

Dossier technique

Master 2 droit de l'économie numérique

Année universitaire 2024-2025

« L'innovation en intelligence artificielle peut-elle se développer sur une réglementation éthique et de transparence algorithmique ? »

Théo BARTZEN

Sous la direction de Aurélie KLEIN

21 décembre 2024

Table des matières

ABSTRACT FRANÇAIS	6
ABSTRACT ANGLAIS	7
INTRODUCTION	8
CHAPITRE 1 : APPRÉHENSION TECHNIQUE DE L'INTELLIGENCE ARTIFICIELLE	9
Section 1 : Les origines de l'intelligence artificielle	9
Section 2 : Fonctionnement technique de l'intelligence artificielle	12
Section 3 : Reproduction du système neuronal biologique : le réseau neuronal convolutif	13
Section 4 : Les différents types d'apprentissage en intelligence artificielle	19
CHAPITRE 2 : RÉFLEXIONS GÉNÉRALES SUR L'INTELLIGENCE ARTIFICIELLE	25
Section 1 : L'intelligence artificielle et l'intelligence humaine	26
Section 2 : L'intelligence artificielle une chance ou un danger ?	32
Section 3 : La comparaison entre l'émergence d'Internet et de l'Intelligence artificielle	35
CHAPITRE 3 : L'INNOVATION RESPONSABLE À TRAVERS LE RÈGLEMENT EUROPÉEN SUR L'INTELLIGENCE ARTIFICIELLE	37
Section 1 : Le RIA et le développement d'une intelligence artificielle responsable	38
Section 2 : Le RIA et l'intégration de la nécessité de l'innovation	43
Section 3 : La quête d'un équilibre paradoxale entre innovation et réglementation	45
CHAPITRE 4 : LE DÉVELOPPEMENT D'UNE GOUVERNANCE DES DONNÉES RESPONSABLE DES « SYSTÈMES D'INTELLIGENCE ARTIFICIELLE »	48
Section 1 : Les données synthétiques : une conciliation parfaite entre innovation et protection des données	50
Section 2 : Le talon d'Achille de l'IA : exploration des biais algorithmiques dans l'intelligence artificielle	58
Section 3 : Le cheminement complexe de la transparence algorithmique	64
CONCLUSION	70

GLOSSAIRE FRANÇAIS

TERMES	DÉFINITION
Algorithme	Suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée
Apprentissage auto-supervisé	Modèle extrait de l'information à partir de données non étiquetées, en créant ses propres tâches de supervision
Apprentissage automatique	Procédé par lequel les informations pertinentes sont tirées d'un ensemble de données d'entraînement
Apprentissage non supervisé	Procédé d'apprentissage automatique dans lequel l'algorithme utilise un jeu de données brutes
Apprentissage par renforcement	Système autonome qui apprend les actions à réaliser à partir d'expériences, en maximisant la somme des récompenses au cours du temps.
Apprentissage profond	Apprentissage automatique utilisant des réseaux de neurones
Couches de neurones	Organisation des neurones dans un réseau dont uniquement les couches sont interconnectées
Donnée brute	Donnée n'ayant subi aucune transformation
Donnée de sortie	Valeur représentant tout ou partie de l'opération effectuée par le système d'IA
Donnée d'entrée	Donnée utilisée pour l'apprentissage automatique
Explicabilité	Capacité de rendre compréhensible les éléments pris en compte par le système
Robustesse	Résilience d'un système d'IA à maintenir ses exigences de performances et de sécurité avec des données d'entrée extérieures
Surapprentissage	Modèle qui correspond trop précisément à une collection spécifique de données d'entraînement
RIA	Règlement sur l'intelligence artificielle

GLOSSAIRE ANGLAIS

TERMS	DEFINITION
Algorithm	Sequence of steps to produce a result from input elements
Self-supervised learning	Model extracting information from unlabeled data by creating its own supervision tasks
Machine learning	Process by which relevant information is extracted from a training dataset
Unsupervised learning	Machine learning process where the algorithm uses a set of raw data
Reinforcement learning	Autonomous system that learns actions to perform from experiences, maximizing cumulative rewards over time
Deep learning	Machine learning using neural networks
Neural layers	Organization of neurons in a network where only the layers are interconnected
Raw data	Data that has not undergone any transformation
Output data	Value representing all or part of the operation performed by the AI system
Input data	Data used for machine learning
Explainability	Ability to make the elements considered by the system understandable
Robustness	Resilience of an AI system to maintain its performance and safety requirements with external input data
Overfitting	Model that fits too precisely to a specific collection of training data
RIA	Artificial Intelligence Act (AI Act)

AVANT-PROPOS

La rédaction de ce dossier technique a été motivé par la volonté de mieux cerner, tant sur l'aspect juridique que technologique, les contours multiples de l'intelligence artificielle. Il constitue un guide complet et synthétique des enjeux soulevés par l'IA.

Le dossier technique a pour objectif d'offrir une approche la plus complète possible sur tous les aspects (juridiques, techniques, sociétaux, philosophiques et éthiques) pouvant se rattacher à l'intelligence artificielle.

La rédaction de ce dossier technique est subdivisé en 4 phases distinctes :

- **Analyse de l'aspect purement technique de l'IA**
- **Analyse des diverses pistes de réflexions sur l'IA**
- **Analyse de la réglementation européenne sur l'IA**
- **Analyse des solutions techniques et juridiques vers une IA éthique**

La rédaction du dossier technique n'a pas été reprise par l'utilisation de l'intelligence artificielle, excepté pour les parties suivantes : « **Introduction** » et « **Abstract Anglais** ».

La lecture du dossier technique doit être faite en suivant l'ordre établi au sein de la « **Table de matières** ».

La police utilisée est « **Georgia** » en taille **12** et avec un espacement des paragraphes de **1,5**.

J'espère que la lecture de ce dossier technique sera agréable et permettra à son lecteur de nouvelles pistes de réflexions, notamment sur l'articulation technico-juridique nécessaire s'agissant de la réglementation par le droit de l'intelligence artificielle.

ABSTRACT FRANÇAIS

Certains diront que l'intelligence artificielle n'est pas une innovation technologique, qu'elle reprend la vieille idée que le calcul puisse remplacer le raisonnement et que la connaissance suffise sans compréhension.

Cette critique explicite le contexte dans lequel s'inscrit l'intelligence artificielle. Les technologies du numérique font craindre à la société le dépassement de la machine face à l'humain. Les peurs à ce sujet sont souvent irrationnelles et font état d'un cruel manque d'éducation sur le sujet.

Cette crainte justifie donc en partie la volonté de l'Europe à y apporter une réponse adéquate en édifiant un cadre réglementaire européen à l'égard de la mise sur le marché des systèmes d'intelligence artificielle.

La volonté de réguler par le droit une technologie est un réel défi et questionne aussi quant à l'efficacité du droit à pouvoir appréhender des technologies aux spécificités techniques complexes tel que l'intelligence artificielle

La difficulté est encore plus importante lorsque le secteur à réguler est nouveau et qu'il représente une part de marché conséquente mais difficile à estimer. Les estimations du marché mondial de l'IA fluctuent de 500 milliards à plus de 1000 milliards selon certaines sources.

Dès lors, la problématique est celle d'édicter une réglementation européenne protectrice des droits fondamentaux des citoyens européens tout en stimulant les possibilités d'innovation des acteurs privés et publics.

Le RIA (Règlement sur l'Intelligence artificielle) réussit en partie ce défi en arborant une volonté de légiférer sur ce qui existe et sur ce qui est en devenir.

Malgré tout, le droit ne demeure l'unique solution. Il est essentiel de connaître le fonctionnement purement technique de l'intelligence artificielle. Cette formalité évitera notamment l'élaboration de normes juridico-techniques incohérentes.

ABSTRACT ANGLAIS

Some might argue that artificial intelligence is not a technological innovation, as it builds upon the old idea that computation could replace reasoning and that knowledge could suffice without understanding.

This critique highlights the context in which artificial intelligence operates. Digital technologies often evoke societal fears of machines surpassing human capabilities. These concerns are frequently irrational and reflect a significant lack of education on the topic.

This apprehension partly explains Europe's determination to provide an adequate response by establishing a regulatory framework for the market deployment of artificial intelligence systems.

Regulating a technology through law is a genuine challenge and raises questions about the law's capacity to address technologies with such complex technical specificities, like artificial intelligence.

The challenge is even greater when the sector to be regulated is both new and represents a significant yet difficult-to-estimate market share. Estimates of the AI market, for instance, vary widely, ranging from 500 billion to over 1,000 billion, depending on the source.

Thus, the issue lies in drafting European regulations that protect the fundamental rights of European citizens while fostering innovation opportunities for both private and public actors.

The AI Act (Artificial Intelligence Regulation) partially achieves this goal by demonstrating a commitment to legislating both existing systems and those yet to emerge.

Nonetheless, the law alone is not a comprehensive solution. It is essential to first understand the technical functioning of artificial intelligence. Such an understanding will help avoid drafting legally inconsistent standards.

INTRODUCTION

L'essor de l'intelligence artificielle marque l'une des évolutions technologiques les plus significatives de notre époque. Longtemps confinée au domaine de la science-fiction, l'IA est aujourd'hui une réalité qui redessine les contours de nos sociétés, de nos économies, et de nos cadres juridiques. Mais qu'est-ce que l'intelligence artificielle exactement ? Comment fonctionne-t-elle ? Et comment peut-elle être réglementée tout en stimulant l'innovation ? Ces questions sont au cœur de ce travail.

Dans un premier temps, il est nécessaire de s'intéresser à la dimension technique de l'IA pour mieux comprendre ses fondements et son fonctionnement. Le chapitre initial se penche sur les origines de cette technologie, les différents types d'apprentissage qu'elle mobilise, ainsi que la façon dont elle s'inspire du fonctionnement neuronal humain à travers des modèles comme les réseaux neuronaux convolutifs.

Ensuite, une analyse plus philosophique et sociétale s'impose. Le deuxième chapitre explore les relations entre l'intelligence artificielle et l'intelligence humaine, en questionnant les opportunités et les risques que cette technologie peut engendrer. Une comparaison entre l'émergence d'Internet et celle de l'IA met en lumière des parallèles intéressants et invite à une réflexion critique sur les enseignements que l'on peut tirer du passé.

Dans cette dynamique, la troisième partie de ce travail se concentre sur l'IA Act, un cadre réglementaire européen qui cherche à promouvoir une innovation responsable. Ce chapitre analyse comment ce texte vise à concilier la nécessité de stimuler le progrès technologique avec celle de protéger les droits fondamentaux, tout en soulignant les tensions qui peuvent naître entre innovation et réglementation.

Enfin, le quatrième chapitre se tourne vers une question essentielle : celle de la gestion des données dans le cadre des systèmes d'IA. Il explore notamment le rôle des données synthétiques, perçues comme une solution idéale pour réconcilier innovation et protection des données personnelles, tout en mettant en évidence les biais algorithmiques qui restent le talon d'Achille de l'IA. Enfin, il s'attarde sur les enjeux de la transparence algorithmique, un sujet complexe et encore largement débattu.

Ce mémoire vise à fournir une compréhension globale de l'IA et des défis émergents.

CHAPITRE 1 : APPRÉHENSION TECHNIQUE DE L'INTELLIGENCE ARTIFICIELLE

Section 1 : Les origines de l'intelligence artificielle

Pamela McCorduck¹ raconte dans l'un de ses livres que l'histoire de l'intelligence artificielle commence avec « *le vieux souhait de jouer à Dieu* ».

Depuis bien longtemps, l'homme cherche constamment à créer des automates donnant l'illusion de s'apparenter à l'humain.

L'intelligence artificielle trouve sa dénomination depuis la conférence de Dartmouth de 1956, organisé par Marvin Minsky et John McCarthy.

La logique a été la première des choses envisagées par les intellectuels de l'époque. Certains ont notamment montré la voie avec le programme « Logic Theorist » qui démontre des théorèmes simples de mathématiques.

Les recherches sur l'IA se sont ensuite atténuées lorsque l'agence américaine du Département de la Défense (l'ARPA) coupe les budgets de recherche en IA en 1970, ce qu'on appellera : « l'hiver de l'IA ».

La commercialisation des systèmes experts, à la mode des années 1960 à 1990, a été une tâche particulièrement ardue. MYCIN incarne cet échec. Il s'agissait d'un moteur d'inférence à chaînage arrière qui appliquait des règles aux faits et en déduisait d'autres. Concrètement, MYCIN était chargé de diagnostiquer des méningites, des infections aiguës ou des maladies du sang. En réalité, le fait que des ingénieurs et des médecins aient collaboré pour diagnostiquer les maladies ne fut pas concluant.

Le saut technologique de l'IA a surtout été permis lorsque les scientifiques ont tenté de retranscrire les principes biologiques de nos neurones aux mathématiques.

¹ Autrice américaine qui a écrit plusieurs livres sur l'histoire et la signification philosophique de l'intelligence artificielle.

Si le perceptron² de Frank Rosenblatt des années 1950 ne contenait qu'une seule couche, les neurones profonds de Yoshua Bengio, Yann Le Cun et Geoffrey Hinton en contiendront au moins trois.

En réalité, les réseaux de neurones profonds ont été la découverte technologique ayant permis à l'IA d'atteindre les performances que l'on connaît aujourd'hui.

L'histoire de l'IA est semée d'échecs, de réussites partielles comme de réussites fabuleuses. De conception plutôt simpliste à son départ avec l'approche symbolique (les systèmes experts), l'IA s'est progressivement améliorée avec les réseaux de neurones profonds.

En 2006, Microsoft Research effectue les premières expériences d'entraîner les réseaux de neurones avec l'utilisation des GPU (cartes graphiques).

En 2010, on observe les premiers résultats de l'apprentissage profond avec des bases de données plus importantes permettant l'application des grands réseaux de neurones profonds.

En 2012, l'IA prend un nouveau tournant avec l'efficacité des réseaux convolutifs.

En 2016, Lee Sedol champion de go sud-coréen s'incline contre AlphaGo construit par DeepMind. Il perd 4 parties sur cinq contre la machine. Comment la machine a-t-elle fait ? Elle a intégré les techniques de réseaux convolutifs, l'apprentissage par renforcement et le « Monte Carlo Tree Search³ ».

Claude Shannon, un des pères de l'IA, a établi ce que l'on appelle le nombre de Shannon de : 10^{120} . C'est une estimation du nombre de parties différentes possibles ayant un « sens échiquéen ». Ce nombre passe à 10^{40} si l'on prend en compte uniquement les parties « plausibles » qui excluent les coups absurdes.

² Première machine apprenante inspirée par la théorie cognitive de Donald Hebb (psychologue/neurobiologiste canadien)

³ Méthode de recherche arborescente aléatoire basé sur un algorithme de recherche heuristique

A titre de comparaison, le jeu de go malgré qu'il soit d'une moindre complexité, comporte entre 10^{600} et 10^{800} possibilités. Cet exemple illustre la capacité d'une « IA faible⁴ » à surpasser dans un domaine bien spécifique l'humain.

N'oublions pas que Lee Sedol a su qu'il avait été battu par DeepMind. A contrario, DeepMind n'a jamais su qu'il avait battu Lee Sedol à quatre reprises.

Cette amélioration technologique s'est donc tout naturellement accompagnée d'une certaine complexification.

Une complexification qu'il s'agira de nécessairement vulgariser afin de mieux comprendre et appréhender cette technologie sous le prisme du droit. Comment légiférer sur l'intelligence artificielle si l'on ne comprend pas la « langue des probables »⁵ ?

A titre d'exemple, une proposition de loi du 12 septembre 2023⁶ suggérait que soit mentionné le nom des auteurs lorsque leur contenu a été utilisé par les systèmes d'IA génératives.

Les législateurs ont oublié que la technicité de ces systèmes est souvent bien trop complexe pour assurer cette obligation. Il est très difficile d'établir la proportion exacte que le contenu de tel auteur a été utilisé pour générer tel résultat. L'entraînement de ces systèmes, basé sur les réseaux de neurones ne permet pas de « *maintenir le lien avec le matériel d'origine* »⁷.

Dès lors, il s'agira d'exposer une vue à 360 degrés du fonctionnement technique des systèmes d'intelligence artificielle. (**Section 2**)

⁴ Qualification donnée à une IA compétente dans un domaine extrêmement restreint (ex : échecs)

⁵ Yannick Meneceur : « *IA générative et professionnels du droit – Comprendre et s'appropriier la langue des probables* » LexisNexis 2024

⁶ https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi

⁷ Yannick Meneceur : « *IA générative et professionnels du droit – Comprendre et s'appropriier la langue des probables* » LexisNexis 2024

Section 2 : Fonctionnement technique de l'intelligence artificielle

L'intelligence artificielle est une technologie qui possède ses propres logiques de fonctionnement. Elles sont essentiellement issues des mathématiques.

Avant tout raisonnement juridique, il faut pouvoir établir simplement ce qu'est l'intelligence artificielle.

