombre croissant de créateurs rencontre en effet des difficultés à savoir si leur œuvre peut être protégée ou non par le droit d'auteur et quelles sont les conditions de cette protection. Pourquoi s'investiraient-ils alors dans leur création ? En outre, les entreprises ne seraient pas non plus incitées à créer des algorithmes d'auto-apprentissage pour l'IA, car leur développement (très coûteux) n'assurerait pas que les futures créations de l'IA soient protégées par le droit d'auteur.

Par exemple, l'IA se développe beaucoup dans le secteur de la mode : le centre de data science de LVMH est très développé et des collaborations avec les plus grands centres d'IA et de vision par ordinateurs se développent. De plus en plus d'approches algorithmiques sont utilisées pour leurs sites web et lors de la création de nouvelles collections de mode (imaginer le DALL·E de la mode et les conséquences économiques d'une telle création !). Cependant, les limites entre l'apport de l'IA et les choix humains impliqués sont de plus en plus minces, ce qui rend de plus en plus difficile pour les entreprises d'avoir la certitude que leurs œuvres seront protégées par le droit d'auteur.

Conclusion

Selon la législation française et européenne actuelle, les systèmes d'IA (de plus en plus autonomes) sont difficilement protégeables par quelque procédé que ce soit. En effet les critères d'expression de la créativité du créateur et d'originalité nécessaires à l'attribution d'un droit d'auteur ou d'un brevet sont régulièrement présentés. De même les œuvres générées par IA ne sont pas suffisamment originales et le lien entre l'IA et son programmeur humain est trop faible pour qu'elles soient considérées comme protégées par le droit d'auteur. En outre, l'IA n'est pas considérée comme une entité juridique, elle ne peut pas demander la propriété de l'œuvre qu'elle a créée, ni être son propre créateur.

Actuellement, il semble donc qu'une IA et ce qu'elle génère ne soit que difficilement protégeable et rentre dans le domaine public dès sa conception. Cela peut poser un problème, car les entreprises ne seront pas incitées à créer des modèles basés sur l'IA, sachant que ni le modèle et ni les œuvres générées par l'IA ne seront réellement protégeables. Cela pourrait ralentir les évolutions technologiques et avoir un impact sur des secteurs tels que les start-ups et les entreprises technologies et pharmaceutiques, et donc sur les consommateurs.

Enfin, un autre enjeu clé en matière de propriété intellectuelle repose sur l'utilisation de données potentiellement protégées par les algorithmes dans leur processus d'apprentissage, telles que des images ou textes. À titre d'exemple, le RGPD, l'une des réglementations dites les plus strictes en matière de protection des données à caractère personnel, ignore complétement certains aspects de l'apprentissage autonome comme celui de la mémorisation des données.

C'est un fait : la propriété intellectuelle et industrielle telle que nous les connaissons sont inappropriées aux nouvelles évolutions de l'IA et pourrait être à la fois un frein à la création de valeur et un enjeu majeur de protection des données dans les années à suivre, en l'absence d'adaptation des lois en vigueur.