

Progressistes pour la social-démocratie

Intelligences Artificielles

L'essentiel

On utilise le terme d'Intelligence Artificielle (IA) pour désigner tout système pourvu de capacités habituellement associées à l'intelligence humaine et amplifiées par la technologie comme celles de simuler un raisonnement, de traiter de grandes quantités de données, d'interagir avec l'homme, d'apprendre en autonomie et d'améliorer ses performances de manière continue. Anthropomorphiser le phénomène de l'IA est cependant une source de confusions qui ouvre la porte aux craintes de la voir dépasser les humains et même de les exterminer comme aux rêves de la voir résoudre tous les problèmes de la planète. L'IA d'aujourd'hui, dite étroite, est avant tout un processus statistique adaptatif qui reste concentré sur des tâches spécialisées et bornées dans leur contexte d'exécution. Etant donné la diversité des cas d'utilisation, des technologies et des sciences qui nourrissent la discipline, on peut considérer qu'il n'y a pas une mais plusieurs Intelligences Artificielles.

L'IA est née avec l'informatique et se développe depuis lors. En 2023, l'IA est en passe d'intervenir dans toutes les activités de notre vie et aucun secteur économique n'est épargné : une ère de l' « IA partout » (AI everywhere) vient de s'ouvrir. Elle est en particulier marquée par la popularisation de l'IA générative par ChatGPT d'OpenAI, qui porte une mini-révolution tous azimuts des usages du numérique.

L'IA prend corps dans l'espace numérique sous forme de puissants algorithmes, notamment les algorithmes d'apprentissage automatique (Machine Learning), qui exploitent de grands volumes de données (BigData) et s'exécutent sur d'immenses fermes de calcul équipées de processeurs graphiques GPU et disponibles à la demande (Supercalculateurs combinés au Cloud).

Les IA ne sont pas vraiment intelligentes, ne sont pas vraiment artificielles, et ne sont pas non plus des intelligences augmentées. Il s'agit d'outils numériques qui contribuent, d'une certaine façon, à augmenter notre propre intelligence et dont les potentialités sont vertigineuses : optimisation des processus métiers, gestion des systèmes complexes, robotique collaborative et jumeaux numériques, recherche scientifique, personnalisation d'expérience, outils d'aide à la création artistique...

Les IA ouvrent la voie à une véritable collaboration entre les machines et les êtres humains, permettant à ces derniers de libérer tout leur potentiel de créativité et leur capacité d'innovation. Il faut toutefois prendre garde à certains risques que peuvent introduire ou amplifier les systèmes d'IA dans de multiples domaines : désinformation et manipulation des opinions par les hypertrucages, biais et discriminations, atteinte à la propriété intellectuelle et à la vie privée, impact environnemental, menaces sur l'emploi, effet de plateformisation et de concentration du pouvoir autour d'une poignée d'acteurs, repli malthusien des populations face au progrès ... Ces risques posent des défis sociétaux, économiques et réglementaires importants exacerbés par l'émergence récente des grands modèles génératifs et par l'ampleur et la rapidité sans précédent avec lesquelles les outils d'IA sont adoptés pour tout type d'usage.

Aux États-Unis, l'approche réglementaire est celle de la light touch regulation. Cette approche, encouragée par les GAFAM, a eu tendance à s'imposer jusqu'à présent. Elle pose un problème à l'Europe qui entend défendre ses valeurs en matière de sécurité et de droits fondamentaux et endiguer la perte de souveraineté qu'elle a subie dans le domaine du numérique. L'Union Européenne a donc adopté une démarche réglementaire plus musclée qui se traduit par de multiples réglementations progressives et coercitives sur le numérique et l'IA, avec pour point d'orgue le projet d'IA Act récemment adopté au

Parlement européen qui repose sur l'idée d'une IA de confiance et qui impose des règles à tous les systèmes d'IA avec des contraintes renforcées sur ceux jugés à « haut risque » (infrastructures critiques, éducation, ressources humaines, maintien de l'ordre). Cette stratégie fait débat car elle impose des contraintes potentiellement fortes (financières, procédurales) aux entreprises qui innovent dans le domaine et tentent de pénétrer un marché déjà largement dominé par les géants américains ou chinois.

L'IA comme source de puissance est un enjeu de leadership entre les États-Unis et la Chine ce qui contraint l'Europe à s'engager sur le même chemin même si elle le fait avec des moyens et des visées différents. La dimension européenne est nécessaire à la bonne exécution de la stratégie française en matière d'IA : l'Europe intervient de manière substantielle pour réguler les usages, fixer les règles pour la gestion des données, financer la recherche et favoriser l'émergence de champions industriels. Mais la prise de conscience des enjeux de l'IA a eu lieu avec un certain retard en Europe alors que les États-Unis et la Chine ont réalisé des investissements considérables dans la technologie dès 2015. Malgré la qualité de sa recherche et ses talents de niveau mondial, la France et l'Europe sont restées en retrait. L'Europe manque en particulier d'acteurs industriels majeurs du numérique et de l'électronique et les investissements dans les domaines pertinents restent insuffisants pour combler le retard.

L'imprégnation réussie de tous les pans de l'économie par l'IA représente à la fois une fenêtre d'opportunité pour la France et l'Europe, et un risque de déclassement en cas d'échec, à une échelle de temps que l'on peut estimer à ce jour entre 5 et 10 ans. En d'autres termes, on peut considérer qu'en 2035 au plus tard, les jeux seront faits en grande partie.