Pourquoi est-il essentiel de connaître les tenants et les aboutissants techniques d'une technologie que l'on souhaite réguler ? Définir les caractéristiques intrinsèques de cette technologie est primordial pour avoir une bonne compréhension des enjeux juridiques de l'intelligence artificielle.

L'utilisation de l'intelligence artificielle, que l'on veuille ou non, sera omniprésente. Il est important de connaître ce que l'on utilise. Tout comme une voiture ne se conduit pas sans permis, l'intelligence artificielle ne s'utilise pas sans à minima quelques connaissances techniques

Il est alors indispensable, pour le régulateur, de bien saisir comment ces modèles apprennent et traitent les données, afin de pouvoir définir des règles pour prévenir les discriminations (biais algorithmiques) garantir la transparence des décisions, et protéger les droits des utilisateurs.

Ainsi, cet exemple montre que l'on ne peut établir des règles juridiques efficaces sans une connaissance minimale des mécanismes techniques de l'IA, car ces mécanismes impactent directement les risques et enjeux juridiques, tels que la protection des données personnelles, l'équité et la non-discrimination.

Malgré une absence de consensus autour de la définition de l'intelligence artificielle, la Commission européenne a proposé une définition.

L'intelligence artificielle serait un *"ensemble de logiciels qui est développé au nom d'une ou plusieurs techniques et approches, qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit."*

Si l'on veut simplifier, l'intelligence artificielle est une technologie, spécifiquement le machine Learning (apprentissage automatique) qui représente statistiquement un environnement défini.

Première observation, « l'environnement défini » peut varier. Un système traitant des radiographies médicales ne va pas fonctionner de la même manière qu'un système de reconnaissance vocale. Les techniques d'apprentissage devront alors s'adapter à cet environnement. Ainsi, lorsque l'on parle d'intelligence artificielle, il faut différencier différents types d'apprentissage.

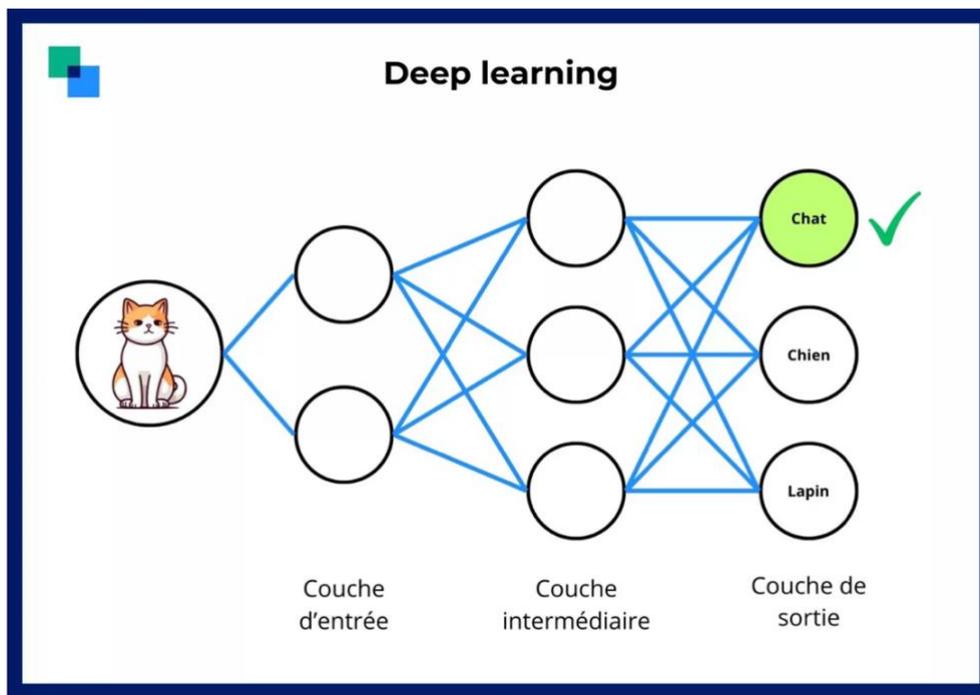
L'humain (Yann LeCun, Yoshua Bengio et Geoffrey Hinton notamment) s'est donc totalement inspiré du fonctionnement et de la structure du cortex visuel pour créer les CNN (Convolutional Neural Network). Ce type de réseau neuronal (**Section 3**) artificiel a notamment permis la reconnaissance d'images tel qu'on le connaît actuellement sur Internet.

Section 3 : Reproduction du système neuronal biologique : le réseau neuronal convolutif

L'intelligence artificielle fait appel à des principes mathématiques extrêmement complexe. En effet, les mathématiques ou la physique s'inspirent souvent de la biologie pour établir de nouveaux principes. Tout comme l'avion s'est inspiré de l'oiseau, l'intelligence artificielle s'est inspirée de notre cerveau et plus précisément de nos neurones.

Par exemple, pour l'identification d'images volumétriques tel que la vidéo, la parole, la musique ou le texte on fait appel au réseau neuronal convolutif.

Il a donc fallu adapter le principe de fonctionnement de nos neurones grâce à une articulation de règles et de principes mathématiques. On cite souvent l'inférence bayésienne (probabilité) et les neurones formels (représentation mathématique d'un neurone biologique).



Le Deep Learning trouve donc son nom du fait de la superposition importante de plusieurs couches de neurones interconnectées (jamais entre eux mais toujours à la couche suivante) tel que sur l'image ci-dessus.

L'organisation de ces couches peuvent varier. Un réseau de neurones convolutif est un type d'organisation bien spécifique de ces couches. L'apprentissage profond (deep learning) est l'entraînement de neurones multicouches.

Les couches d'entrée et de sortie sont dites visibles (on connaît leurs « valeurs ») tandis que les couches intermédiaires sont dites cachées. Déterminer les sorties des « couches cachées » est toute la difficulté du deep learning.

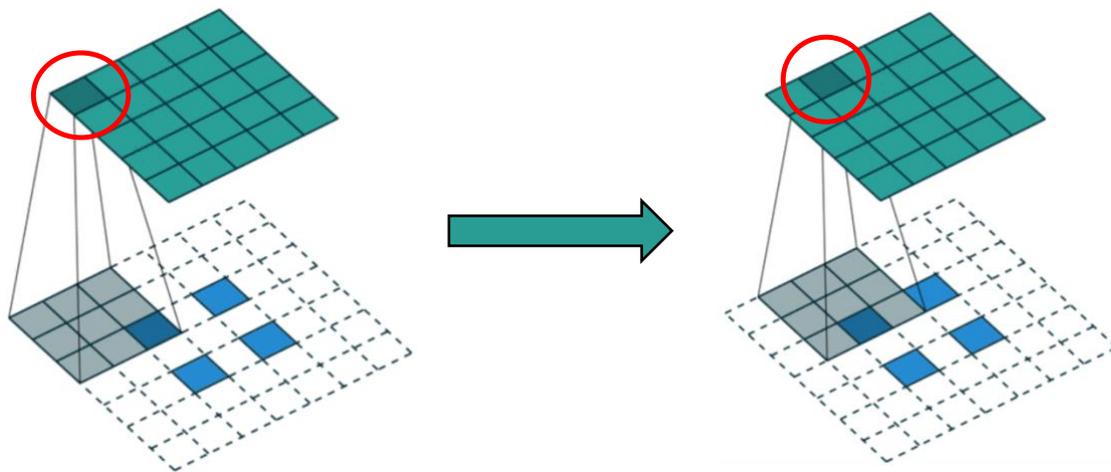
La convolution fonctionne en construisant un réseau de neurones avec une architecture spécifique. En lui présentant des images labellisées ou étiquetées (ex : image zèbre ou image lion) et à travers son entraînement le réseau va réussir à apprendre quelles caractéristiques sont les plus importantes pour classer des images de zèbres et de lions.

Pour cela il faut passer par plusieurs étapes que sont : les **filtres** (I) et le **pooling** (II). L'objectif n'est pas de rentrer dans les principes mathématiques qui sont bien trop complexes mais dans le principe de fonctionnement primaire du réseau neuronal convolutif.

Sous-section 1 : Les filtres : les yeux du réseau

Les filtres fonctionnent, au départ, purement aléatoirement. Au fur et à mesure, ils vont reconnaître selon ce qu'on leur demande, de reconnaître tel ou tel spécificité de l'image qui va permettre la distinction entre un zèbre et un lion (rayures zébrées versus crinière)

Voici comment cela fonctionne. Ce sont deux tableaux bidimensionnels, le blanc au-dessous et le vert émeraude au-dessus. Le tableau blanc représente l'image brute (notre zèbre 🦓) et le tableau vert représente la couche de réseau. Le filtre est le carré gris.



Son rôle va être de passer sur chaque pixel de l'image du zèbre (comme sur cet exemple, sa taille est de 3x3). Le filtre va passer plusieurs fois sur des pixels qui ont déjà été analysés. C'est ce processus qui permet de filtrer l'image afin de mieux la traiter.

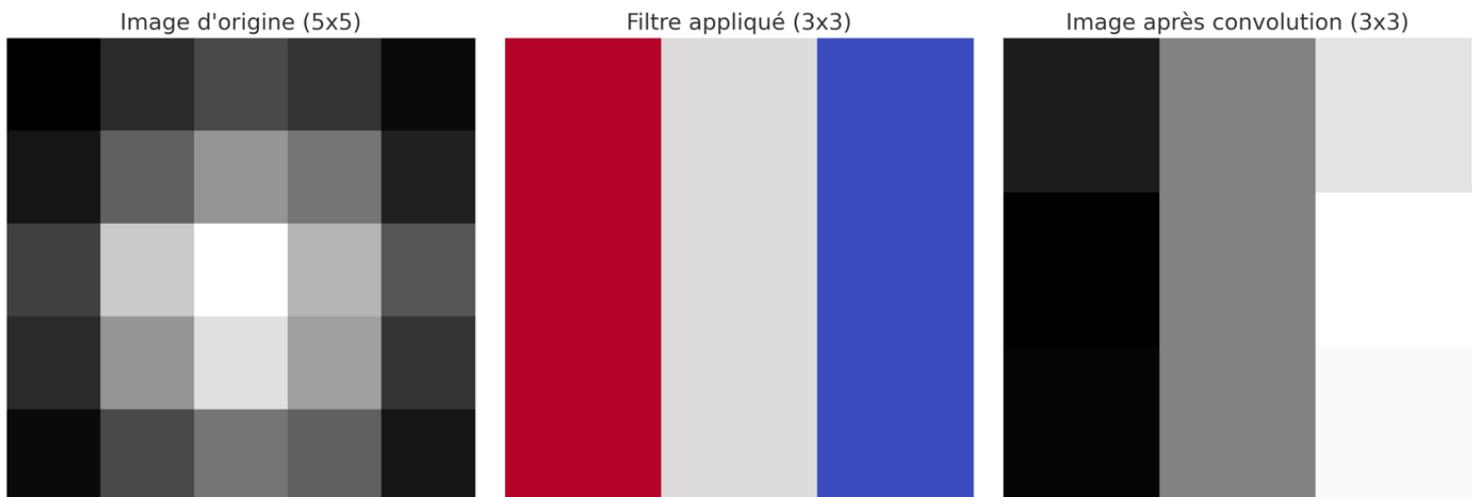
On transmet cette valeur à une couche suivante du réseau qui effectuera le même processus. Le cerveau humain fonctionne de la même façon.

Pour chaque filtre, l'image ou la valeur transmise correspond à une transformation de l'image d'entrée et où certaines formes spécifiques auront été mises en évidence.

Le pooling est une couche de réduction. La case 3x3 de l'image du zèbre (tableau blanc) devient une case 1x1 (il devient le pixel en vert foncé sur l'image). On obtient donc le même nombre de « feature maps⁸ » en entrée. Celles-ci sont simplement réduites.

⁸ Données de l'image qui rapportent aux contours, formes ou textures.

Voici un exemple :



- **L'image d'origine (5x5)** : est le pixel d'un zèbre. Pour chaque carré, on peut estimer « l'intensité » du gris entre 0 et 255. Le 0 correspond au noir et 255 correspond au blanc.
- **Le filtre appliqué 3x3** : agit comme si l'on regardait dans des anciennes lunettes « Anaglyphe » de couleur rouge et bleu devant un film. Avec ces lunettes on distinguerait mieux certaines couleurs que d'autres. Le filtre ou « kernel » fonctionne de la même manière en filtrant les couleurs qu'il souhaite distinguer.

Dans cet exemple, le filtre est utilisé pour détecter des contours verticaux. Ce filtre contient des « poids prédéfinis » qui permettent de renforcer le contraste en distinguant les transitions claires/sombres.

- **L'image après convolution (3x3)** : fait appliquer le filtre aux sous-régions de l'image comme vu précédemment. On multiplie pour chaque position les valeurs des pixels (0-255) avec les poids du filtre. Puis on fait la somme pour obtenir une valeur dans l'image de sortie

L'image convolué contient les informations que l'on a voulu extraire avec le filtre et seront réutilisées pour alimenter les couches suivantes du réseau ! On voit très bien que le résultat est une sorte de mélange des couleurs.

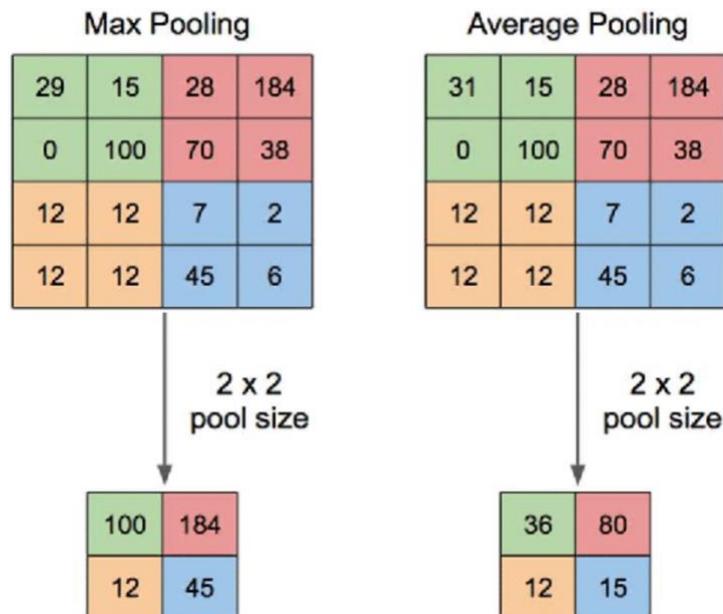
Sous-section 2 : Le « pooling » : réduire l'information pour mieux la traiter

Par-la-suite, intervient le Pooling qui consiste à réduire la taille d'une image en mappant des pixels. Cela permet entre autres de diminuer le nombre de données à traitées par le système neuronal sans pour autant perdre en pertinence d'analyse.

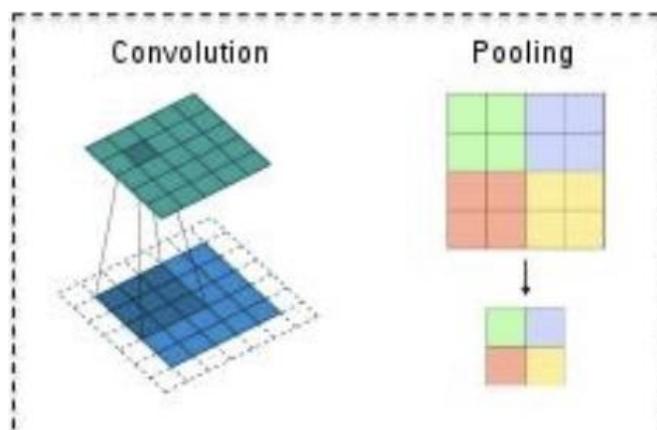
Il existe deux principaux types de couches de regroupement

Le **Max-pooling** : retiendra la valeur maximale dans une petite région (ex : carré de 2x2)

L'**Average-pooling** : calculera la moyenne des valeurs dans chacune des régions.



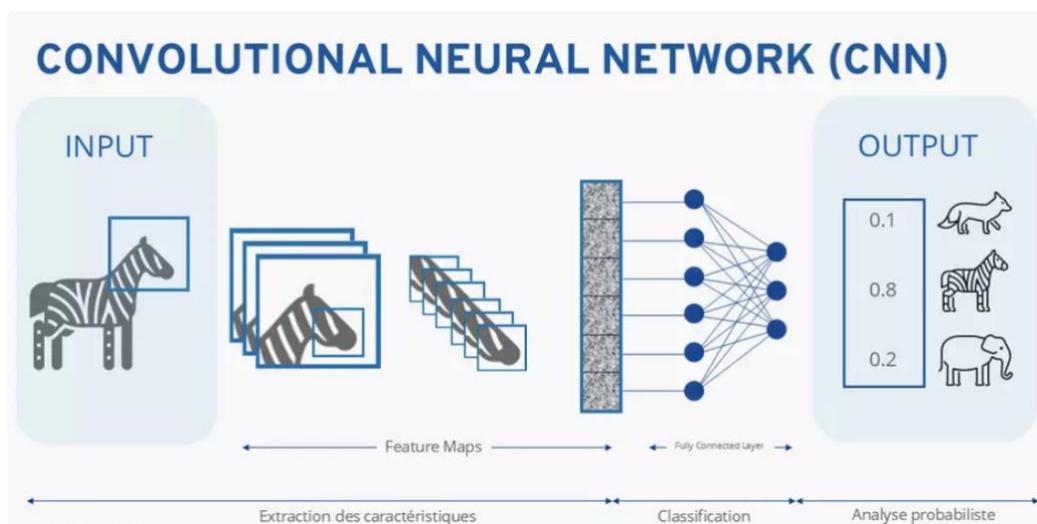
Schématisation du « pooling » en comparaison à la convolution :



Ce processus préserve les caractéristiques importantes mais réduit la taille de l'image. On passe d'un tableau 4x4 pixels (16 pixels) à un tableau 2x2 (4pixels). La donnée résultante est donc identique à la première mais celle-ci est moins volumineuse.

Le pooling n'est pas qu'une technique de réduction dimensionnelle. Il permet aussi de rendre le modèle plus robuste aux petites variations et aux déplacements des objets dans l'image.

Un zèbre dans différentes poses pourra alors être reconnu.



L'objectif de la convolution (filtre et pooling) est d'agir comme un filtre qui analysera chaque partie de l'image pour en extraire des caractéristiques spécifiques. Ce processus permet de se concentrer sur des aspects différents de l'image. On arrive ainsi à reconnaître des motifs indépendamment de leur position dans l'image.

Par-la-suite, l'image réduite est aplatie en un vecteur unidimensionnel (exemple pour un carré de 4 pixels : [200, 180, 150, 120]).

Ce vecteur passe dans les différentes couches où il est transformé en probabilités pour différentes classes. Ex : Classe Zèbre : 50%, Classe Lion 20% et Classe Singe 30%.

Les CNN ne sont pas universels. Ils excellent dans les tâches liées aux données visuelles ou aux grilles structurées, mais d'autres architectures sont mieux adaptées à des tâches comme la génération de texte, l'analyse de séries temporelles ou le traitement de données tabulaires. Le choix de l'architecture dépend donc de la nature de la tâche et des données.

Ainsi, il existe d'autres types d'apprentissage (**Section 4**) utilisés par l'intelligence artificielle.

Section 4 : Les différents types d'apprentissage en intelligence artificielle

La dénomination « intelligence artificielle » n'est pas que trompeuse en s'auto qualifiant « intelligente ». Elle cache aussi en réalité plusieurs types d'apprentissage. Par-exemple, Chat GPT fonctionne avec 4 différents types d'apprentissage machine.

Toujours dans cet objectif de mieux comprendre l'aspect technique de l'intelligence artificielle, il s'agira d'observer 3 différents types d'apprentissages : l'apprentissage supervisé (**Sous-section 1**), l'apprentissage par renforcement (**Sous-section 2**) et l'apprentissage autosupervisé (**Sous-section 3**).

Sous-section 1 : L'apprentissage supervisé : une limite de l'apprentissage machine

L'apprentissage supervisé est une sous-catégorie du machine learning et de l'intelligence artificielle. Il est aussi connu sous le nom de machine learning supervisé. Il n'est pourtant qu'un pâle reflet de l'apprentissage humain.

Simplement, l'apprentissage supervisé se calque sur des jeux de données qui sont nécessairement annotés initialement par un humain (ex : images d'un zèbre).

L'annotation permet à l'algorithme de savoir quel résultat est attendu. L'objectif est d'entraîner un algorithme à classer correctement les données ou même de pouvoir prédire un résultat. L'algorithme repose sur le principe d'une architecture où les réseaux s'ajustent progressivement pour approcher le plus possible la tâche demandée. L'apprentissage supervisé utilise le principe du réseau neuronal convolutif expliqué précédemment.

Ce système comprend des entrées et des sorties correctes, le modèle s'améliore donc au fil du temps. L'algorithme s'auto évalue en mesurant par le biais de la fonction de perte l'écart d'erreur et la réduit si nécessaire pour obtenir le résultat souhaité.

Le premier souci de cet apprentissage est qu'il est nécessaire de fournir à l'algorithmes une quantité énorme d'images (plusieurs millions) pour reconnaître les différents objets. Par-exemple, une banane verte reste une banane même si celle-ci n'est pas jaune. L'algorithme s'il a accès à un petit jeu d'images de bananes vertes pourra l'assimiler très facilement à une pomme verte, ce qui est problématique.

L'autre difficulté est qu'il faut étiqueter chacune de ces images. Comment la machine peut-elle savoir que l'image fournit est une banane ou un chien. La machine ne comprend rien de notre monde, il faut donc le lui dire avec le langage machine.

Ainsi, cet apprentissage est certes efficace mais il le demeure uniquement dans un environnement spécifique. Chaque changement peut perturber l'apprentissage effectué.

Par-exemple, des études ont montré que la modification légère d'un panneau de stop pouvait conduire à ce que la voiture ne le détecte pas. Luc Julia chercheur en IA utilise cet exemple : un homme marchant avec un panneau stop le long d'une route. La voiture « autonome » s'arrêtera toutes les trois secondes. L'humain comprendrait instantanément la supercherie.

Autre exemple, si l'on fournit à une machine l'image d'un lion et d'un poisson, il est possible en modifiant les valeurs de certains pixels de la photo d'obtenir le résultat poisson alors que l'image est un lion. Pour l'être humain, l'image est visuellement un

lion mais pour la machine ce n'est pas le cas, car elle analyse les valeurs des pixels sous le prisme des mathématiques. Une étude du MIT a ainsi démontré que des changements invisibles à l'œil nu de l'image ont conduit la machine à assimiler la tortue à une arme à feu⁹.

Comment la machine peut-elle se tromper ?

Chaque pixel se compose de 3 sous-pixels : rouge, vert et bleu. Chacun de ces sous-pixels peuvent aller de 0 à 256 valeurs de luminosité de leur couleur respective.

Il y a donc « 256 puissances 3 » valeurs possibles soit : 16 777 216 possibilités de combinaisons de couleurs qu'un pixel peut prendre.

Voici une image composée de 4 pixels :



Il y aurait alors pour une image de 4 pixels : $16\,777\,216 \times 16\,777\,216 \times 16\,777\,216 \times 16\,777\,216 = 8^{28}$ possibilités d'images si cette image comporte 4 pixels.

On aurait donc 723 trillions d'images que l'on pourrait générer avec uniquement 4 pixels.

Ainsi, un ensemble d'apprentissage de 1 milliard d'exemples ne couvrirait seulement qu'une infime partie des possibilités.

Il demeure une certaine fragilité de ces systèmes, sensibles à de petites modifications.

⁹ <https://news.mit.edu/2019/why-did-my-classifier-mistake-turtle-for-rifle-computer-vision-0731>

L'apprentissage supervisé ne permet pas de construire des machines réellement intelligentes. Il ne constitue qu'une petite pièce du puzzle. L'apprentissage par renforcement est un apprentissage davantage spécifique mais tout aussi nécessaire à élaborer un apprentissage complet pour un système d'intelligence artificielle.

Sous-section 2 : L'apprentissage par renforcement : le roi des échecs

L'apprentissage par renforcement fonctionne différemment de l'apprentissage supervisé. Il permet d'entraîner une machine sans lui donner la réponse attendue, en lui indiquant seulement si la réponse qu'elle a produit est juste ou non. C'est tout simplement un système de « punition-récompense ».

L'approche est différente. Cet apprentissage fonctionne uniquement lorsque l'on est capable d'évaluer la qualité de la réponse du système et ce, sans qu'on lui fournisse la bonne réponse.

Par-exemple, il serait fastidieux de programmer un robot pour qu'il puisse monter les marches d'un escalier. Le programme serait bien trop compliqué. Il faudrait dire à chaque instant quel moteur le robot doit activer pour accomplir la tâche.

Par-contre, il est facile d'évaluer si le robot a réussi à monter la première marche et ou l'escalier dans sa totalité. Le robot va donc essayer une stratégie, observer si elle fonctionne et en essayer une autre si elle ne fut pas concluante. Le robot peut avoir réussi à monter la première marche mais il peut chuter. Il faut alors revoir la stratégie.

Cet apprentissage par essais et erreurs avec évaluation du résultat sans que l'on donne la réponse correcte à la machine s'appelle « l'apprentissage par renforcement ». L'évaluation du résultat pourra aussi être fait automatiquement par le robot, le système va auto apprendre. Il est impossible de modifier la sortie pour amener le système vers la récompense.

Ce système est surtout utilisé lorsqu'il s'agit d'un jeu ou de contrôler un robot.

Prenons l'exemple des échecs: la machine va utiliser le deep learning et le renforcement. Le programme ordonnera à la machine de jouer avec des règles prédéfinies et détermine quels sont les coups qui l'amèneront avec le plus de probabilité vers la victoire. Ainsi, la machine pourra jouer des milliards de parties contre elle-même. Parfois les stratégies seront hasardeuses mais les probabilités statistiques mèneront la machine à un entraînement presque infaillible.

Quelles sont les limites de l'apprentissage par renforcement ?

Ce type d'apprentissage est surtout efficace dans un environnement bien défini et dont les règles sont quasi immuables. L'application de l'apprentissage par renforcement dans sa forme la plus pure est inapplicable dans le monde réel.

Il serait impossible d'apprendre à une voiture de conduire en utilisant l'apprentissage par renforcement. Dès le premier virage, la voiture finirait dans le premier arbre sur sa route. Imaginer le coût pour un seul essai. Le deuxième essai elle déciderait donc de tourner légèrement plus à gauche. La voiture retaperait l'arbre quelques centimètres un peu plus à gauche. Il faudrait des millions d'heures de conduites et des milliers d'accidents pour que la voiture puisse apprendre à rester simplement sur une route et l'on ne prend pas en compte les voitures présentes sur la route.

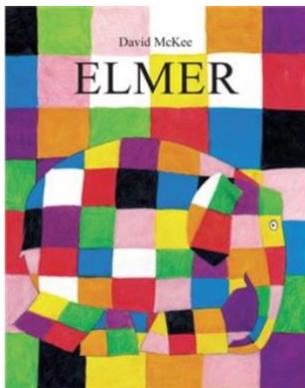
Ce qui permet aux hommes de savoir conduire en une vingtaine d'heures ne se trouvent pas dans les principes de l'apprentissage par renforcement ou de l'apprentissage supervisé. Les simulateurs sont une piste mais encore faut-il qu'ils soient assez puissants, reflète assez fidèlement la réalité et que l'on puisse transposer tout cela au monde réel.

L'apprentissage par renforcement ne constitue qu'une infime partie de l'intelligence artificielle et n'est applicable que lorsque les règles sont bien définies en amont. Cette technique est en réalité peu adapté lorsqu'il se présente une masse d'interactions liées à un environnement.

Sous-section 3 : L'apprentissage « autosupervisé » : idéal de l'apprentissage humain

L'apprentissage « autosupervisé » est une amélioration de l'apprentissage « supervisé ». C'est notamment l'un des apprentissages majeurs de ChatGPT.

Prenons l'exemple du cours de la vie d'un enfant jusqu'à ses 9 mois. Au cours de ses premiers mois, l'enfant apprend énormément. A ses 2 mois l'enfant est conscient qu'il existe des objets animés et des objets inanimés. Si l'on présente une dizaine d'images d'éléphant à cet enfant, il sera capable de reconnaître un éléphant sur n'importe quelle image et ce, même s'il s'agit d'Elmer l'éléphant !



Un système basé sur l'intelligence « supervisé » aura du mal à détecter un éléphant sur cette photo. Sauf si son corpus de données contient des photos du célèbre Elmer étiqueter comme représentant un éléphant.

Rappelons-le, le système n'a aucune compréhension de notre monde.

L'enfant va ainsi développer une faculté de prédiction, qui est essentielle dans le cadre de notre raisonnement. Les machines actuelles en sont totalement dénuées.

Notre intelligence est la combinaison des 3 systèmes d'apprentissages énoncées préalablement. Un randonneur sait que s'il ne s'arrête pas au bord de la falaise, il fera une grosse chute et qu'il décèdera sûrement. Il n'a pas besoin d'essayer pour le savoir. Nous sommes ainsi dotés de modèles prédictifs qui nous sommes propres. La voiture « autonome » n'est pas doté de ce modèle prédictif.

L'idée de l'apprentissage « autosupervisé » est de prendre une donnée d'entrée, de masquer une partie de cette entrée et d'entraîner la machine à prédire la partie masquée à partir de la partie visible. Contrairement à l'apprentissage « supervisé », le résultat souhaité est une partie de l'entrée qui a été préalablement masquée.

La réelle difficulté est de représenter l'incertitude. Prenons l'exemple d'un clip vidéo, la machine doit alors compléter la vidéo d'un monsieur qui est entrain de marcher.

L'homme pourrait très bien commencer à courir ou bien tomber par terre car il n'a pas fait ses lacets.

S'agissant de la prédiction d'une machine pour deviner le mot suivant d'un texte (ex : ChatGPT), il est facile de représenter l'incertitude. La sortie du système est alors une liste de probabilités des mots les plus adéquats pour tel structure du texte.

La limite actuelle est lorsque l'entrée est une donnée continue et de grande dimension tel que les images d'une vidéo.

Si l'on devait résumer une limite actuelle de l'intelligence artificielle, il s'agirait de dire qu'elle manque d'une compréhension générale de notre monde et que la représentation mathématique de cette perception, du fait de sa grande diversité, la limite à des tâches simplistes. Du moins, à des tâches qui soient transcribibles mathématiquement à la machine.

La conférence de Darmouth et ses illustres participants ont certes le mérite d'avoir fait émerger la dénomination « d'intelligence artificielle ».

Malgré tout, les mots choisis dépassent, à l'heure actuelle, la réalité. Il est plus raisonnable de parler d'intelligence « augmenté ». Une intelligence au service de l'humain et sous son contrôle qui lui permet d'augmenter son potentiel.

Ainsi, il ne s'agit pas d'exposer une vision pessimiste de l'avenir de l'IA, au contraire. Malgré tout, cela ne justifie pas pour autant que l'on adopte une vision hors réalité et idéalisée de l'IA.

Le **Chapitre 2** invite au questionnement et à un état des lieux sur la véritable nature « intelligente » de l'intelligence artificielle.

En effet, si l'on souhaite construire une réglementation innovante, il est aussi nécessaire de placer l'intelligence artificielle sous le prisme de certaines réflexions issues des sciences sociales.

CHAPITRE 2 : RÉFLEXIONS GÉNÉRALES SUR L'INTELLIGENCE ARTIFICIELLE

Section 1 : L'intelligence artificielle et l'intelligence humaine

Les algorithmes ne comprennent rien à notre monde et les humains ne comprennent rien aux algorithmes. Les algorithmes deviennent de plus en plus complexes et il est souvent difficile d'expliquer pourquoi tel résultat a été généré.

En réalité, rien ne devrait fonder la comparaison entre l'IA et l'intelligence humaine. L'intelligence humaine demeure une intelligence à part, dont la comparaison avec toute autre forme d'intelligence peut s'avérer délicate.

Malgré tout, la comparaison reste inévitable. Nier cette comparaison, c'est en quelque sorte nier les prémices fondamentales de l'intelligence artificielle qui trouve son existence même dans l'objectif d'imiter l'intelligence humaine.

« Toujours imité, jamais égalé ». A l'heure actuelle, difficile de savoir si l'intelligence artificielle égalera l'intelligence humaine.

Certains l'espèrent, d'autres le craignent. Mieux vaut le craindre.

Malgré tout, l'essentiel est avant tout de se questionner sur « l'intelligence pure » de l'intelligence artificielle comparé à l'intelligence humaine (**Sous-section 1**)

Sous-section 1 : L'intelligence artificielle est-elle intelligente ?

L'intelligence artificielle a été autrement définie comme « *la science qui consiste à faire faire par des machines des choses qui nécessiteraient de l'intelligence si elles étaient faites par des hommes* ».

Cependant, l'intelligence artificielle ne dispose pas d'une conscience propre. Ce que nous appelons aujourd'hui intelligence artificielle ce sont des agents intelligents formés par des modèles d'apprentissage automatique. On peut citer l'apprentissage

supervisé, l'apprentissage non-supervisé, l'apprentissage profond et l'apprentissage par renforcement.

L'apprentissage profond étant une sous-branche de l'apprentissage automatique qui a été introduite dans le but de rapprocher l'apprentissage automatique de l'un de ses objectifs initiaux, à savoir l'intelligence artificielle. L'intelligence artificielle générative a ainsi profité des avantages récents de l'apprentissage automatique, des ensembles de données massives et des augmentations substantielles de la puissance des ordinateurs. Les intelligences artificielles génératives tel que Chat GPT, démontrent leur capacité à produire des contenus divers et variés.