Nous, progressistes pour la social-démocratie, reconnaissons que la technologie d'IA est un sujet politique et pas seulement par ses implications éthiques ou sociales. Nous prenons position pour une IA sûre et digne de confiance, pour la promotion de l'innovation et le renforcement de la compétitivité européenne, pour le soutien à l'éducation, à la formation et à l'emploi dans un souci d'excellence, d'équité et d'inclusion, pour le respect des droits fondamentaux et de la vie privée, pour la sobriété d'usage et le respect de l'environnement et enfin, pour une régulation adaptative et une gouvernance d'implémentation efficace susceptibles de hisser la France et l'Europe au rang des leaders mondiaux du domaine sur les plans juridique et économique.

Groupe de travail IA

Intelligences Artificielles

Table des matières

Définition et concepts généraux	5
Définition	5
Dimension historique	5
Les systèmes experts et les premiers robots dotés d' <i>intelligence</i>	5
La période glaciaire de l'IA et l'avènement du Machine Learning.....	6
Internet et l'émergence des GAFAM	6
AI everywhere.....	6
Dynamique d'évolution	7
Les catégories d'IA.....	7
Les principales fonctions	8
Les technologies sous-jacentes	8
Collecte, acheminement, stockage et organisation des données	8
Calculs complexes et à grande vitesse sur les données.....	9
De puissants algorithmes	9
Intelligences Artificielles et intelligence humaine	11
Domaines d'application et potentialités.....	13
Optimisation des processus métiers	13
Gestion des systèmes complexes	14
Robotique collaborative et jumeaux numériques	15
Recherche scientifique	15
Création artistique.....	15
Personnalisation d'expérience.....	16
Education et inclusion	16
Les risques	17
Désinformation et manipulation des opinions.....	17
Biais et discriminations.....	18
Opacité des algorithmes et propriété intellectuelle.....	18
Protection de la vie privée.....	19
Impact environnemental	19
Emploi et précarité du travail	19
Sûreté et sécurité	20
Concentration du pouvoir et plateformisation.....	21
Perception culturelle	21

Les enjeux réglementaires.....	22
L'IA de confiance.....	22
La lutte contre l'exterritorialité.....	22
Les évolutions de la réglementation européenne.....	23
L'IA Act.....	24
Pour une régulation efficace.....	25
Vers une convergence des régulations ?	25
Les capacités de mise en œuvre en France et en Europe	27
La domination américaine et chinoise	27
Le soutien de l'Europe aux stratégies nationales	28
Les programmes budgétaires de soutien à la recherche.....	28
La construction d'un écosystème <i>AI-friendly</i>	29
Le contrôle des acquisitions étrangères dans l'UE.	29
La stratégie française.....	30
Les capacités de mise en œuvre.....	30
Les mesures à prendre, les orientations stratégiques.....	33
Pour une IA sûre et digne de confiance.....	33
Pour la promotion de l'innovation	34
Pour le soutien à l'éducation, à la formation et à l'emploi dans un souci d'excellence, d'équité et d'inclusion	35
Pour le respect des droits fondamentaux et de la vie privée.....	36
Pour la sobriété d'usage et le respect de l'environnement.....	36
Pour une régulation adaptative et une gouvernance d'implémentation efficace.....	37

Définition et concepts généraux

Définition

La paternité du terme **Intelligence Artificielle (IA)** est attribuée à John McCarthy du MIT (Massachusetts Institute of Technology) qui a organisé durant l'été 1956 au Dartmouth College à Hanover (New Hampshire, États-Unis) une conférence considérée comme l'acte de naissance de l'IA en tant que discipline de recherche autonome. Cette discipline, jeune d'une soixantaine d'années, est un ensemble de sciences, de théories et de techniques (logique mathématique, statistiques, probabilités, neurobiologie computationnelle, informatique) qui ambitionne d'imiter les capacités cognitives d'un être humain.

De nos jours, on utilise le terme d'Intelligence Artificielle pour désigner **tout système pourvu de capacités habituellement associées à l'intelligence humaine et amplifiées par la technologie** comme celles de simuler un raisonnement, de traiter de grandes quantités de données, d'interagir avec l'homme, d'apprendre en autonomie et d'améliorer ses performances de manière continue.

Ces capacités demeurent toutefois loin d'une intelligence humaine au sens strict, ce qui rend **la dénomination Intelligence Artificielle critiquable**. L'IA dite *forte*, une IA capable de contextualiser des problèmes spécialisés très différents de manière totalement autonome n'existe pas pour l'instant. L'IA d'aujourd'hui, dite *étroite*, ne raisonne pas et consiste en un processus statistique adaptatif qui, bien qu'extrêmement performant, reste concentré sur des tâches spécialisées et bornées dans leur contexte d'exécution.

L'IA prend corps dans l'espace numérique sous forme de puissants algorithmes, notamment les algorithmes d'apprentissage automatique (Machine Learning), qui exploitent de grands volumes de données (Big Data) et qui sont capables d'accomplir des tâches *à la manière* des humains.

Etant donné la diversité des cas d'utilisation, des technologies et des sciences qui nourrissent la discipline, on peut considérer qu'il **n'y a pas une mais plusieurs Intelligences Artificielles**. Dans la suite de cette note, nous utiliserons tout de même le terme Intelligence Artificielle ou IA pour désigner la discipline dans son ensemble et ses différents champs d'application.

Dimension historique

L'IA est née avec l'informatique et se développe depuis lors.