A cet égard, OpenAI a annoncé son tout nouveau modèle « o1 », plus performant que la version 4.0 sortie en mai 2024. Les grands acteurs dans le domaine de l'intelligence artificielle ne font que repousser les limites du progrès technologique. Il est certain que l'avancée à laquelle l'intelligence artificielle se développe est impressionnante et ne fait que renforcer la vision d'une intelligence artificielle comparable à l'intelligence humaine.

Certains l'affirment, l'intelligence artificielle est intelligente, elle raisonne et permet de résoudre des problèmes extrêmement complexes. En réalité, cette théorie attribue une qualité que « l'intelligence artificielle ne dispose pas.

D'autres empruntent une théorie davantage basée sur l'essence même du fonctionnement de l'intelligence artificielle. Ils avancent notamment le fait que l'intelligence artificielle ne procède que par le biais d'algorithmes finement ajustés, qu'elle ne fait preuve d'aucune conscience et n'a donc aucune remise en question ou faculté de différencier ce qui est éthiquement correcte ou non par-exemple.

L'intelligence artificielle fonctionne comme un lave-vaisselle, un micro-onde ou une machine qui mout le café, à la différence qu'elle fonctionne avec un algorithme ultra développé et un jeu de données vaste.

En réalité, l'intelligence artificielle générative est certes une innovation technologique. Malgré tout, elle reprend la vieille idée que le calcul peut remplacer le raisonnement et

que la connaissance suffit sans compréhension. C'est sur ce point précis que subsiste le débat sur l'intelligence artificielle. Le risque est de créer une machine qui recycle et organise un ensemble de données. Une technologie qui reste figé dans une dimension tournée vers le passé où les données traitées sont pour une partie déjà « dépassées » par l'actualité.

Sous-section 2 : Doit-on se préparer à une IA plus intelligente que nous ?

Si l'on devait placer sur un curseur le stade de développement de l'IA à l'heure actuelle, elle se placerait à ses prémices.

L'intelligence artificielle est souvent comparue, à tort, à notre intelligence humaine. Ce raccourci est malheureusement dû à une méconnaissance du fonctionnement technique réel de l'intelligence artificielle. L'intelligence à proprement parler ne peut aujourd'hui se rapporter à ce qu'est capable l'intelligence artificielle pour l'instant.

Dans le cas de l'humain, l'intelligence est généralement définie comme la capacité à comprendre, résoudre des problèmes et s'adapter à des nouvelles situations. Dès lors, comment peut-on comparer une machine conçue pour traiter des informations à un cerveau humain doté d'une conscience. Il est difficile d'attribuer à une intelligence artificielle des qualités aussi abstraites que la créativité ou l'intuition.

Cependant, l'intelligence artificielle n'a aucune capacité de raisonnement ou de prédiction. Elle peut déduire certaines choses grâce à son vaste corpus de données mais elle ne peut créer quelque chose d'elle-même sans pouvoir se baser sur ses données d'entraînement. L'intelligence artificielle n'est pas consciente de ce qu'elle génère. Elle exécute des tâches selon des modèles probabilistiques et statistiques.

Prenons un exemple : ChatGPT n'aurait aucun mal à vous détailler les étapes à suivre pour commettre l'assassinat de votre voisin du dessus sans vous faire attraper par la police. C'était d'ailleurs une possibilité, avant que les développeurs n'aient restreint les prompts à caractère illicite.

L'intelligence artificielle générative, comme ChatGPT est certes une avancée considérable. Elle est capable de générer de la voix, des vidéos dans toutes les langues et créer des images avec différents styles artistiques. Malgré tout, ce n'est pas parce que les machines réalisent à l'heure actuelle des tâches qui nous semblent complexes qu'elles pourront demain faire aussi ce qui nous est simple.

Hans Moravec¹⁰, a souligné justement un paradoxe en lien avec l'évolution de l'IA : « *le plus difficile en robotique est souvent ce qui est le plus facile pour l'homme* ».

Un exemple flagrant est celui de la voiture à conduite automatisée, dont sa démocratisation est sans cesse reportée. Cela illustre la double difficulté de lever les verrous technologiques et déployer effectivement les systèmes automatisés en situation réelle.

La réponse serait donc de dire : non l'intelligence artificielle n'est pas plus intelligente que nous et elle n'est pas encore prête à le devenir

Le cerveau humain comporte 86 milliards de neurones, celui d'un chien 2,2 milliards et celui d'un chat 760 millions.

Pour évaluer où l'IA se situe en termes d'intelligence pure, il est bon de rappeler que le cerveau humain peut effectuer $1,5 \times 10^{18}$ opérations par seconde. Une carte GPU peut en effectuer 10^{13} . Il en faudrait 100.000 pour approcher la puissance du cerveau.

La reproduction de la puissance de calcul d'un cerveau humain par la technologie est un défi de tout autre nature. On estime qu'un système d'intelligence artificielle est moins intelligent que le chat en termes de « puissance de calcul ».

Certes, comme évoqué précédemment avec le réseau neuronal convolutif, nous comprenons les principes de l'apprentissage dans le cerveau et sa structure.

Malgré tout, la puissance de calcul qui est requise pour en reproduire le fonctionnement est incommensurable.

¹⁰ Enseignant-chercheur au centre de robotique de l'université Carnegie-Mellon.

Pour obtenir la puissance d'un cerveau humain il serait nécessaire de connecter une centaine de milliers de processeurs (carte GPU) au sein d'un ordinateur géant qui consommerait au moins 25 mégawatts. `

Le cerveau humain consomme 25 watts de puissance et une seule carte GPU en consomme 250 watts. L'électronique a une consommation 1 million de fois plus élevée que la biologie si la consommation atteignait 25 mégawatts.

L'intelligence artificielle est à l'heure actuelle, un million de fois moins efficace que notre cerveau ! Yann Le Cun s'amuse même à dire une phrase qui fait état de ce contraste saisissant : « *L'IA reste 50 fois moins intelligente qu'un enfant de 5 ans* ».

Actuellement, Google ou Facebook ont accès à cette capacité de calcul mais il est extrêmement difficile de la mobiliser en totalité sur une tâche bien définie.

Il y a alors un réel défi technologique mais aussi scientifique si la technologie veut réellement jouer jeu égal avec l'humain.

L'informatique quantique sera peut-être la solution à cette puissance de calcul manquante. Malgré tout, il sera aussi nécessaire de se pencher sur de nouveaux principes mathématiques.

Sous-section 3 : La question de savoir si « l'intelligence artificielle est intelligente » est-elle pertinente ?

Il est vrai que la question de savoir si : « ***L'intelligence artificielle est-elle intelligente ? (Sous-section 1)*** » ne permet pas de saisir à elle seule l'entièreté des problématiques émergentes.

Néanmoins, l'interrogation permet de rappeler qu'il ne faut pas rentrer dans l'anthropomorphisme des machines en leur attribuant des qualités humaines, telles que l'émotion, la pensée ou la conscience.

En effet, l'intelligence artificielle est continuellement comparée à l'intelligence humaine. La comparaison entre le raisonnement humain et le raisonnement artificiel est tout à fait compréhensible. Le but ultime de l'intelligence artificielle étant d'atteindre ou du moins d'imiter l'intelligence humaine, la comparaison reste inévitable.

Cependant, certains critiquent le fait que ce débat est vain et qu'il n'apporte aucune plus-value intellectuelle.

Un expert tel que Mark Stevenson à l'université de Sheffield, a permis de recentrer le débat et de rappeler les compétences actuelles de l'intelligence artificielle. Selon lui, les algorithmes de machine Learning sont des « champions en probabilité, pas en raisonnement ». Pour résumer, les algorithmes utilisés sont excellents pour anticiper le mot suivant dans une phrase ou résoudre des équations mathématiques. Il rappelle à juste titre qu'elles ne sont pas du tout conçues pour comprendre réellement le sens ou les implications des problèmes qu'elles résolvent. C'est surtout leur efficacité en précision statistique qui est mis en avant, que leur réelle capacité de raisonnement.

La question de l'efficacité des algorithmes ouvre une réflexion bien plus pertinente et en phase avec la réalité que de vouloir évaluer une technologie sous le prisme de l'intelligence humaine.

On peut citer à cet égard ce que disait Albert Einstein : « Si vous jugez un poisson par sa capacité de grimper à un arbre, il vivra toute sa vie en croyant qu'il est stupide ».

L'IA n'a donc pas pour vocation, actuellement, d'égaliser la capacité de raisonnement de l'humain. L'humain raisonne et fournit à l'IA des données issues de son raisonnement. Dès lors, l'IA a vocation d'être avant tout efficace grâce à sa grande capacité d'analyse des données qui lui sont fournies.

Le débat est peut-être ailleurs. Le débat n'est peut-être pas tant celui de savoir si l'intelligence artificielle est empreinte d'une intelligence humaine.

La réelle préoccupation est sûrement l'avenir incertain de la manière dont l'intelligence artificielle impactera positivement ou négativement notre monde.

Section 2 : L'intelligence artificielle une chance ou un danger ?

Pour répondre à cette interrogation, on pourrait simplement dire que tout dépend de l'utilisation que l'on fait de l'intelligence artificielle.

Premièrement, une première observation est que, tout comme l'humain, l'intelligence artificielle a besoin de connaissances pour pouvoir fonctionner.

Si l'intelligence artificielle se met à apprendre à partir de données approuvées par un consensus, l'intelligence artificielle est alors une réelle aide pour l'homme. La vitesse de traitement de données par l'IA est à un stade où elle peut assimiler des données bien plus vite et efficacement que l'Homme qui lui, est sujet à fatigue. Ainsi, si l'on se place dans l'aspect purement technique de l'intelligence artificielle, il peut être affirmé qu'elle est sûrement une chance pour l'homme.

Malgré tout, l'impact de l'intelligence artificielle au sein de la société, est profond et contrasté. Les préoccupations tiennent notamment en ce que l'intelligence artificielle nécessite une grande quantité d'informations personnelles pour fonctionner efficacement. Éthiquement, les éventuelles discriminations que peut opérer l'intelligence artificielle est un réel défi à surmonter.

Concernant les domaines impactés par l'intelligence artificielle, le secteur de l'emploi est souvent évoqué. La crainte majeure étant le remplacement des humains par la technologie, que cela soit l'intelligence artificielle ou l'automatisation par les robots.

En réalité, cette crainte suit l'évolution naturelle de notre monde et du progrès technologique.

Prenons l'exemple du métier de falotier qui consistait en l'allumage des réverbères des rues pendant la révolution industrielle. L'avènement de l'éclairage électrique l'a fait

disparaître. Sûrement que sa disparition eut suscité la colère de ces allumeurs de réverbères.

Pour autant, est-il pertinent de s'inquiéter de sa disparition ? Une fois la vague transformationnel et sociétale accompagnant l'intelligence artificielle passée, les craintes principales se seront sûrement dissipées.

Il est davantage essentiel de réfléchir comment accompagner le changement qui s'annonce avec l'intelligence artificielle. Cet accompagnement a déjà commencé avec l'établissement de normes juridiques européennes (AI Act) règlementant l'intelligence artificielle de son entraînement à son déploiement sur le marché.

Il est plus intéressant d'envisager l'intelligence artificielle sous le prisme d'une collaboration.

L'European Parliament Think Thank¹¹ estimait en 2020 que l'IA permettrait une augmentation de la productivité au travail allant de 11% à 37% d'ici 2035.

Cette collaboration pourra recentrer le travail sur l'essentiel en accélérant les tâches chronophages, répétitives et fastidieuses. Dans le domaine médical, les médecins profiteront d'une analyse rapide des vastes ensembles de données médicales pour poser des diagnostics plus précis. Dans le domaine de l'industrie, les ouvriers pourraient collaborer avec des robots d'IA pour ne plus à effectuer les tâches dangereuses. Sous l'angle de la collaboration, le potentiel humain sera décuplé et stimulera l'innovation à travers divers secteurs.

L'IA doit aussi répondre à des problématiques, et surtout éviter d'en créer, notamment sur la question environnementale.

L'intelligence artificielle réduirait les émissions mondiales de gaz à effet de serre de 1,5% à 4% d'ici 2030. L'heure étant à la sobriété énergétique et l'utilisation d'énergies

¹¹ <https://yperia.fr/lintelligence-artificielle-va-t-elle-booster-la-productivite-des-entreprises/>

renouvelables. Il serait dangereux de s'aventurer dans le développement de l'intelligence artificielle sans prendre en compte ces paramètres.

La réglementation d'une IA respectueuse de l'environnement doit effectivement être envisagée. L'IA Act mentionne cette obligation malgré tout faut-il encore que le cadre proposé soit sérieux et qu'il soit respecté par les acteurs concernés.

Le réel problème étant que l'IA engendra peut-être tout autant de pollution qu'elle n'en fera l'économie.

Il ne faut pas oublier que l'intelligence artificielle se paye par un coût énergétique colossale. Un chiffre pour l'illustrer : en 2026 la consommation d'électricité générera un surplus de 37 milliards de tonnes de CO2 dans l'atmosphère (Rapport Agence International de l'énergie 2023- Electricity Market Report – centre de données / IA / cryptoactifs). A titre de comparaison, c'est la consommation annuelle d'un pays comme le Japon.

L'intelligence artificielle pourrait ainsi tenter de remédier à notre surconsommation d'électricité. Il faut néanmoins prendre le sujet avec prudence. D'autant plus que la principale préoccupation principale des acteurs majeurs sur le marché de l'intelligence artificielle est plutôt axée sur la finance. Sam Altman, patron d'OpenAI avait déclaré début janvier 2024 ne pas « *savoir mesurer les besoins en énergie de cette technologie* ».

Le danger c'est aussi d'être convaincu que l'IA rendra les technologies actuelles obsolètes et donc qu'il y aura une économie énergétique dans l'expansion de l'IA avec la disparition d'autres technologies ou usages ou une amélioration de ces derniers.

En réalité, l'histoire a pu nous dire que cela est généralement faux. Les usages s'additionnent mais ne se remplace pas. Il s'agit d'une superposition croissante des usages des nouvelles technologies. L'IA ne va pas remplacer les usages existants elle va surtout créer de nouvelles opportunités. C'est un schéma itératif dans l'histoire du numérique : chaque avancée technologique s'intègre dans l'écosystème existant et enrichit les pratiques plutôt que de les éliminer complètement.

Pour l'instant, difficile de donner une réponse précise et il serait insensé de vouloir y répondre avec justesse tant le sujet est fluctuant. Ce qu'il faut retenir, c'est que l'intelligence artificielle prendra la place que l'on voudra bien lui donner. Si elle en prend de trop, c'est que l'humain l'aura décidé.

Notre objectif commun est de trouver un équilibre entre les avantages et les inconvénients de l'intelligence artificielle. Un défi auquel la société doit faire face à mesure que cette technologie continue de se développer en posant ainsi davantage de défis.

Finalement, l'intelligence artificielle pourrait se comparer au feu : lorsqu'il est bien maîtrisé, il nous éclaire et nous réchauffe, mais hors de contrôle, il peut s'avérer destructeur.

Section 3 : La comparaison entre l'émergence d'Internet et de l'Intelligence artificielle

La comparaison est souvent faite entre Internet et l'émergence récente de l'intelligence artificielle. Souvent, on veut démontrer que l'intelligence artificielle a, tout comme Internet, bouleversé nos quotidiens, nos usages et les perspectives de notre monde. En effet, Internet a ouvert un monde des possibles qui était avant son apparition inimaginable pour certaines personnes.

Le plus grand changement est sûrement celui de la transformation de la communication. Peu à peu, nos relations se sont ouvertes exponentiellement vers l'international. Du fait de la connexion instantanée entre les personnes, peu importe où elles se trouvent, la façon dont les gens interagissent entre eux a radicalement changé.

Il est par-ailleurs difficile de trouver un domaine qui n'ait pas été impacté par l'émergence d'internet, ayant transformé notre monde papier par un réseau mondial interconnecté par la numérisation de nos données.

Des thèmes communs s'observent en mettant en parallèle le développement de l'IA avec les débuts de l'Internet. Dès que le thème de l'intelligence artificielle est évoqué, souvent la question sous-jacente est la suivante : « Comment l'intelligence artificielle va-t-elle bouleverser notre monde ? ». Les craintes à ce sujet ne tardent souvent pas à se poser, elles aussi.

Malgré tout, cette analogie est-elle réellement pertinente ?

Il serait plus juste de comparer l'IA à des révolutions de l'ordre technologique tel que les microprocesseurs ou les ordinateurs personnels. En effet, Internet n'est rien sans ses supports matériels, tel que les ordinateurs tout comme la 4G/5G n'est rien sans les smartphones. On parle souvent de technologies immatérielles mais le numérique est le domaine le plus matériel qui existe à l'heure actuelle. L'IA suit cette logique : sans data center l'IA n'existerait pas.

Ainsi, l'IA pourrait s'observer comme un nouveau type de « super ordinateur ». Un ordinateur doté de capacités d'apprentissage et de traitement des données à très grande échelle et d'une rapidité sans pareil. L'intelligence artificielle est davantage une nouvelle puissance de fonctionnement qu'une technologie matérielle à proprement parler.

D'un autre côté, l'émergence de l'IA n'est pas strictement résultante à une recette mathématique bien concocté combinant inférence bayésienne et neurone formel. Elle doit aussi son apparition au progrès technique des puces informatiques tel que les GPU (Graphics Processing Unit) signifiant « unité de traitement graphique ». Plus le progrès technique s'est développé, plus les principes mathématiques intégrés dans l'intelligence artificielle ont pu fonctionner efficacement.

L'IA est tout autant une technologie qui a vocation à fonctionner en autonomie qu'une technologie qui s'intègre dans un dispositif technique précis tel que l'intégration de l'IA dans une voiture Tesla pour la conduite sans chauffeur.

Ainsi, le développement du marché de l'IA pourrait suivre un chemin similaire à celui de l'informatique, qui au début était centralisé puis s'est dispersé.

Cette phase de dispersion de la technologie de l'IA s'est déjà largement amorcée. Cette perspective éclaire sur le potentiel transformationnel de l'IA.

L'IA a complètement bouleversé le domaine du droit et si elle ne l'a pas encore complètement fait, cela arrivera. Dans tous les cas, qu'on le veuille ou non, l'intelligence artificielle fera partie intégrante de notre future. Ainsi, il vaut mieux l'accepter et décider d'agir dès à présent que de rester simple spectateur.

A ce titre, il s'agira de voir de quelle manière l'Europe a décidé d'encadrer juridiquement l'intelligence artificielle tout en assurant l'innovation en la matière.
(Chapitre 3)

CHAPITRE 3 : L'INNOVATION RESPONSABLE À TRAVERS LE RÈGLEMENT EUROPÉEN SUR L'INTELLIGENCE ARTIFICIELLE

L'idée de réglementer l'IA s'observe d'une part comme un besoin plus que nécessaire mais aussi à un défi de taille. Pionnière en la matière, l'Europe s'est ainsi attelée rapidement au problème et a fait émerger un règlement européen : l'IA Act.

Édicter un cadre réglementaire européen via l'IA Act fait écho à ce que j'avais écrit Rabelais dans son premier chef d'œuvre « *Pantagruel* » : « *Science sans conscience n'est que ruine de l'âme* » 1532, F. Rabelais.

La question se pose de savoir de quelle manière l'IA Act entreprend cette réglementation de l'intelligence artificielle (**Section 1**) tout en assurant un terreau favorable à l'innovation (**Section 2**) et ce, malgré une balance d'intérêts difficile à maintenir en équilibre (**Section 3**)

Section 1 : Le RIA et le développement d'une intelligence artificielle responsable

L'IA Act européen voté le 13 mars 2024, devient le fondement incontournable vers la responsabilisation des fournisseurs d'IA et des divers acteurs intervenant dans le cycle de déploiement de l'IA.

La base juridique du projet de règlement est l'article 114 TFUE qui est relatif aux mesures destinées à garantir un bon fonctionnement du marché intérieur.

Le RIA s'inscrit dans le contexte du marché unique numérique. L'objectif du RIA est davantage de garantir des règles harmonieuses quant au développement et à l'utilisation de produits ou services exploitant l'IA. Le principal risque étant la fragmentation du marché européen. La motivation principale n'est donc pas tant la protection des droits des citoyens européens, tel que la protection des données personnelles. L'objectif est en grande partie commercial. La confiance est entre autre la base du commerce. Le RIA a pour but de développer cette confiance pour favoriser le développement d'IA éthiques acceptées.

Malgré tout, le RIA vise à adopter une posture influente sur la scène mondiale en contribuant à la définition de normes mondiales et à la promotion d'une IA en conformité avec les intérêts de l'Union.

Le défi n'est pas mince puisque l'objet de la réglementation suscite des craintes notamment de complexité, d'imprévisibilité et d'opacité de certains « systèmes d'IA » dont les biais algorithmiques sont au centre du débat.

Certains auteurs¹² proposent une analyse de l'intelligence artificielle autour de la « théorie des baïonnettes intelligentes ». Cette métaphore permet de mieux rendre compte de la tendance actuelle à la déresponsabilisation des acteurs de l'IA.

¹² Dominique Boullier et Aurélie Jean : « L'IA, l'éthique et la théorie des baïonnettes intelligentes »

Le RIA met ainsi l'accent sur une certaine prévention et gestion des risques, obligeant les fournisseurs d'IA à se questionner continuellement sur l'impact de leur conception dans un contexte de risques croissants et de croissance fulgurante des acteurs tel que OpenAI (Article 9 RIA)

Si prévenir les risques est déjà une tâche ardue, elle devient irréalisable s'il est impossible de les prédire. Comment prédire ces risques ? Le RIA a fait le choix d'une pyramide de Kelsen des risques des systèmes d'IA. Le point névralgique réside donc dans la manière dont le RIA a choisi de prédire et structurer le degré de ces risques.

Il s'agira de comprendre en quoi l'approche par les risques ne permet pas complètement d'effacer les craintes entourant l'IA (**Sous-Section 1**)

Sous-section 1 : Le choix d'une classification par les risques

La réelle particularité du RIA (Règlement sur l'intelligence artificielle) est d'être construit sur une approche par les risques. Le règlement établit 3 différents risques : les systèmes d'IA à hauts risques, à risques acceptables et à risques minimaux.

Les exigences établies au sein du RIA concerne presque exclusivement les systèmes d'IA à hauts risques. Un choix du RIA est donc de se focaliser sur ce qui est le plus risqué.

La classification d'un système d'IA à haut risque dépend donc de la fonction que ce système remplit, de sa finalité et de ses modalités d'utilisation.

En somme, l'approche par les risques consiste à ` autoriser sur le marché européen des systèmes d'IA sous réserve du respect de certaines exigences obligatoires et d'une évaluation ex ante de la conformité.

L'article 9 du RIA évoque à ce titre un système de gestion des risques (documenté) qui consiste à analyser les divers risques qui sont dits « *raisonnablement prévisibles* ». Il est aussi évoqué aussi l'évaluation des risques qui seraient « *susceptibles d'apparaître* » lorsque le système est utilisé dans de mauvaises conditions d'utilisation.

La Commission se réserve le droit de modifier l'annexe III listant les différents systèmes d'IA à haut risque.

On peut légitimement se questionner sur la pertinence d'une telle approche par les risques envisagée par le RIA, mot cité 177 fois dans le RIA.

Le problème est le suivant : Comment le droit peut-il appréhender une violation des droits fondamentaux lorsqu'il s'agit de décisions algorithmiques d'un système d'IA à haut risque ?

La complexité des algorithmes rend difficile l'identification des causalités de la décision algorithmique prise. Prenons un exemple : « Une IA autonome porte un préjudice. Qui est responsable ? Le fournisseur d'IA. Le RIA répond par un pragmatisme uniformisé alors que l'interrogation, suscitant des enjeux diversifiés, nécessiterait une réponse sur-mesure.

N'est-il pas risqué d'avoir un système d'évaluation basé sur les risques alors même que certains risques sont totalement imprédictibles ?

Cette interrogation en entraîne une autre, notamment celle de savoir s'il est réellement possible de prédire avec justesse les risques que pourrait engendrer une technologie si mouvante et complexe que l'IA.

En effet, l'approche par les risques du RIA ne prend uniquement en compte les risques identifiés.

Certes, il est possible de déterminer le risque de certains systèmes d'intelligence artificielle lorsqu'elles interviennent dans des domaines sensibles tel que la santé ou qui dont leur finalité est éthiquement douteuse (scoring social).

D'un autre côté, il est tout à fait possible qu'une intelligence artificielle soit non risqué lors de son élaboration mais le devienne dans un second temps.

Le RIA ne fait-il pas le jeu de l'IA en abordant une conception prédictive des risques ?

Un système d'intelligence artificielle fonctionne par prédiction. Il existe une part importante de probabilité et de hasard. L'intelligence artificielle établit ainsi la probabilité qu'un tel résultat soit associé à telle entrée.

Voici la définition du « risque » au sein du RIA : « la combinaison de la probabilité d'un dommage et de la gravité de ce dommage ».

Ainsi, il est légitime de se demander s'il est réellement possible de prédire les risques ?

La conception du risque demeure extrêmement vague et subjective. L'absence de délimitations objectives des systèmes d'intelligence artificielle rend l'encadrement juridique incertain et problématique. En effet, les notions mobilisées dans l'IA Act, comme celle de "systèmes d'intelligence artificielle", s'apparentent davantage à des constructions abstraites qu'à des réalités tangibles. Ces fictions juridiques, déconnectées des dynamiques techniques et pratiques, aboutissent à des régimes juridiques hasardeux et souvent inadaptés.

Ce flou conceptuel, couplé à une régulation précipitée, confère une légitimité et une légalité à des technologies dont les impacts réels restent imprécis, voire imprévisibles.

Par conséquent, de telles réglementations risquent de faciliter le déploiement de dispositifs aux effets sociétaux ou éthiques insuffisamment anticipés. Ces dérives entraînent un cycle récurrent d'ajustements réglementaires pour pallier les lacunes des cadres initiaux, témoignant de l'instabilité et de l'impréparation de cette régulation.

Cette critique invite à repenser les bases conceptuelles et méthodologiques de la régulation de l'IA. Il faut éviter que l'encadrement juridique ne devienne lui-même une source de vulnérabilités pour les individus et les institutions.

Si la question devait être : « Légiférer sur ce qui existe est ou sur ce qui va devenir », le RIA a plutôt choisi la voie de : « légiférer sur ce qui pourrait être ».

Un paradoxe apparaît. Comment appliquer le RIA dans des entreprises à des experts de l'IA qui, bien qu'informés, n'ont pas encore les moyens et les repères nécessaires pour agir efficacement dans un secteur qui évolue constamment.

Ainsi, les termes utilisés par la Commission sont certes larges et ne permettent pas une appréhension précise de la réglementation. Malgré tout, ce choix permet au règlement d'être flexible et de pouvoir s'adapter aux éventuelles évolutions technologiques.

Les « systèmes d'IA », terme choisi par la Commission, est défini au sein de l'article 3 comme des logiciels qui « (...) *pour des objectifs explicites ou implicites, déduit, à partir des données qu'il reçoit, comment générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer des environnements physiques ou virtuels* »

La réglementation établie par le RIA s'apparente davantage à un « bac à sable » réglementaire, qu'à une réglementation contraignante envers les fournisseurs et utilisateurs des systèmes d'IA. Cela s'explique par le court délai pour établir cette réglementation et l'évolution rapide des IA génératives.

A titre d'exemple, le RIA devait initialement interdire strictement l'utilisation de l'IA s'agissant de la reconnaissance faciale. Par-la-suite, les États membres, dont la France a choisi d'assortir ce principe d'interdiction à certains cas d'usages.

Le Considérant 53 dispose à ce titre que : « *Cette interdiction ne devrait pas s'appliquer aux systèmes d'IA mis sur le marché pour des raisons strictement médicales ou de sécurité* »

Le RIA demeure une réglementation souple et préventive. Il s'agira de voir dans le Chapitre 2, au sujet des biais algorithmiques, que le principe de responsabilité est très difficile à appliquer lorsqu'il s'agit d'intelligence artificielle.

Cet encadrement législatif s'explique d'autre part par la nécessité de ne pas « tuer dans l'œuf » une technologie qui se développe encore. Le RIA essaye donc de résoudre cette équation complexe entre protection des principes éthiques et du soutien à l'innovation
(Section 2)

Section 2 : Le RIA et l'intégration de la nécessité de l'innovation

Le contexte international en matière de progression technologique de l'intelligence artificielle oblige l'Europe à ne pas s'enfermer dans un cadre réglementaire qui empêcherait toute innovation.

Du fait de son évolution rapide, les acteurs mondiaux se précipitent d'investir dans l'IA et ce, du fait de la crainte de se voir sinon, dépasser par des pays voisins.

Les États-Unis, la Chine, le Japon ou encore la Corée du Sud ont adopté des stratégies importantes visant à en faire des acteurs de premier plan sur la scène mondiale.

La Commission européenne établit chaque année une liste des 2500 premiers investisseurs en R&D du monde. Dans le domaine des technologies de l'information et des télécommunications, l'Union européenne dispose de 4 fois moins d'entreprises que les États-Unis. De plus, les entreprises américaines investissent 6 fois plus en R&D que les européens.

La souveraineté de l'Europe est une problématique de plus en plus préoccupante. Que devons nous faire pour préserver la culture française et européenne si les outils d'IA sont construits sur des référentiels étrangers ?

L'Europe est donc dans l'urgence d'agir. La quasi-inexistence d'un marché européen sur les semi-conducteurs n'arrange pas la situation. Cette nécessité plus que primordiale d'assurer l'innovation explique aussi la raison pour laquelle l'Europe a adopté une réglementation extrêmement prudente avec le RIA.

Au sein du RIA, ce sont les articles 57 à 63 qui exposent les diverses mesures permettant aux acteurs européens de ne pas s'enfermer dans une tour d'ivoire et de créer le terreau favorable à la création d'une innovation responsable.

Le RIA vise notamment à alléger la charge réglementaire concernant les start-ups et les PME. C'est dans ce cadre juridique, propice à l'innovation, qu'a été instauré des bacs à sables réglementaires (Sand box) pour instaurer un cadre sûr pour l'expérimentation.

L'article 57 du RIA dispose que les états sont dans l'obligation de mettre en place au moins « *un bac à sable réglementaire en matière d'IA* ». La date butoir étant fixé au 2 août 2026. Le réel avantage de ces bacs à sables est de placer l'innovation européenne dans une « *pole position* » en leur permettant d'être dès le départ en conformité avec le droit de l'union. Cela paraît logique mais c'est aussi une stratégie pour booster le marché européen une fois le RIA en vigueur.

A cet égard, l'article 57 dispose clairement qu'une documentation détaillant les activités menées dans le bac à sable doit être établie. Cette documentation servira notamment de preuve de la conformité du système d'IA à la réglementation.

S'agissant du traitement des données personnelles dans le cadre du « *bac à sable réglementaire* », l'article 59 autorise une certaine souplesse selon certaines conditions. Ainsi, si le système d'IA est développé dans l'intérêt public (santé, sécurité publique, protection de l'environnement...) des données à caractère personnel « *légalement collectées à d'autres fins* » pourront être traitées.

Cette exception ressemble à l'exception posée par le Considérant 53 concernant l'usage de l'intelligence artificielle pour la reconnaissance faciale.

Il faut tout de même mentionner que le traitement de données personnelles collectées à d'autres fins est soumis à de nombreuses conditions préalables. Notamment, les données ne devront pas quitter le cadre du « *bac à sable réglementaire* » et seront complètement supprimées une fois le projet terminé. L'objectif étant certes d'autoriser des données initialement interdites au traitement mais uniquement dans le but de l'entraînement des systèmes d'IA. D'autre part, le considérant 1(c) impose la mise en place de mécanismes de suivi des risques élevés pouvant survenir au cours de l'expérimentation. Une sorte de procédure d'urgence devra être mis en place à ce titre.

S'agissant de l'article 60 du RIA, il est en quelque sorte une exception à l'exception. Initialement, il s'agit d'innover dans un bac à sable réglementaire. Ici, le RIA autorise de sortir complètement de ce bac à sable pour tester une IA à haut risque dans des « *conditions réelles* ». Le RIA impose toute une liste d'obligations et de conditions à

remplir pour effectuer ce test. Ces obligations s'inscrivent dans cette volonté d'assurer une balance équitable entre innovation et réglementation. La possibilité offerte par le RIA demeure une opportunité avantageuse pour les fournisseurs d'IA. Le RIA adopte une vision dualiste. Elle rappelle à juste titre que tout « incident grave » suspendrait immédiatement les essais et ce, jusqu'à la résolution du problème.

En effet, le RIA n'oublie pas que les systèmes d'IA ne peuvent se développer au détriment des droits fondamentaux des individus.

Malgré tout, dans cette volonté louable du RIA d'esquisser ce qui lui semble être une balance parfaite entre réglementation et innovation, la critique reste présente. L'adage est connu, Rousseau formulait ainsi que : « Qui veut plaire à tout le monde ne plaît à personne ».

Le constat est nuancé, tant les start-ups que les associations de défense des libertés estiment ne pas trouver leur compte dans cette balance instaurée par le RIA.

Ainsi, le texte du RIA semble courir après une chimère en ayant pour quête un équilibre presque paradoxale entre d'une part la volonté d'innover et d'autre part la volonté de réglementer l'intelligence artificielle (**Section 3**)

Section 3 : La quête d'un équilibre paradoxale entre innovation et réglementation

La réglementation de l'IA est tout autant une affaire politique qu'une affaire de droit.

En réalité, la puissance des acteurs du numérique est telle (Meta, OpenAI, Microsoft...) qu'il n'est plus envisageable de ne pas prendre en compte leur avis sur la question. L'Europe est consciente de cela mais d'un autre côté, elle ne peut se laisser aveuglement influencé par ces acteurs qui posent une diversité de problèmes éthiques en matière de traitement des données.