L'engouement récent du grand public pour l'IA est lié à l'essor des **plateformes conversationnelles** de nouvelle génération (dont la plus connue est **ChatGPT d'OpenAI**) qui se sont ouvertes à partir de novembre 2022 aux internautes du monde entier désireux de les essayer et d'éprouver leurs surprenantes capacités. L'adoption rapide et massive par le grand public et les entreprises de ces nouvelles solutions, en substitution aux moteurs de recherche classiques, pour la recherche ou la génération d'information, constitue une étape importante dans l'évolution de l'informatique et plus généralement dans la gestion du savoir humain.

Les systèmes experts et les premiers robots dotés d'*intelligence*

Il y a quelques décennies ce que l'on appelait IA était l'automatisation des raisonnements logiques par des **systèmes experts**, des programmes informatiques qui s'exécutent selon des règles fixes. Ces programmes ont donné naissance à plusieurs générations de robots, souvent peu convaincants en matière d'intelligence et peu susceptibles de motiver les investissements publics ou privés. Il s'agissait des robots frustes de la décennie 60 dotés d'une autonomie mécanique mais d'une très faible capacité logique, de ceux de la décennie 70, dotés de moyens de perception mais insuffisamment performants

et enfin de ceux de la décennie 80, un peu plus performants car ils ont commencé à apprendre c'est-à-dire à modifier de manière autonome un programme inscrit dans un **réseau de neurones**, une des fondations algorithmiques de l'IA moderne.

La période glaciaire de l'IA et l'avènement du Machine Learning

Dans les années 80, la sécurité des réseaux de neurones a été mise en question, de sorte que les systèmes experts ont été plus largement plébiscités. L'IA est alors entrée dans une période glaciaire, jusque vers 2010, période caractérisée par la **disponibilité de volumes de données colossaux au sein des majors d'Internet (GAFAM)**, et par l'accès facilité et à bon marché à la puissance de calcul extraordinaire des **calculateurs haute puissance (HPC) utilisant des processeurs graphiques d'ordinateurs (GPU)**. Cette période a été propice à l'avènement des **machines apprenantes (Machine Learning)** qui s'appuient sur des réseaux de neurones pour résoudre des problèmes inaccessibles aux systèmes experts.

En 2011, Watson (IBM) est devenu champion incontesté du jeu *Jeopardy*. En 2012, AlexNet a changé la donne en reconnaissance d'image, et en 2016 Alphago (Google) est devenu champion du monde de jeu de go. Alphago a marqué un immense saut qualitatif par rapport à Deep Blue, champion d'échec en 1997 et capable de calculer l'ensemble des combinaisons possibles du jeu d'échec à partir d'une position donnée : AlphaGo n'a pas eu besoin du soutien humain pour développer ses compétences, il s'est entraîné contre lui-même et a appris de son expérience. *A la manière* d'un humain, Alphago n'est pas capable de calculer l'ensemble des combinaisons possibles du jeu de go à partir d'une position de jeu donnée mais il est capable d'évaluer pour chaque position de jeu le mouvement à suivre le plus prometteur.

Internet et l'émergence des GAFAM

Plus que le micro-ordinateur, **Internet porte une 3ème révolution technologique** qui fait suite à celles de la vapeur et de l'électricité. Tandis que l'électricité a permis de répandre l'énergie et la force motrice, **Internet a permis de répandre l'information et la force logique**. Il existe toutefois une différence essentielle dans les catalyseurs de propagation de ces deux révolutions technologiques : il aura fallu des instances puissantes, des états en particulier, pour financer et structurer les capacités de création et de distribution de l'énergie électrique, tandis que l'expansion d'Internet et de ses applications s'est réalisée essentiellement à l'initiative d'acteurs privés grâce aux standards d'interopérabilité des systèmes informatiques en réseau. Et ces mêmes acteurs privés (les **GAFAM** essentiellement), soucieux de tirer pleinement bénéfice de l'économie numérique connectée et dont le leadership s'est renforcé par des effets d'agglomération (services, infrastructures, données), ont pris une place prépondérante dans les développements qui ont suivi, c'est-à-dire celui des objets connectés (smartphones et IoT), et celui de l'IA un peu plus tard.

Le poids économique et industriel de l'informatique s'est donc largement accru au début des années 2000 et s'est largement **concentré sur un panorama restreint d'acteurs privés**, américains en premier chef, **aux dépens des forces étatiques**. Nous reviendrons plus tard sur les capacités de la France et de l'Europe dans le domaine de l'IA en comparaison aux deux superpuissances que sont les Etats-Unis et la Chine.

AI everywhere

1995 aura vu l'avènement du Web reposant sur l'infrastructure de réseaux Internet, 2007 aura vu l'apparition de l'Internet mobile et du smartphone. En 2023, l'IA est en passe d'intervenir dans toutes les activités de notre vie et aucun secteur économique n'est épargné : une ère de l'« IA partout » (**AI everywhere**) vient de s'ouvrir. Elle est en particulier marquée par la popularisation de l'**IA générative**

par **ChatGPT d'Open AI**, qui porte une mini-révolution tous azimuts des usages du numérique. Bill Gates, le fondateur de Microsoft, écrivait sur le réseau social Twitter les 31 mars 2023 : « Le développement de l'IA est aussi fondamental que la création du microprocesseur, de l'ordinateur personnel, de l'internet et du téléphone portable. Il changera la façon dont les gens travaillent, apprennent, voyagent, se soignent et communiquent entre eux. ».