Les Big techs sont eux davantage favorable à ce que le droit n'interfère pas avec l'intelligence artificielle. On retrouve dans ce discours notamment le père fondateur de l'IA : Jurgen Schmidhuber qui plaide pour une non-règlementation de l'IA.

Selon lui, même si une réglementation existerait, les entreprises seraient presque contraintes à ne pas la suivre tant le potentiel économique est important.

Le droit est réellement vu comme étant indésirable. Ainsi, le droit s'apparente à un obstacle à l'innovation et donc à l'intelligence artificielle. Le discours tenu est souvent teinté de mépris en soulignant la folie de vouloir réglementer quelque chose qui entame tout juste son évolution. A cet égard, Yann Le Cun, Marck Zuckerberg ou encore Daniel Ek (Spotify) ont rédigé une lettre ouverte en pointant du doigt la nécessité d'une « certitude réglementaire en matière d'IA ».

D'après Yann le Cun : les sociétés prennent un « énorme risque à limiter l'utilisation de ces technologies » qui pourrait les mener à leur perte et devrait à la place stimuler la recherche en IA. En réalité, ce discours n'étonne en rien puisqu'il s'agit souvent de personnes dont l'objectif n'a cessé d'être de réfléchir à améliorer cette technologie. Un discours souvent empreint d'un « technosolutionnisme » bien affirmé.

L'union européenne quant à elle, croie fermement que l'intelligence artificielle doit étroitement pouvoir se développer avec une conception éthique. La conception européenne est fervente d'une conciliation possible entre l'IA et l'innovation avec la protection des droits. L'argument du RIA est de dire aux Big techs que l'innovation est tout à fait possible si le système en cause présente le moins de risques possibles. En effet, plus les risques sont faibles, moins il y a d'encadrement juridique. Malgré tout, les systèmes d'IA sur le marché vont davantage vouloir flirter avec la limite de l'interdiction. Il y a une réelle corrélation entre innovation et risques. Plus le système d'IA sera innovant, plus les risques seront potentiellement importants.

Ainsi, en appréhendant les systèmes d'IA sectoriellement suivant le degré de risque, le RIA souhaite rendre difficile la conformité des « IA à haut risque ». L'approche par les risques se fonde sur la potentialité du préjudice subi par les personnes physiques.

Tendre à assurer l'innovation et la réglementation demeure un idéal difficilement atteignable. Il s'agit davantage d'une conciliation des principes.

L'approche par les risques est donc l'approche qui s'apparente comme étant celle la plus justement conciliatrice de l'intérêt d'innover et de l'intérêt de réglementer.

Le cadre normatif proposé par la Commission se situe clairement à l'intersection de l'éthique et du droit, ce qui représente, du point de vue stratégique, la méthode la plus habile pour encourager son adoption par les citoyens européens.

En effet, le développement de l'IA en Europe dépend de son acceptabilité, laquelle ne peut être atteinte que si les systèmes d'IA sont perçus comme fiables.

Afin que l'UE puisse se positionner comme un acteur majeur dans le domaine de l'innovation, il est crucial que l'IA européenne se distingue des modèles concurrents en mettant en place un cadre réglementaire qui soutient la création d'un "écosystème de confiance unique en son genre" (livre blanc sur l'IA).

« L'Europe est-elle à un point de bascule entre surréglementation et volonté d'innover ?¹³ »

Le droit peut tout à fait stimuler l'innovation en invitant les principaux acteurs à repenser la méthode d'entraînement de leurs algorithmes. Le développement d'une IA éthique basée sur jeu de données responsables (**Chapitre 3**) intégrant des données synthétiques (**Section 1**) tout en minimisant les biais algorithmiques (**Section 2**) pourrait créer une meilleure efficacité, interopérabilité des systèmes et ce, en contribuant à l'élaboration d'une meilleure explicabilité des algorithmes (**Section 3**).

¹³ <https://www.usine-digitale.fr/article/ia-generative-l-europe-a-un-point-de-bascule-entre-surreglementation-et-volonte-d-innover.N2219037>

CHAPITRE 4 : LE DÉVELOPPEMENT D'UNE GOUVERNANCE DES DONNÉES RESPONSABLE DES « SYSTÈMES D'INTELLIGENCE ARTIFICIELLE »

Les données représentent un actif de plus en plus lucratif à condition d'en obtenir en grand nombre. Les Big techs ont ainsi une capacité de collecte, de stockage et de traitement des données telle que ce « marché des données » est estimée à 271,83 milliards de dollars¹⁴. Il faut poser des limites car la tension émerge avec la protection de la vie privée du fait de la volonté des entreprises à massifier les collectes de données.

Les entreprises cherchent à redéfinir leurs positions en mettant en œuvre des technologies capables de protéger la confidentialité des données et leur permettre de les utiliser pour leurs divers objectifs.

Le nerf de la guerre dans la bataille contre les IA à « haut risque », ce sont donc la protection des données personnelles utilisées lors de l'entraînement de leurs algorithmes. L'IA nécessite une quantité faramineuse de données. L'IA générative quant à elle nécessite souvent davantage de données culturelles que de données personnelles. Cela pose alors davantage la préoccupation de la protection intellectuelle des données récoltées.

L'évolution de ces modèles tend davantage à les rendre moins gourmand. Les systèmes d'IA mettent davantage l'accent sur la qualité des données que sur la quantité. La puissance de ces modèles tendra aussi alors à se réduire.

Il faut retenir que la qualité des données utilisées dans l'entraînement des systèmes d'IA est la solution pour obtenir une réelle fiabilité des systèmes.

A titre d'exemple, l'IA a permis la découverte d'un nouvel antibiotique contre le staphylocoque doré après avoir entraîné une IA avec des structures moléculaires existantes d'antibiotiques. Les données utilisées n'étaient qu'au nombre de 40.000.

¹⁴ <https://www.fortunebusinessinsights.com/big-data-analytics-market-106179>.

Les données d'entraînement étaient suffisamment de qualité pour permettre aux chercheurs de ne pas fournir davantage de données. De la même manière, il est plus efficace de cibler publicitairement 1000 acheteurs déjà intéressés par une marque que 10.000 personnes au hasard.

Deux difficultés peuvent être soulevées s'agissant de la difficulté d'accéder à des données, relater par les acteurs de l'IA mais aussi les chercheurs dans le domaine de la santé.

D'une part, le cadre actuel établi par le RGPD est souvent moins exigeant que des pratiques françaises en matière de récolte de données à caractère personnel. A titre d'exemple, s'agissant des données de santé, il existe une autorisation préalable qui peut impliquer jusqu'à 3 niveaux d'approbation. Il existe une procédure simplifiée mais elle est peu répandue. Pour les chercheurs, cette procédure est très contraignante contrairement au cadre de responsabilité/liberté institué par le RGPD.

D'autre part, on observe un écart croissant entre la protection des données personnelles et l'utilisation des données massives. Seul les données anonymisées échappent au régime de protection du RGPD. Les avancées technologiques rendent malheureusement de plus en plus facile la réidentification de ces données, l'anonymisation est moins fiable. Le RGPD apparaît lui aussi, difficile à appliquer face au système d'IA notamment s'agissant du responsable de traitement ou encore de la finalité du traitement. Le RGPD aurait pu mettre en avant une gestion davantage collective des données, ce qui renforcerait l'effectivité de la garantie des droits.

La génération de données synthétiques est l'une des solutions ici envisagée. Les données synthétiques constituent une conciliation favorable tant à l'innovation qu'à la protection des données à caractère personnel des individus (**Section 2**). La piste offerte par les données synthétiques vise à minimiser les risques et maximiser la conformité au regard du RGPD.

Section 1 : Les données synthétiques : une conciliation parfaite entre innovation et protection des données

Sous-section 1 : Définition de la donnée synthétique

L'OCDE propose la définition suivante : « *Les données synthétiques sont générées à partir d'un modèle construit sur des données réelles. Elles devraient être représentatives des données réelles originales* ».

Les données synthétiques ne sont plus des données réelles mais elles possèdent la même « représentativité » que les données réelles.

Le Data Act, au sein du huitième considérant précise qu'elles sont issues du : « *recours à des technologies [...] qui permettent d'appliquer des algorithmes aux données et d'obtenir des informations précieuses sans transmission de ces données entre les parties ni copie inutile des données brutes ou des données structurées elles-mêmes* ».

En effet, les données synthétiques sont obtenues de manière « artificielle ». L'objectif est de modifier leur apparence tout en conservant leur qualité.

La Commission européenne complète les usages des données synthétiques en précisant qu'elles favorisent « *l'augmentation de données qui implique des techniques permettant de générer artificiellement des points de données supplémentaires à partir de données existantes* ».

Les données synthétiques sont certes des données générées artificiellement à partir d'un ensemble de données mais elles demeurent tout aussi représentatives que les données originales. Elles peuvent même compléter un jeu de donnée qui lui est resté originel en effectuant un « suréchantillonnage synthétique ». Ainsi, s'il existe une minorité au sein de ces données, celle-ci sera maintenue dans la même proportion de représentativité, tout en augmentant le jeu de données.

Les données synthétiques peuvent donc produire des données de haute qualité avec un coût bien plus réduit que s'il fallait recourir à l'étiquetage ou à labellisation des données qui demandent des moyens financiers et techniques importants.

Exemple concret de génération de données synthétiques dans le domaine médical et dans le domaine bancaire :

Les données synthétiques dans le domaine de la santé représentent une réelle aubaine pour les chercheurs.

Les « National Institutes of Health (NIH) » des États-Unis ont collaboré avec une start-up spécialisée dans les services IT, nommé *Syntegra*. Lors de l'épidémie de la Covid, *Syntegra* a récupéré les données issues de plus de 2,7 millions de personnes et dont plus de 413 000 patients positifs ont consulté pour suspicion de COVID-19¹⁵.

La start-up a généré une copie synthétique de ces données, avec la même représentativité statistique. Cela a permis à des chercheurs des quatre coins de la planète d'utiliser ce jeu de données synthétiques pour accélérer la mise au point de traitement et vaccins, sans risque de réidentification des informations initiales.

Autre exemple, dans le domaine bancaire où ces données sont de nature extrêmement confidentielle.

Si l'on souhaite détecter grâce à l'intelligence artificielle les cas de fraude bancaire, il faudrait se baser sur des données relatant l'identification de fraudes qui se sont produites. Le problème étant que le taux de fraude bancaire s'établit à environ 0,076% en 2023¹⁶. Ainsi, la quantité de données issues de la fraude bancaire est bien trop faible pour entraîner correctement un système à les détecter efficacement.

Une intelligence artificielle (génératrice de données synthétiques) augmenterait ainsi les cas de fraude bancaire. Par-exemple, elle générerait des virements bancaires à des

¹⁵ <https://www.accenture.com/fr-fr/insights/artificial-intelligence/synthetic-data-speed-security-scale>

¹⁶ https://www.banque-france.fr/system/files/2024-02/240125_OSMP-Statistiques-de-fraude-du-S1-2023.pdf

horaires inhabituelles, avec des localisations suspectes (New-York, Japon...) et avec des montants anormalement élevés etc...

L'objectif est tout simplement de fournir des données supplémentaires lorsque le système n'en a pas assez pour s'entraîner. Les entreprises peuvent alors accélérer les phases de test et de déploiement de leur solution d'IA.

Les raisons d'un accès restreint à certaines données peuvent être très diverses. Dans ces deux exemples, les dossiers médicaux des personnes ou leurs informations bancaires sont des données sensibles et confidentielles. Les protections et les contrôles sont souvent si exigeants qu'il s'agit d'un véritable parcours du combattant pour accéder à ces dites données. Les délais pour les obtenir peuvent être long (protocoles de sécurité) et le nombre de données que les entreprises sont autorisées à traiter dans des quantités infimes au regard de la nécessité d'entraînement d'un système d'IA. Ainsi, l'obtention des données et leur utilisation devient tout aussi difficile que chronophage.

Les données synthétiques représentent donc une option pertinente et sécurisée pour l'entraînement des systèmes d'intelligence artificielle (**Sous-section 2**)

Sous-section 2 : Donnée synthétique et protection des droits attachés à la personne

Le premier avantage des données synthétiques, c'est leur contribution à protéger efficacement les données personnelles et ou sensibles. S'agissant des méthodes de chiffrement ou d'anonymisation, elles sont certes efficaces, mais tant que les données originelles sont présentes, il y a un risque de pouvoir les réidentifier.

Cette meilleure protection des données personnelles s'explique tout simplement par le fait que ces données sont purement fictives. Elles ne sont pas rattachées à une personne. C'est une donnée de la donnée. Les méthodes classiques obligent à intervenir directement sur les données originelles. La modification des données originales est permise en ajoutant ou en modifiant la donnée brute.

Prenons l'exemple de l'interview d'une personne dénonçant des faits de pédophilie à la télévision mais qui souhaite garder l'anonymat. Avec les méthodes classiques de chiffrement c'est un peu comme si l'on flouter le visage de cette personne et qu'on remplace sa voix par la fameuse voix « robotique ». Avec la conception des données synthétiques, le témoin ne serait pas du tout interviewé. On prendrait à la place une personne qui retranscrirait au mot près ce que souhaitais dire le témoin. Cet exemple illustre qu'il est en effet « impossible » de retrouver l'identité du témoin et pourtant, le contenu même du discours est exactement le même

Pour les données synthétiques, c'est exactement la même chose. Les données sont issues, tout comme le discours de la tierce personne, à partir des données réelles. La seule corrélation qui existe est la représentativité des données initiales.

Ce qu'il faut retenir, c'est que les données personnelles issues des données originales ne sont pas intégrées dans la génération des nouvelles données. Les données synthétiques remplacent donc totalement les données d'origine, tout comme la personne ayant remplacé le témoin lors de l'interview.

Ainsi, à titre d'illustration, un projet nommé « SYNTHIA » du Health Data Hub¹⁷ a utilisé des données synthétiques dans le cadre des informations du système national des données de santé (SNDS).

A l'origine, l'intérêt d'utiliser de données synthétiques était surtout présent lorsqu'il s'agissait de données de santé. Leur utilisation s'est progressivement étendue à d'autres secteurs d'activité.

Un autre avantage est que les données synthétiques représentent également des gains en termes de scalabilité, de rapidité et d'innovation (**Sous-section 3**)

¹⁷ <https://www.health-data-hub.fr/projets/generation-dun-jeu-de-donnees-synthetiques-partir-des-donnees-dusnds-synthia>.

Sous-section 3 : Données synthétiques : scalabilité et équitabilité

Les données synthétiques inspirent la confiance auprès des entreprises. Elles se savent en conformité avec la protection des données des personnes. Malgré tout, si le seul avantage des données synthétiques n'était qu'une assurance de conformité avec les réglementations européennes, les entreprises n'y verraient sûrement qu'une moindre utilité à les générer.

La scalabilité permise par les données synthétiques est ainsi l'un des avantages les plus disruptif. En réalité, la scalabilité est une résultante qui découle directement du fait de la sécurité et de la rapidité permise par les données synthétiques. Cela permet aux entreprises de traiter des volumes de données plus conséquents. Souvent, les entreprises sont prisonnières des seules données dont elles disposent. Les données synthétiques leur permettent de compléter les données dont elles disposent déjà.

La mise en commun de données synthétiques permettrait ainsi cette scalabilité. Actuellement, les données sont « sectorisées » de sorte que chaque entreprise possède des données et qu'il est souvent difficile de les rassembler et d'en tirer un avantage quelconque. Les spécificités d'une entité à une autre pouvant être très différentes. Une banque par-exemple travaille indépendamment des autres banques puisqu'elle est la seule autorisée à analyser ses données pour y identifier d'éventuelles activités suspectes. Cette exigence est notamment instaurée par les organismes de régulation eux-mêmes. Dès lors, l'utilisation de données synthétiques permettrait à chacune des banques de pouvoir librement les partager. Cette mise en commun de chaque jeu de donnée avec leurs compères du secteur leur offrirait une vision à 360° de l'ensemble des personnes interagissant avec des banques de tel ou tel pays. Le processus de détection des fraudes s'y verrait grandement amélioré. L'identification des fraudes serait accéléré tout en permettant d'en identifier davantage.

A l'heure actuelle, cette idée de mise en commun n'est pas encore instituée au sein de notre manière de fonctionner. Si l'ensemble des institutions publiques ou encore les établissements de santé mettaient en commun de telles données, de grandes opportunités se créeraient. Le domaine de la santé aurait accès à jeu de données qui lui permettrait de surmonter des défis jusqu'alors insurmontables.

Le partage des données y serait facilité et créerait une réelle économie des données.

Les données synthétiques peuvent également jouer un rôle dans la création d'ensembles de données plus équilibrés. Pour ce faire, on va contraindre le processus d'optimisation à respecter une contrainte telle que celle de l'équité. Les données résultantes auront été générées en conformité avec la contrainte d'une équité stricte dans leur représentabilité.

Le RIA au sein de son article 10 sur la « Gouvernance des données » précise que : « *Les ensembles de données de formation, de validation et d'essai sont pertinents, suffisamment représentatifs et, dans la mesure du possible, exempts d'erreurs et complets au regard de l'objectif visé* » et ce en recourant à des « *opérations de traitement de préparation des données pertinentes, telles que l'annotation, l'étiquetage, le nettoyage, la mise à jour, l'enrichissement et l'agrégation ;* »

Les données synthétiques vont donc enrichir le jeu de données, maintenir la représentativité des diverses parties tout en réduisant les biais pouvant avoir une incidence sur une population spécifique.

Ainsi, ces ensembles de données ajustés peuvent être utilisés pour entraîner des modèles sans avoir recours à des approches d'atténuation des biais qui peuvent parfois compromettre la précision. À titre d'exemple, l'entreprise autrichienne MOSTLY AI a démontré que des données synthétiques conçues pour être équitables pouvaient remédier aux biais d'un jeu de données utilisé par un logiciel nommé COMPAS.

Ce logiciel est notamment employé dans des systèmes judiciaires américains pour évaluer les risques de récidive d'un accusé et il avait produit des résultats discriminatoires, notamment envers les Afro-Américains. Grâce à son approche, MOSTLY AI a réussi à réduire considérablement cet écart, le ramenant à seulement 1 %, et ce, tout en préservant la précision prédictive des modèles.

Actuellement, les données synthétiques sont encore bien trop cantonnées au domaine de la data science et rencontrent divers obstacles expliquant une utilisation encore peu répandue (**Sous-section 4**)

Sous-section 4 : Les données synthétiques, un levier d'innovation difficile à actionner

L'intérêt de recourir à des données synthétiques est de permettre une expérimentation interne facilitée et donc une innovation plus efficace. Les données synthétiques n'étant pas soumises aux mêmes protections juridiques, il est plus facile pour les entreprises de réaliser des expérimentations. Les données synthétiques peuvent se comparer au « bac à sable règlementaire » institué par le RIA. L'expérimentation y est facilitée.

Malgré tout, une première raison à la difficulté d'actionner ce levier est que, même si les données synthétiques ne sont pas soumises à des protections juridiques spécifiques, le jeu de donnée d'origine lui est soumis à ces protections. Un premier obstacle, qui peut l'être pour certaines entreprises, est celui de veiller à ce que la génération des données synthétiques s'effectue sur des données d'origines dont leur traitement est licite et légitime.

Ainsi, dès lors que le traitement initial est légitime et que toutes les protections juridiques ont été respectées, les données synthétiques ne seront en principe soumises à aucune protection juridique. Respecter cette première condition permettra à l'entreprise de trouver dans ces données générées un réel avantage tant du point d'un point de vue de l'innovation que d'un point de vue commercial (revente de données « *safe* »).

La réelle difficulté est davantage de l'ordre technologique et technique que de l'ordre juridique. En effet, générer des données synthétiques ne se fait pas simplement en effectuant quelques prompts sur ChatGPT en y insérant les ensembles de données de l'entreprise. Pour générer des données synthétiques qualitatives il faut en amont posséder des connaissances poussées en intelligence artificielle. Il faut aussi établir un cadre détaillé et rigoureux qui permettra à l'entreprise de générer les données effectivement attendues. Cela paraît relever d'une évidence pure mais ce point demeure essentiel. Une entreprise chargée de la production de données synthétiques doit pouvoir démontrer explicitement à son client que ces dites données reflètent avec exactitude les données d'origine. Sinon, comment prouver qu'il s'agit bien de données synthétiques issues des données d'origine ? Sans cette explication, il y aurait des dérives qui amèneraient à des risques importants suivant la nature des données. Un

établissement de santé pourrait ainsi créer un système d'IA totalement déphasé de la réalité et engendrait tous les risques imaginables.

Le potentiel des données synthétiques reste encore largement sous-estimé, avec un important travail de démocratisation à mener pour qu'elles soient perçues par les diverses institutions non plus comme une curiosité excentrique, mais comme une solution concrète et prometteuse.

Tout comme l'IA, les données synthétiques ont une faiblesse, elles ne sont pas adéquates pour toutes les utilisations. Les données synthétiques sont très utiles lorsqu'il s'agit de données sensibles que l'on souhaite partager tel que les données de santé.

Lorsqu'il s'agit d'IA générative, c'est une tout autre finalité souhaitée. Il est nécessaire d'avoir les données les plus diversifiées possibles. Les données synthétiques peuvent empoisonner une base de données. Un cercle vicieux s'installe où les modèles sont nourries de leur propre probabilité et produiront alors un contenu de plus en plus uniforme. Yannick Meneceur évoque à juste titre ce « *besoin vital de nouvelles données non artificielles pour prévenir l'effondrement des modèles*¹⁸ ».

Une autre difficulté demeure, celle du biais algorithmique, qui peut facilement s'instiller dans les modèles d'IA entraînés sur des jeux de données qui contiennent des biais humains historiques profonds. Ainsi, créer un jeu de données qui est simplement la copie de l'original engendrera tout simplement la duplication des biais préexistants.

Il est donc tout à fait primordial de s'assurer à ce que les systèmes d'IA, tiennent compte des biais algorithmiques et génèrent un jeu de donnée le plus éthique et représentatif (**Section 2**)

¹⁸ Yannick Meneceur – « IA générative et professionnels du droit » Lexis Nexis

Section 2 : Le talon d'Achille de l'IA : exploration des biais algorithmiques dans l'intelligence artificielle

Les biais algorithmiques sont aux algorithmes ce que les vers sont à la pomme. Ils constituent des éléments indésirables qui « empoisonnent » la qualité d'un ensemble de données.

Concrètement, les biais algorithmiques sont des erreurs récurrentes et systématiques dans les algorithmes de machine learning et qui produisent des résultats discriminants et injustes. Ils sont souvent issus de préjugés socioéconomiques, ethniques et sexistes préexistants.¹⁹

Les biais algorithmiques peuvent apparaître de plusieurs manières et à différents stades de développement de l'algorithme.

La première cause est lorsque les données d'entraînement de l'algorithmes sont biaisées. Ces données « erronées » feront produire à l'algorithmes des résultats biaisés qui seront amplifié. Ceci est dû au fonctionnement des algorithmes et de la boucle de rétroaction. Un cycle dans lequel l'algorithme apprend et répète en permanence les mêmes modèles biaisés ce qui rend l'algorithme de plus en plus « mauvais ».

Il est aussi possible que les biais apparaissent non pas par l'existence de données de mauvaises qualité (données non représentatives comportant des biais sociaux) mais par une mauvaise corrélation entre les données qui lui sont fournies. Ce sont les biais de corrélation. Un algorithme entraîné avec des données de qualité peut produire des biais par le simple fait que les liens entre les données sont complètement faux. Par exemple, un algorithme pourrait établir un lien de causalité entre l'augmentation des ventes de glaces et l'augmentation des attaques de requins. En réalité, les glaces sont majoritairement consommées l'été et les attaques de requins surviennent l'été lorsque les personnes vont davantage se baigner, donc en été. Un humain comprend rapidement que ce lien de corrélation est peu sérieux. Pour autant, l'algorithme lui est capable d'effectuer ce lien car il n'a pas pris en compte que d'autres facteurs peuvent être plus importants que les données en elle-même.

¹⁹ <https://www.ibm.com/fr-fr/think/topics/algorithmic-bias>

Les biais algorithmiques peuvent aussi apparaître lors de la conception des algorithmes. Le plus souvent, le concepteur de l'IA va créer un programme en introduisant dedans des règles subjectives basées sur ses propres préjugés. La pondération peut aussi en être la cause. Le concepteur de l'IA va mal pondérer certains facteurs dans la prise de décision dans le système d'IA.

Les proxys dans les systèmes d'IA peuvent aussi introduire d'importants biais. Leur utilisation est très fréquente mais nécessite une attention toute particulière. Les proxys sont utilisés pour faciliter l'entraînement des modèles. Ils remplacent des variables complexes par des variables corrélées plus simples.

Par-exemple, utiliser le code postal en proxy pour le statut socio-économique ou employeur le nombre de clics sur une publicité comme proxy pour l'engagement de l'utilisateur.

En effet, il est difficile de représenter la « satisfaction du client » ou le « bonheur » au sein d'un système. Malgré tout, on distingue bien la dangerosité d'un tel outil, certes pratique. Ainsi, un système pourrait utiliser le revenu d'un client d'une banque pour l'accès au crédit. Il pourrait utiliser le proxy du « nombre de visite à l'hôpital » pour établir « l'état de santé des patients » ou comme cité précédemment, utiliser le code postal pour avantager ou désavantager certains groupes ethniques.

Les biais peuvent aussi apparaître lors de l'étape de l'évaluation, il s'agit alors de biais d'évaluation. Ces biais ne sont pas présents dans le système à proprement parler. C'est tout simplement une interprétation fautive des résultats produits par le système par une personne qui les interprète dans un sens tout autre.

Les exemples illustrant les préjudices qu'entraînent l'existence de biais algorithmiques ne sont pas rares.

En 2021, le gouvernement néerlandais a été démis à la suite de l'existence d'algorithmes discriminatoires et où Amnesty internationale avait dénoncé une utilisation de « machines xénophobes » dans le secteur public. Il avait ciblé de façon

péjorative des milliers de familles comme ayant commis une fraude sociale et leur a demandé de restituer d'importantes sommes d'argent. Les familles en question étaient majoritairement des foyers à faibles revenus et davantage de minorités ethniques.

La même année, Amazon a lui aussi été contraint d'abandonner son outil de recrutement qui discriminait systématiquement les candidates féminines.

Souvent lorsque l'on veut éviter des biais algorithmiques il arrive d'en créer d'autres ou de créer du contenu non désiré.

Un exemple l'illustrant est notamment lié à l'algorithme de Google Gemini, dans son effort pour prendre en compte les problèmes d'égalité raciale, qui a fini par produire des résultats inverses. Plusieurs utilisateurs se sont plaints de cette tendance à la diversité, la percevant comme une forme de racisme antiblanc ou de propagande « woke ».

Un dernier exemple est l'expérimentation de justice prédictive aux États-Unis avec l'algorithme COMPAS, évoqué précédemment, qui s'était révélé particulièrement « raciste ». Elle attribuait un taux de récidive deux fois supérieur pour les afro-américains tandis que pour les Blancs, ce risque était sous-estimé.

Les biais peuvent être présents à tous stades de l'élaboration d'un système d'IA et peuvent être présents dans tous les secteurs dont une IA est utilisée. Les systèmes de reconnaissance faciale, de recrutement, de police prédictive, de génération d'images...

Le droit doit pouvoir appréhender cette problématique souvent inhérente et parfois inévitable que représente les biais algorithmiques dont ceux « *implicites*²⁰ » et « *explicites*²¹ ».

²⁰ Préjugés ou des stéréotypes inconscients qui influencent leurs comportements et décisions sans que les individus ne le perçoivent, plus difficiles à détecter.

²¹ Préjugés ou des stéréotypes dont les individus sont conscients et qu'ils expriment ouvertement

Sous-section 1 : La réglementation à l'égard des biais algorithmiques

Le RIA entend bien mettre fin à ces pratiques discriminatoires et à lutter contre l'apparition récurrente des biais algorithmiques.

Le règlement insiste bien sur la nécessité de respecter les droits fondamentaux et de lutter contre les biais discriminatoires. L'article 10 du RIA dispose qu'il faut mettre en place « des mesures appropriées pour détecter, prévenir et atténuer les éventuels biais conformément au point f) ».

Malgré tout, même si la détection et la prévention des biais systémiques est évoqué par la réglementation, en réalité il n'existe pas de moyen de mesurer précisément la discrimination sauf aux États-Unis où le « disparate impact » est juridiquement consacré.

A cet égard, certains chercheurs issus du fair machine learning (FairML) ont mis en place des outils tels que les *erreurs conditionnelles*²², les *rapports de cote conditionnels*²³ ou la *parité statistique/démographique*²⁴ pour prévenir des risques et identifier l'existence de biais algorithmiques.

A noter que le RIA n'adopte à ce sujet aucune méthode technique et concrète pour lutter contre ces biais, contrairement aux États-Unis.

Comme le dit Annaëlle Martin, docteur en droit public – éthique du numérique : « *Si l'équité ne se résume pas à une définition juridique, elle ne se réduit pas non plus à une équation mathématique. Pas plus que les mathématiques, le droit n'offre de*

²² Calcul d'un taux d'erreur de prévision sur la probabilité de se tromper pour le groupe sensible (minorité ethnique) sur la probabilité de se tromper pour l'autre groupe.

²³ Rapports des taux de faux positifs du groupe sensible (minorité ethnique) sur le taux de faux positifs de l'autre groupe et rapport des taux de vrais positifs.

²⁴ La parité statistique exige que la réalisation de la prédiction ne soit pas corrélée à une variable sensible (comme l'origine ethnique). Le système doit donner un résultat favorable à chaque groupe démographique.

réponse toute faite lorsqu'il s'agit d'appréhender les risques inhérents aux systèmes d'IA caractérisés par leur opacité et leur complexité croissantes ».

Ce postulat évoqué par la doctorante, illustre la difficulté et souvent l'impuissance du droit à identifier les causalités et chercher la responsabilité des différents acteurs impliqués dans la chaîne décisionnelle.

Dès lors, comment assurer l'équité des algorithmes qui ne fait même pas l'objet d'une définition au sein du RIA. Assurer l'équité algorithmique c'est tout simplement éliminer les discriminations.

Pour accomplir cette tâche, il faudrait alors procéder à établir des discriminations positives pour éviter que les modèles ne renforcent les biais sociaux. Le problème est juridique, les discriminations positives sont interdites dans la plupart des états européens.

Ainsi, comment établir une responsabilité des acteurs s'il est difficile d'identifier la causalité de ces derniers face au préjudice observé. Le RIA a fait le choix pragmatique de rendre responsable le fournisseur lors de la mise sur le marché et ou service d'un système d'IA à haut risque.

Sous-section 2 : Une solution partielle offerte par la directive 2024/2853

Récemment, un grand changement de paradigme a eu lieu avec la directive 2024/2853²⁵ en étendant la responsabilité du fait de produits défectueux aux logiciels et donc aux systèmes d'IA. Initialement, la directive ne visait seulement les biens meubles.

L'article 4 de la directive 2024/2853 définit la notion de produit et précise que cela : « *comprend l'électricité, les fichiers de fabrication numériques, les matières premières et les logiciels* ».

²⁵ https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=OJ:L_202402853

A cet égard, le champ des dommages réparables s'est agrandi, comprenant aussi les dommages du fait de la : « destruction ou la suppression ou la corruption de données non utilisées ».

Le réel apport de cette directive réside dans l'élargissement de la responsabilité des opérateurs économiques. L'article 8 de la directive 2024/2853 redéfinit la responsabilité des fabricants de produits ou services d'IA en voyant leur responsabilité engagée en cas de défectuosité du système d'IA inclut dans le produit ou le service.

Certains diront qu'en théorie même si la responsabilité est élargie en visant les « acteurs de l'IA », en réalité la complexité de prouver un préjudice du fait d'un défaut dans un système d'IA rend cette responsabilité peu opérante.

A cet égard, un assouplissement du lien de causalité a été effectué.

L'article 10 de la directive dispose que lorsque le demandeur a des : « difficultés excessives » du fait de « la complexité technique ou « scientifique » à prouver la « défectuosité du produit », le juge pourra présumer le lien de causalité entre la défectuosité et le dommage.

L'Union européenne adopte ainsi une directive qui suit, tout comme le RIA, cette approche par les risques en assurant une utilisation éthique des systèmes d'IA. Les biais d'algorithmes pourront s'analyser comme représentant une défectuosité du produit.

L'existence des biais algorithmiques s'expliquent aussi du fait de l'opacité des algorithmes souvent nommés « black box ». Ainsi, la transparence des algorithmes (**Section 3**) représente l'une des conditions nécessaires à l'éradication totale des biais dans les algorithmes.

Section 3 : Le cheminement complexe de la transparence algorithmique

L'algorithme est un peu comme la recette de cuisine de notre grand-mère. Elle est très secrète. Les géants de la tech n'ont pas du tout à cœur de la dévoiler. Dès lors, la volonté d'une meilleure « transparence des algorithmes » n'est pas vraiment inscrite dans leurs objectifs. L'argument de défense est souvent celui de dire que leurs algorithmes sont devenus bien trop complexes pour pouvoir les expliquer. En réalité, il y a une part de vérité. Les neurones profonds sont d'une technicité telle qu'il est souvent difficile d'expliquer les causes ayant produit telle sortie. On utilise souvent le terme « boîte noire » pour désigner le fait qu'on ne saurait expliquer les raisons exactes d'une décision d'un algorithme.

Le RIA intègre plusieurs mesures visant à renforcer la transparence globale des algorithmes (**Sous-section 1**). Cette dynamique évolue progressivement vers une approche axée sur l'explicabilité des algorithmes (**Sous-section 2**)

Sous-section 1 : L'appréhension du concept de transparence par le RIA

L'article 50 du RIA est l'un des seuls articles du règlement qui évoque explicitement le mot « transparence » avec l'article 13. Une nuance doit être faite à ce sujet. La transparence évoquée au sein de l'article 50 ne fait pas référence à l'objectif d'une transparence des algorithmes.