Dynamique d'évolution

L'IA se développe à une vitesse fulgurante. En 2023, une frénésie d'investissement, en particulier dans la Silicon Valley américaine, soutient une myriade de startups nouvellement créées dans le domaine. Les capacités des infrastructures Big Data et des Supercalculateurs continuent de croître de manière exponentielle, les algorithmes se diversifient, sont optimisés et de nombreux plugins sont proposés pour permettre l'utilisation de l'IA dans des systèmes tiers. En outre, de nombreuses solutions d'IA sont proposées en distribution libre (open source) afin de rendre leur développement plus transparent, plus qualitatif et susceptible de passer à grande échelle.

L'IA opère une colonisation de tout l'existant numérique. Citons pour illustration :

- Le standard **ONNX (Open Neural Network Exchange)** initié en 2017 par Facebook, sous licence libre Apache, qui permet de développer des applications de Machine Learning destinées aux smartphones,
- La librairie **FKB (Fortran-Keras Bridge)** qui permet de découpler l'efficacité de recherches en modélisation et simulation en transférant des réseaux de neurones dans les codes de simulation¹, et inversement.

Cette dynamique entropique de l'IA rend les efforts d'appréhension politique du phénomène et la construction de cadres réglementaires particulièrement difficiles, d'autant que peu de décideurs politiques ou institutionnels sont aujourd'hui en réelle capacité de porter un regard clairvoyant sur les enjeux à venir.

Les catégories d'IA

Les chercheurs ont classifié les intelligences Artificielles en trois catégories selon leur capacité à résoudre des problèmes et à effectuer des tâches *à la manière* des humains :

- L'**IA étroite** (ANI, Artificial Narrow Intelligence) également appelée IA faible, est une IA entraînée et concentrée pour effectuer des tâches spécifiques. L'intelligence artificielle étroite régit l'essentiel de l'IA actuelle et contribue au fonctionnement de certaines applications très robustes : joueurs virtuels d'échec, assistants vocaux, solutions de reconnaissance d'images.
- L'**IA générale** (AGI, Artificial General Intelligence) est une forme théorique d'**IA forte** qui aurait une intelligence égale à celle des humains : elle aurait une conscience d'elle-même, serait capable de résoudre des problèmes, d'apprendre et de planifier l'avenir dans un nombre illimité de domaines ou de situations.
- La **super intelligence artificielle** (ASI, Artificial Super Intelligence) est une IA forte qui dépasserait l'intelligence et les capacités du cerveau humain.

Seule l'intelligence artificielle étroite est, à date, une réalité.

¹ <https://doi.org/10.1155/2020/8888811> présente une application de la librairie Fortran-Keras Bridge à la simulation météo

Les principales fonctions

Les principales fonctions des Intelligences Artificielles sont la prédiction (IA prédictive), la classification (IA discriminative) et la génération de données (IA générative) et **les systèmes d'IA actuels utilisent communément une combinaison de ces trois fonctions.**

L'**IA prédictive** est une intelligence artificielle qui recueille et analyse des données afin de prédire des événements futurs. Elle vise à comprendre les modèles et les relations entre les données historiques et à faire des prédictions éclairées sur de nouveaux jeux de données. Citons parmi les cas d'usage de l'IA prédictive la détection des fraudes financières ou les prévisions météorologiques.

L'**IA discriminative** est une intelligence artificielle qui se concentre sur la classification des données en fonction d'étiquettes spécifiques. L'étiquetage des données consiste à ajouter une ou plusieurs étiquettes (labels) à des données brutes (images, fichiers texte, vidéos) dans l'objectif d'aider les algorithmes d'apprentissage automatique à comprendre et à classer les informations qu'ils traitent. La **vision par ordinateur** ou encore **le traitement automatique du langage naturel (NLP)** sont des cas d'usage de l'IA discriminative.

L'**IA générative** est une intelligence artificielle qui se concentre sur l'estimation d'une distribution de probabilité sur un ensemble de données. Elle utilise des modèles génératifs pour générer de nouvelles instances de données, en utilisant des points de données préexistants et des instructions (prompts) comme point de départ. Ces modèles utilisent un ensemble de règles pour générer de nouvelles données. L'IA générative est depuis moins d'une année largement connue du grand public grâce à la diffusion des solutions de **LLM (Large Language Models)**, des IA entraînées à partir d'une multitude d'articles, d'entrées Wikipédia, de livres, de ressources Internet et d'autres données qui ont la faculté de produire des réponses (textes, images, vidéos, sons) semblables à celles des humains à des requêtes en langage naturel (**ChatGPT** d'OpenAI, **Google Bard**, **Llama 2** de Meta, **MidJourney** laboratoire de recherche indépendant, **Mistral AI** la licorne française dans ce domaine, etc.).

Les technologies sous-jacentes

La progression fulgurante des Intelligences Artificielles est rendue possible par une combinaison de trois facteurs :

- La capacité de collecter, d'acheminer, de stocker et d'organiser de vastes quantités de données
- La capacité d'effectuer des calculs complexes sur les données à des vitesses élevées
- La disponibilité de puissants algorithmes basés sur le Machine Learning

Collecte, acheminement, stockage et organisation des données

Afin d'exceller dans les tâches qui leur sont confiées, les Intelligences Artificielles ont besoin de se nourrir de grandes quantités de données (Volume), de les traiter à grande vitesse (Vitesse), de gérer leurs formes diversifiées et leur hétérogénéité (Variété), de compter sur leur fiabilité et leur qualité (Véracité) et sur leur valeur intrinsèque (Valeur). Les 5 V précédents constituent les piliers des systèmes **Big Data** qui ont pour mission de collecter, d'acheminer, de stocker et d'organiser les données à des fins d'analyse ou d'exploitation par d'autres systèmes ou programmes, comme les algorithmes d'apprentissage des Intelligences Artificielles.

Avec l'**Internet mobile**, on peut considérer qu'une fraction importante de l'humanité accède déjà aux services d'Internet via leurs smartphones. Cet usage génère un volume colossal de données stockées dans les datacenters et exploitables par l'IA.

L'Internet Of Things (IoT) ou en français Internet des Objets (**IdO**) qui représente l'écosystème des objets et équipements connectés au réseau internet constituera progressivement un formidable gisement de données pour les applications d'Intelligences Artificielles de demain. D'ici 2025, 42 milliards d'appareils seront connectés au réseau Internet et les cas d'usage de l'Intelligence Artificielle appliquée aux données de l'IoT (AIoT : Artificial Intelligence of Things) vont se développer de manière exponentielle : analyse de données et prise de décision en temps réel, prédiction et prévention de pannes, amélioration des processus, personnalisation d'expériences.

Les infrastructures Big Data et celles supports à l'Internet mobile et à l'IoT (nouvelles générations de réseaux mobiles (5G), plateformes de Edge Computing) constituent donc des actifs clés de la chaîne de valeur des solutions d'Intelligences Artificielles.

Calculs complexes et à grande vitesse sur les données

Le calcul est le cœur de la mise en œuvre des algorithmes d'IA. Etant donné la quantité de données à traiter et la complexité des algorithmes, il convient de disposer de puissants microprocesseurs capables de réaliser des tâches complexes en un temps limité. Les **Supercalculateurs** combinés au **Cloud** constituent les infrastructures de calcul nécessaires à l'exécution des applications d'IA.

Également appelé **HPC**, pour **High Performance Computing**, un Supercalculateur est constitué de centaines voire de milliers de serveurs de calcul mis en réseau. Chaque serveur représente un nœud et les nœuds fonctionnent en parallèle, ce qui accélère la vitesse de traitement et fournit un calcul dit de haute performance. Les supercalculateurs les plus rapides de l'industrie sont équipés de **GPU (Graphics Processing Unit)**, des processeurs graphiques qui excellent dans le calcul parallèle et matriciel à haute performance.

Le modèle d'hébergement en **Cloud** (accès via internet aux capacités de calcul, stockage centralisé et traitement Big Data, paiement à l'usage) est un modèle particulièrement adapté à la fourniture de capacités de calcul haute performance aux particuliers, aux entreprises et aux administrations de toute taille

L'IA en Cloud est donc un marché prometteur qui devrait permettre à des acteurs souverains en France et en Europe, de proposer des offres de plateformes spécialisées susceptibles de concurrencer, sur le segment de l'IA, celles des hyperscalers américains ou asiatiques.

A un horizon de temps plus lointain, nous pouvons conjecturer que **la technologie de calcul quantique**, grâce à sa puissance combinatoire et la promesse d'un parallélisme massif, permettra d'augmenter radicalement la puissance de calcul nécessaire aux modèles d'apprentissage automatique et propulsera l'IA dans une nouvelle ère. Au moment où nous écrivons cette note, la création d'un ordinateur quantique pleinement fonctionnel reste toutefois un défi de taille.

De puissants algorithmes

Les Intelligences Artificielles s'appuient sur de puissants algorithmes.

Le **Machine Learning**, ou **l'apprentissage automatique**, est le premier sous-domaine algorithmique de l'Intelligence Artificielle. Il s'agit d'un programme ayant la capacité d'apprendre et de résoudre des problèmes par lui-même. Le procédé se déroule en deux phases : celle de l'apprentissage à proprement parler, au cours de laquelle le programme établit un modèle à partir d'un grand volume de données de référence, puis celle de la décision, où le programme se base sur son apprentissage pour définir automatiquement la réponse à apporter à un événement donné.

Le **Deep Learning**, ou apprentissage profond, est un sous-domaine du Machine Learning. L'apprentissage profond est un apprentissage automatique s'appuyant sur un réseau de neurones qui permet d'affiner l'analyse de manière itérative grâce à un traitement par couches.

Un **réseau de neurones** désigne l'architecture du Deep Learning qui s'inspire des synapses et des neurones du cerveau humain. En pratique, il s'agit d'une technique de calcul statistique (régression logistique) basée sur l'association d'un grand nombre de cœurs processeurs travaillant en parallèle, chacun sur une partie extrêmement simple d'un ensemble complexe, et disposés en couches successives. Le réseau de neurones est adaptatif. Il se modifie par lui-même au fur et à mesure qu'il apprend pour améliorer en continu les résultats fournis.

Le Machine Learning se décline sous différents types de modèles, qui emploient chacun des techniques algorithmiques différentes. Selon la nature des données et le résultat souhaité, l'un de ces quatre **modèles d'apprentissage** peut être utilisé : **supervisé**, **non supervisé**, **semi-supervisé** ou **par renforcement**. Dans chacun de ces modèles, une ou plusieurs techniques algorithmiques peuvent être appliquées. Tout dépend des ensembles de données qui seront utilisés et de l'objectif visé au niveau des résultats. Par nature, les algorithmes de Machine Learning sont conçus pour classer des éléments, repérer des motifs ou structures (patterns), prévoir des résultats et prendre des décisions fondées sur leur apprentissage. Les algorithmes peuvent être mis en œuvre individuellement ou en groupe dans le but d'atteindre la plus grande précision possible lorsque les données utilisées sont complexes et imprévisibles.

Le panorama des technologies nécessaires à l'IA est vaste et donne une idée de la diversité des filières stratégiques d'investissements de la France et de la Europe en matière d'IA s'il s'agit pour ces dernières de tirer pleinement bénéfice de la révolution en cours et de se faire une place de premier plan dans la compétition mondiale.