Ainsi, l'article 50 érige des « obligations de transparence pour les fournisseurs et les déployeurs de certains systèmes d'IA » quant à l'information communiquée aux utilisateurs. Les fournisseurs d'IA ont pour obligation d'indiquer explicitement aux utilisateurs lorsqu'ils interagissent avec une IA sauf si cela apparaît de manière flagrante.

L'article 50 vise davantage à encadrer les dérives qui amèneraient l'utilisateur à croire à tort que le contenu provient d'une source humaine alors qu'il s'agit en réalité d'un contenu généré artificiellement.

Le RGPD s'est lui attelé plus profondément à cette problématique au sein de son article 22. LE RGPD impose d'informer la personne lorsqu'une décision entièrement automatisée a été prise à son encontre mais donne aussi le droit de connaître la logique et les critères employés pour prendre la décision. Là encore, lorsqu'il s'agira d'un algorithme basé sur les neurones profonds, difficile de respecter à la lettre les droits évoqués par le RGPD

L'article 50 offre une certaine transparence minimale. Il n'a pas pour finalité d'offrir une réelle transparence algorithmique à l'égard des utilisateurs. Cela peut s'expliquer car il concerne surtout les systèmes qui ne sont pas considérés comme à haut risque

En effet, la réelle transparence algorithmique abordée par le RIA concerne essentiellement les systèmes d'IA à haut risque.

L'article 11 du RIA oblige aux fournisseurs d'IA à haut risque d'ériger une documentation technique. L'article indique que ladite documentation doit contenir au minima les informations indiquées au sein de l'annexe IV.

L'annexe IV décrit de manière extrêmement détaillée divers obligations de mentions qui peuvent se rattacher à cet objectif de transparence.

Il s'agit notamment d'être transparent sur l'entraînement des systèmes d'IA en mentionnant : « *les méthodes et les étapes effectuées pour le développement du système d'IA* » (annexe IV 2.a).

La transparence algorithmique pourrait être ainsi assurée en diffusant « *les spécifications de conception du système* » tout en explicitant « *la logique générale du système d'IA et des algorithmes* » en allant jusqu'à expliquer « *les principaux choix de conception, y compris la justification et les hypothèses faites* ».

Le RIA aborde ici pleinement la question de la transparence. L'article 10 mentionne par-ailleurs l'obligation que la documentation technique comporte aussi des mesures visant à faciliter l'interprétation des résultats produits par les systèmes d'IA à l'égard des déployant. Cette exigence est relative à l'article 13 du RIA.

En effet, il serait paradoxal d'assurer une transparence non comprise par les déployeurs des systèmes d'IA. L'article 13 du RIA impose alors que l'ensemble des informations communiquées aux déployeurs d'IA soient compréhensibles. En réalité, il faut voir l'article 13 comme l'obligation pour les fournisseurs d'IA d'établir un manuel d'information et d'utilisation du système d'IA. L'article aborde tant l'obligation de fournir les performances et les robustesses du système que les dangers connus dû à son utilisation.

L'article 49 du RIA traite quant à lui de l'obligation pour le fournisseur d'une IA à haut risque d'inscrire son système dans la base de données de l'union. Le point intéressant dans l'article est le fait que cette inscription s'effectue en amont de toute commercialisation (point 1 Art. 49). Cette obligation d'enregistrement est en lien avec l'article 71 du RIA relatif à la « base de données de l'UE pour les systèmes d'IA à haut risque ».

Un dernier article mérite d'être mentionné, il s'agit de l'article 12 du RIA qui prévoit l'obligation de tenir des registres. C'est l'article qui contribue le plus à l'objectif de transparence des algorithmes. Ce registre pourrait se comparer à la « boîte noire » d'un avion. Il enregistre automatiquement chaque événement pendant toute « *la durée de vie du système* ». Il y a donc une réelle « *traçabilité* » pour identifier les « *situations qui peuvent entraîner un risque* ». Le RIA utilise la transparence au service d'une vision préventive des risques.

Un autre point intéressant se situe au 3)c de l'article 12 disposant que les systèmes fournissent : « les données d'entrée pour lesquelles la recherche a abouti à une correspondance ». Il s'agit en réalité d'améliorer la traçabilité des décisions prises par un système d'IA.

L'article vise à garantir une trace numérique lorsque par-exemple l'image d'un visage capturé par une caméra (une donnée d'entrée) trouve une correspondance à un profil dans la base de donnée.

En réalité, le RIA tente d'aborder une transparence algorithmique sur tout le long de la chaîne des acteurs intervenant dans la mise sur le marché d'un système d'IA. Il est

malgré tout regrettable que le RIA n'ait pas traité de la transparence dans une même section.

Il s'agira de vérifier si les dispositions relatives à la transparence n'entraîneront pas, en pratique, une certaine réticence des acteurs à fournir tant d'informations.

Le RIA est davantage dans une démarche de transparence des produits sur un marché. Tout comme il est obligatoire de mentionner sur un pot de confiture la liste entière des ingrédients et les éventuels additifs.

La réelle critique réside dans le fait que le RIA traite de la transparence de manière éclatée. Le règlement a établi certaines mesures de transparence algorithmique mais cet objectif apparaît caché entre les lignes et non explicitement au sein du RIA.

Malgré tout, les dispositions du RIA tendent non seulement vers une transparence de l'information (art. 50), une transparence algorithmique mais surtout vers une explicabilité algorithmique (**Sous-section 2**).

Sous-section 2 : Le cheminement souhaitable vers l'explicabilité algorithmique

Si certains secteurs imposent des obligations telles que la publication du code source, une partie de la doctrine suggère de dépasser cette approche purement informative en faveur d'une véritable « explicabilité » des algorithmes.

L'explicabilité consiste en une meilleure compréhension interne du fonctionnement des algorithmes. Cette obligation prend tout son sens lorsqu'il s'agit de « justice prédictive » où il est simplement inconcevable d'imaginer qu'un tel procédé émerge sans pouvoir expliquer la raison qui l'aurait mené à telle ou telle décision de justice.

La performativité de la justice prédictive qui « désigne l'absence de barrière étanche entre le fait de dire – présenter la jurisprudence – et d'agir – trancher le litige » pourrait s'illustrer par l'image et le risque d'une « prophétie auto-réalisatrice »²⁶.

L'objectif derrière l'explicabilité algorithmique n'est pas de tomber dans une transparence totale des algorithmes. En réalité, comme dit précédemment, la plupart des « data scientist » ne sauraient expliquer précisément le fonctionnement des algorithmes qu'ils peaufinent.

L'objectif final est de pouvoir expliquer à l'utilisateur les facteurs majeurs intervenant dans la prise de décision, des liens significatifs entre les variables et surtout d'expliquer la logique et les règles guidant les choix du système. La confiance des utilisateurs serait renforcée en leur expliquant concrètement les implications d'une telle utilisation tout en intégrant l'explicabilité.

Ce défi est notamment à mener à l'égard des algorithmes publics (ex : algorithmes d'attribution des aides sociales) et dont le département « Etalab » se charge de réaliser cet accompagnement des administrations publics.

Camille Girard-Chanudet dans le podcast « Le code a changé » sur « les dames de l'algorithme », explique que l'usage d'algorithmes par les administrations publiques reste encore opaque. Ainsi, l'obligation inscrite à l'art. L.312-1-3 du CRPA est rester en quelque sorte « lettre morte » depuis 2016.

Des initiatives de certains chercheurs et doctorants, dont Camille Girard-Chanudet, vise à répondre à cette inaction administrative. Ce sont donc trois femmes, Camille Girard-Chanudet, Estelle Hary et Soizic Pénicaud qui ont réalisé un inventaire qui répertorie les algorithmes publics²⁷

De l'autre côté, Yann Le Cun évoque lui l'argument inverse. Il interroge sur la réelle nécessité d'expliquer des algorithmes. Selon lui, personne ne sait précisément

²⁶ « Quelle transparence pour les algorithmes de justice prédictive » *Élise Mouriesse* - situation où une croyance ou une prédiction, même initialement fausse, finit par se réaliser simplement parce qu'elle est crue et qu'elle influence les comportements

²⁷ <https://odap.fr>

comment fonctionne notre cerveau alors pourquoi cela devrait être le cas pour les algorithmes.

Comme l'explique Etienne Ollion²⁸, l'importance réside davantage dans le bon codage du modèle. Utiliser des outils que l'on ne comprend pas, appelé des boîtes noires est un problème mais qui peut être relativisé.

Si l'on est capable de démontrer et prouver extérieurement que le modèle effectue une bonne prédiction, que le modèle soit capable générer ce résultat par de bonnes inférences statistiques ou « qu'il lit dans les entrailles de boyaux de chats », c'est suffisant.

L'essentiel étant qu'on réussisse à lui faire produire ce qu'on souhaite qu'elle produise.

L'explicabilité des algorithmes n'est donc pas encore chose acquise. Certains se contentent d'une transparence minimal.

Pourtant, l'explicabilité algorithmique permettrait de mieux détecter les biais algorithmiques et les erreurs au sein des algorithmes.

L'intérêt n'est pas uniquement situé à l'égard des utilisateurs. Les fournisseurs de systèmes d'IA ont aussi tout intérêt à gagner leur confiance.

²⁸ Chercheur-professeur en sciences sociales à Polytechnique

CONCLUSION

Marie-Laure Denis, présidente de la CNIL a notamment déclaré récemment dans un entretien avec le journal le Monde que : « *La protection des données personnelles ne doit pas être la variable d'ajustement de l'IA* ». ²⁹

Cette phrase traduit d'une part un constat, et d'autre part un défi à relever pour les juristes.

Premièrement, le droit doit pouvoir encadrer l'émergence souvent des évolutions technologiques, notamment ici l'intelligence artificielle.

Deuxièmement, l'évocation du droit comme la « variable » traduit aussi cette idée que le droit se fait souvent dépasser voire submerger par la technologie. Le droit apparaît souvent comme inefficace ou inadaptée.

Emmanuelle Legrand, magistrate ayant contribué à l'élaboration du RIA a rappelé cette difficulté du droit : « N'oubliez pas que les progrès de l'intelligence artificielle avancent bien plus vite que le droit »

Légiférer sur ce qui est ou légiférer sur ce qui va être ? A ce sujet, « légiférer sur ce qui va être » peut sembler relever de l'utopie de prédire l'avenir en encadrant par le droit ce qui pourrait arriver.

En réalité, il est possible de prévoir, dans un avenir certain, les éventuels risques susceptibles de se produire lorsqu'une technologie n'est pas encore à son paroxysme mais que celle-ci s'y approche à grand pas.

Le droit ne doit donc plus uniquement légiférer sur ce qui existe (ce qui est déjà souvent une difficulté), il doit aussi pouvoir légiférer sur ce qui est en devenir. Il vaut mieux prévenir que guérir dit-on souvent. Le droit doit s'inscrire dans cette vision.

²⁹ Citation tirée d'un article du journal « Le Monde » dans la rubrique « Entretiens de l'IA » interviewant Marie-Laure Denis, président de la CNIL.

L'émergence de l'intelligence artificielle illustre cette difficulté que le RIA tente de surmonter en ayant une réelle approche préventive et prédictive des usages que pourraient être fait de l'intelligence artificielle.

Il est pourtant difficile d'affirmer qu'une réglementation (telle que le RIA) assurera effectivement un encadrement éthique des intelligences artificielles. Des difficultés techniques demeurent. Le prompt injection (formulation des requêtes) peuvent contourner les mesures de sécurité imposées par telle ou telle réglementation.

Tout comme il n'y a pas de consensus autour de la définition de l'IA, il n'y a pas de consensus mondial sur l'élaboration d'un système de valeurs universelles de ce qu'est l'éthique et de pouvoir l'intégrer dans les systèmes d'IA.

Pour « rêver d'IA intrinsèquement éthiques³⁰ », l'éthique devra donc être intégré dans les IA elles-mêmes et surtout dans une gouvernance universel adéquate. \$

Il y a une réelle interrogation s'agissant de l'universalité de l'éthique sur l'IA³¹ : « *Dans quel sens est employé le terme éthique ?* ». Cette difficulté d'avoir le même référent moral empêche de pouvoir élaborer un débat interculturel et international sur les différentes réglementations.

La réelle transparence algorithmique ne sera quant à elle réellement assurée que lorsque l'on obligera les acteurs à trouver des solutions concrètes à l'explicabilité de leurs algorithmes tout en préservant le secret de leur recette.

Même si les modes de régulation sont encore difficiles à établir (contexte d'utilisation variable, multitudes d'acteurs...) le droit, les juristes, les ingénieurs ou les simples utilisateurs ne doivent pas tomber dans le piège du technosolutionnisme ou du techno-idéalisme.

³⁰ https://www.lemonde.fr/sciences/article/2024/12/11/peut-on-rever-d-ia-intrinsiquement-ethiques_6442840_1650684.html

³¹ <https://hal.science/hal-03605668/document>

Le philosophe allemand Günther Anders (1902-1992) relatait déjà la problématique qu'induit la critique de la technologie : « *Rien ne discrédite aujourd'hui plus promptement un homme que d'être soupçonné de critiquer les machines* ».

Patrick Lecomte, professeur à l'Université du Québec, interroge cette problématique : « *Est-il encore possible d'être critique à l'égard de l'intelligence artificielle sans être qualifié de fou ?* »³²

Le droit, dans les pays démocratiques, trouve en partie son émergence des opinions sociétales, non censurées. Si même la critique s'efface devant l'oppression expansive des empires digitaux, elle légitimera leur hégémonie et leur course effrénée vers l'innovation dénué de garde-fou.

L'innovation, portée par les acteurs de l'IA, constitue une dimension que le droit ne peut certes ignorer. Toutefois, cela ne doit pas conduire à une dilution du pouvoir juridique sous prétexte de l'innovation.

Ce travail de recherche amène à la conclusion suivante : réussir à être suffisamment technocritique, pour éviter le technosolutionnisme, sans pour autant tomber dans une forme de technophobie.

Le droit doit pouvoir suivre cette vision, en créant une harmonisation entre réglementation et innovation.

³² https://www.lemonde.fr/idees/article/2024/12/13/est-il-encore-possible-en-2024-d-etre-critique-a-l-egard-de-l-intelligence-artificielle-sans-etre-qualifie-de-fou_6446212_3232.html

Sources web :

- <https://blogdroiteuropeen.com/wp-content/uploads/2022/05/wp-ia-act-final-.pdf>
- https://www.lemonde.fr/idees/article/2024/09/11/intelligence-artificielle-l-innovation-ne-doit-pas-se-faire-au-sacrifice-de-l-ethique-et-de-la-securite_6313697_3232.html
- <https://www.lemagit.fr/actualites/366610004/AI-Act-lessentiel-sur-le-bureau-europeen-de-lIA>
- <https://www.banquedesterritoires.fr/le-cadre-juridique-de-lintelligence-artificielle-encore-loin-detre-finalise>
- <https://es.weforum.org/agenda/2022/11/escalar-la-ia-por-que-hay-que-invertir-primero-en-ia-responsable/>
- <https://datalegaldrive.com/ai-act/quest-ce-que-ai-act/#:~:text=L'AI%20Act%20ne%20se,sont%20utilisés%20dans%20l'UE.>
- <https://dpo101.fr/ia-act-reglementation-europeenne-intelligence-artificielle/>
- <https://digital-strategy.ec.europa.eu/fr/policies/regulatory-framework-ai>
- <https://www.accenture.com/fr-fr/insights/artificial-intelligence/synthetic-data-speed-security-scale>
- <https://artificialintelligenceact.eu/fr/ai-act-explorer/https://www.actuia.com/actualite/investissements-dans-lia-lue-doit-rapidement-passer-a-la-vitesse-superieure/>
- <https://blogdroiteuropeen.com/wp-content/uploads/2022/05/wp-ia-act-final-.pdf>
- <https://usbeketrica.com/fr/article/des-algorithmes-peuvent-ils-etre-ethiques>
- <https://www.amnesty.org/fr/latest/campaigns/2024/01/the-urgent-but-difficult-task-of-regulating-artificial-intelligence/>

Sources littéraires :

- Yann Le Cun – « Quand la machine apprend »
- Yannick Meneceur – « IA générative et professionnels du droit – Comprendre et s'appropriier la langue des probables »
- Ethan Mollick « Co-Intelligence: Living and Working with AI”

Prompts

- “Génère trois schémas traduisant le fonctionnement du réseau neuronal convolutif à partir de cette phrase : « (phrase expliquant le fonctionnement) »
- Utilisation de l’intelligence artificielle pour résumer certains articles juridiques
- Utilisation de l’IA aux fins de compréhension de certains principes mathématiques liés à l’IA afin de rendre l’explication écrite plus claire.

L'innovation en intelligence artificielle peut-elle se développer sur une politique éthique et de transparence algorithmique ?