Intelligences Artificielles et intelligence humaine

Le scientifique britannique Alan Turing s'est demandé en 1950 si les machines pouvaient penser. Il imagina qu'on puisse en créer une capable de dialoguer avec un humain sans que celui ou celle-ci soit capable de détecter la supercherie. Quand le terme Intelligence Artificielle est apparu en 1956, Turing avait déjà *pipé* le sujet en fixant comme critère la ressemblance avec notre intelligence. Or, l'IA étroite, la seule forme incarnée de l'IA, n'imité ni ne reproduit l'intelligence humaine, elle se limite à simuler le comportement humain sur la base d'une gamme étroite de paramètres et de contextes.

Anthropomorphiser le phénomène de l'IA est donc une source de confusions qui ouvre la porte aux craintes de la voir dépasser les humains et même de les exterminer comme aux rêves de la voir résoudre tous les problèmes de la planète. Cela altère les discussions sérieuses et la sensibilisation aux potentiels et aux risques posés par les technologies concernées. Les IA d'aujourd'hui ne sont pas vivantes et ne sont pas sur le point de le devenir. Elles sont avant tout des outils et les concepts d'IA générale comme de super intelligence artificielle doivent donc être abordés avec circonspection, prudence et même scepticisme.

Le fonctionnement de ces outils doit pouvoir être compris au moins pour cette raison. Leurs créateurs doivent rendre ces éléments communicables.

On peut évoquer au moins les trois distinctions suivantes entre intelligence humaine et IA :

- Une **distinction du point de vue du fonctionnement** : l'IA met en œuvre des processus statistiques, donc beaucoup plus simples que le fonctionnement du cerveau humain ;
- Une **distinction sur les finalités** : les projets d'IA sont conçus pour résoudre des problèmes ou relever des défis, tandis que la finalité ou la « mission » de l'intelligence humaine relève du mystère ;
- Une **distinction de nature** : les projets d'IA répondent à des spécifications ou des schémas établis, l'intelligence humaine est connue de manière expérientielle.

Ces distinctions amènent à **suspecter autant les présentations évoquant une convergence entre l'IA et l'intelligence humaine**, que ce soit pour déclencher une sympathie ou une crainte, **que les discours autour de la fameuse notion de singularité**, le fameux jour où l'IA échappera à l'humanité.

Les IA ne sont pas vraiment intelligentes. Le fait qu'elles peuvent l'emporter sur les humains dans certains affrontements spécifiques (échecs ou go) est la preuve qu'elles sont parfois de très bons outils, mais leurs capacités d'intelligence se réduisent à du *pattern matching*, la capacité de retrouver les mêmes motifs et donc de les reproduire en étudiant des myriades de données.

Les IA ne sont pas vraiment artificielles. Les informations récupérées et utilisées par des chatbots comme celui de Google ou de OpenAI-Microsoft sont le fruit du travail d'humains, récompensés ou rémunérés pour cela de manière extrêmement inégale. Les micro-tâches indispensables à l'IA (entraîner l'IA, réparer l'IA, vérifier l'IA, imiter l'IA) sont payées en centimes l'unité, voire rendues sans contrepartie financière². Elles consistent à traiter de grandes masses de données en les fragmentant et en les atomisant en petits projets et cela semble impliquer des millions de personnes.

Les IA ne sont pas non plus des intelligences augmentées car elles gardent la référence humaine. Il s'agit d'outils numériques qui, d'une certaine façon, contribuent à *augmenter* notre propre intelligence.

² C'est le cas par exemple des systèmes de captcha qui permettent un étiquetage gratuit par les utilisateurs du web mondial de grandes masses de données au profit des acteurs du numérique

Toutes ces considérations amènent à porter attention au positionnement de l'IA par rapport à l'humanité et aux humains : même si elle devient un assistant de plus en plus utile, voire indispensable pour de nombreuses d'activités, **les humains ne peuvent déléguer complètement la prise de décision à l'IA car cela reviendrait à abandonner leur responsabilité : l'IA doit toujours proposer à l'humain, et le seconder.** Pour illustration, l'IA est certainement un outil très efficace de préparation des instructions juridiques, mais ne doit en aucun cas se substituer aux décisions de justice qui sont de la responsabilité des juges dans l'exercice de leur intime conviction. De la même manière, l'interprétation des images médicales par IA doit toujours rester sous le contrôle de personnels médicaux compétents.

L'utilisation de l'IA dans les systèmes autonomes, c'est-à-dire des systèmes capables de fonctionner sans intervention humaine, **pose des problèmes spécifiques.** Certes, la robotique nous a habitués à des systèmes autonomes pour des applications *simples* qui ne doivent en aucun cas être confondus avec des systèmes apprenants : le pilotage d'un métro, par exemple, obéit à un code fixe, sans aucune marge de manœuvre et la sûreté repose sur un processus formel très exigeant. Mais de nombreux autres scénarii d'utilisation émergent et posent question : c'est le cas des véhicules autonomes pour lesquels, malgré quelques implémentations locales plus ou moins réussies³, il est peu probable que l'on atteigne le niveau 5 de la conduite sans intervention humaine selon la classification établie par la Society of Automotive Engineers (SAE) pour des questions *éthiques* et de *responsabilités* en cas d'accident dans les environnements complexes comme les villes. C'est le cas également des systèmes militaires létaux autonomes, dont on ne peut que regretter l'existence, pour lesquels un cadre réglementaire international reste à définir.

³ En août 2023, les autorités californiennes ont autorisé les firmes Waymo (filiale d'Alphabet) et Cruise (propriété de General Motors) à faire circuler leurs taxis autonomes 24h sur 24 et 7 jours sur 7 dans l'ensemble des rues de San Francisco. Impliqués dans plusieurs accidents et présentant des comportements parfois erratiques, les taxis autonomes ne font pas encore l'unanimité au sein de la cité californienne.

Domaines d'application et potentialités

Les potentialités de l'IA sont vertigineuses. Les administrations et les entreprises, quels que soient leur taille, leur géographie ou secteur d'activité (industrie, services, agriculture, santé, éducation, armée ...) entendent bénéficier de l'IA pour réussir pleinement dans leur mission, se moderniser, saisir de nouvelles opportunités, gagner en compétitivité et en efficacité. En outre, l'IA est le catalyseur d'une transformation profonde des méthodes de travail et des interactions des travailleurs avec leur environnement. Elle ouvre la voie à une véritable collaboration entre les machines et les êtres humains, permettant à ces derniers de libérer tout leur potentiel de créativité et leur capacité d'innovation.

Afin d'entrevoir les potentialités et les grands domaines d'applications de l'IA, tentons de faire un tour d'horizon des cas d'utilisation d'ores et déjà en vigueur ou pressentis de cette technologie de rupture.

Optimisation des processus métiers

L'Intelligence Artificielle est amenée à prendre une part croissante dans l'optimisation des processus métiers :

Processus de production	Meilleure planification, utilisation plus efficace des ressources, prise de décision plus rapide, optimisation des flux de production.
Automatisation des tâches répétitives	Automatisation des tâches de collecte et de saisie de données, automatisation des tâches de génération d'informations répétitives (synthèses documentaires, comptes-rendus, courriels), planification de rendez-vous.
Maintenance prédictive	Détection d'anomalies et de signes précurseurs de pannes sur les systèmes.
Logistique et chaîne d'approvisionnement	Amélioration du contrôle de l'information dans la chaîne d'approvisionnement, gestion des stocks en temps réel, émission de commandes d'approvisionnement instantanées, prédiction des tendances de consommation, automatisation des opérations de stockage, génération d'itinéraires de transport optimisés.
Idéation, conception de produits et ingénierie	Génération d'idées et d'options de conception, formalisation de problèmes, propositions de solutions (architecture, code informatique), chatbots spécialisés dans le développement de systèmes.
Qualité des produits	Détection des défauts de qualité des produits et identification des causes sous-jacentes.
Sécurité et conformité	Détection de comportements dangereux ou non conformes, détection de fraudes ou d'activités malveillantes.
Expérience client	Chatbots, voicebots, assistants virtuels, interactions temps-réels et recommandations.

Marketing produit	Génération de fiches produit et de contenus promotionnels
Communication et publicité	Création de visuels publicitaires, assistance à la rédaction, génération de textes pour le SEO (Search Engine Optimization).
Service client	Catégorisation, synthèse et routage des messages, analyse des humeurs et des sentiments, retranscription des conversations, proposition de réponses personnalisées et contextualisées, synthèse vocale.
Traitement de l'information et génération de contenu	Recherche bibliographique, résumé de documents, analyse de contenus complexes, reformulation, traduction, génération de contrats, de règlements ou de documents juridiques.

Figure 1 - Les cas d'utilisation de l'IA dans les processus métiers de l'entreprise

De nombreux cas d'utilisation ne seront pas développés par l'entreprise elle-même : les éditeurs des suites bureautiques et collaboratives, du développement logiciel ou encore du marketing digital embarquent progressivement des fonctionnalités innovantes basées sur des modèles d'IA générative. En novembre 2023, Microsoft a par exemple annoncé l'intégration de son outil Copilot, un assistant personnel IA, à toutes les applications de sa suite bureautique et à sa plateforme de développement de logiciels GitHub.

Afin d'exploiter toute la valeur de l'IA et d'en tirer des avantages concurrentiels, les entreprises devront toutefois dépasser les cas d'utilisation génériques et créer des modèles et applications sur-mesure à partir de leur propre patrimoine informationnel.

Gestion des systèmes complexes

On qualifie de **complexe** un système composé d'une multitude d'entités dont les interactions locales font émerger des propriétés globales difficilement prédictibles par la seule connaissance des propriétés de ces entités. Une nuée d'oiseaux, un réseau social, le climat, une ville, Internet sont des exemples de systèmes complexes.

L'un des domaines les plus prometteurs de la recherche en IA est le développement d'algorithmes capables de modéliser et de prédire le comportement de tels systèmes.

L'IA est par exemple utilisée pour modéliser la propagation de maladies infectieuses, telles que le COVID-19. Elle exploite pour cela les données sur les mouvements de population, les interactions sociales et d'autres facteurs. Elle est utilisée pour étudier le comportement des marchés financiers qui se caractérisent par un grand nombre de variables en interaction (cours des actions, taux d'intérêt, indicateurs économiques). Elle est encore utilisée pour optimiser le fonctionnement des réseaux électriques ou de télécommunications à partir de l'exploitation des données de consommation et de production des systèmes.

L'IA est amenée à jouer un rôle de premier plan dans l'étude et la gestion des systèmes complexes, et offre des perspectives nouvelles pour faire face aux défis les plus urgents auxquels notre monde est actuellement confronté comme ceux du réchauffement climatique⁴, des pandémies et des inégalités.

⁴ <https://www.inria.fr/en/ai-climate-change-environment>

Robotique collaborative et jumeaux numériques

La robotique collaborative est un cas d'utilisation de l'IA bénéfique à tous les secteurs où la répétitivité, la précision et la pénibilité des tâches est en jeu.

Les cobots sont des machines alimentées par l'Intelligence Artificielle créées spécifiquement pour aider les humains. Ils sont conçus pour travailler en toute sécurité aux côtés des humains dans un espace de travail partagé et ont vocation à réaliser des tâches précises, répétitives et pénibles pour l'homme : prélever et placer, contrôler la qualité ou lever de lourdes charges par exemple. À la différence des robots classiques, les cobots sont des outils bon marché, polyvalents, capables de réaliser de nouvelles tâches et de les reproduire autant de fois que nécessaire. Ils constituent une alternative au *tout automatique* et sont utilisés lorsque les robots seuls ne peuvent pas remplacer l'homme dans son expertise, son savoir-faire, sa sensibilité, sa perception des choses et de son environnement.

Autres cas d'utilisation de l'IA, **les jumeaux numériques**, des modèles numériques qui reconstituent fidèlement un objet ou un système physique, **sont de plus en plus utilisés dans l'industrie pour optimiser le fonctionnement et la maintenance d'actifs physiques, de systèmes ou de processus de production.** Le modèle numérique se nourrit de manière continue des données d'observabilité du système et les algorithmes d'IA simulent les résultats des actions susceptibles de l'améliorer, de l'optimiser ou encore d'en assurer la maintenance.

Recherche scientifique

L'IA est le nouveau moteur de la recherche scientifique. Les applications d'IA envahissent à grande vitesse les laboratoires pour améliorer les instruments, accélérer les calculs, aiguiller vers des hypothèses fécondes, et optimiser de manière radicale le traitement et le partage de l'information. Grâce à l'apprentissage automatique, les scientifiques collectent, relient et analysent des données à des échelles sans précédent, ils sont capables d'élaborer des modèles de prédiction approfondis englobant de multiples domaines d'étude, ou encore de simuler des phénomènes physiques complexes. Cette dynamique d'adoption peut s'illustrer par quelques exemples.

Dans le domaine de la physique des particules, l'IA est utilisée pour analyser des résultats produits par les collisionneurs. Ces infrastructures de recherche, représentée par le Grand Collisionneur de Hadrons, produisent des données massives. Les algorithmes d'apprentissage automatique permettent de détecter des particules, de classer les événements et d'identifier des signaux faibles dans le bruit de fond.

Dans le domaine de la recherche médicale, une collaboration entre l'Institut Curie et Ibex Médical Analytics a permis début 2023 de démontrer la performance, la fiabilité et la mise en application clinique d'un algorithme d'IA capable de diagnostiquer les cancers du sein lors de biopsies mammaires. 18 spécialistes ont entraîné un algorithme d'IA sur plus de deux millions d'échantillons d'images afin de garantir un diagnostic robuste et fiable.

Création artistique

L'IA générative offre aux créatifs de nouvelles façons de développer leur art. Les générateurs d'images (Dall-E, MidJourney) donnent vie à des concepts visuels inédits tandis que les générateurs de musique (AIVA), capables de composer des mélodies originales, offrent aux musiciens compositeurs de nouvelles sources d'inspiration. En 2018, le portrait fantomatique Edmond de Belamy, créé par le collectif français Obvious, a été la première œuvre réalisée à l'aide d'une intelligence artificielle vendue aux enchères par la grande maison de vente aux enchères Christie's.

L'art est et restera le reflet de l'expérience humaine, de nos émotions et de notre vécu. L'IA vient enrichir la *palette* de l'artiste et la collaboration entre les humains et les IA doit être perçue comme une ouverture vers de nouvelles voies créatives.

Personnalisation d'expérience

Personnalisation du fil d'actualité d'un réseau social en fonction de nos habitudes d'interactions (Instagram, Facebook, LinkedIn, Twitter), recommandations d'œuvres musicales ou de vidéos en fonction de nos goûts (Spotify, Deezer, Youtube), génération de courriels en fonction de nos humeurs ou de nos intentions (Google Bard pour Gmail), l'IA se greffe d'ores et déjà à nos usages numériques quotidiens en proposant une expérience unique et personnalisée qui fidélise et crée une relation émotionnelle durable entre le client et le produit ou service qu'il consomme.

Les outils de personnalisation basés sur l'IA offrent aux entreprises un moyen puissant et efficace de créer une expérience client positive et de se démarquer ainsi de la concurrence.

Education et inclusion

L'IA trouve sa place au sein des systèmes éducatifs par une gamme de cas d'utilisation utiles à l'enseignant/formateur, aux élèves/stagiaires ou aux institutions elles-mêmes.

Au service de l'enseignant/formateur, l'IA pourra s'avérer une aide précieuse pour la détection de plagiat, la curation intelligente du matériel d'apprentissage, l'évaluation sommative automatique ou la création de ressources pédagogiques.

Au service de l'élève/stagiaire, elle permettra par exemple la mise en place de systèmes de tutorat intelligents basés sur le dialogue et d'environnements d'apprentissage exploratoire.

Et enfin au service des institutions, elle peut être utilisée pour faciliter la gestion des admissions ou encore la planification (sites, cours, horaires).

En outre, alors que 12 millions de Français sont en situation de handicap, et que 25 % d'entre eux présentent un niveau d'étude équivalent ou supérieur au bac, **l'IA responsable pourra être un vecteur d'inclusion dans l'enseignement, en tant que moyen de facilitation de l'apprentissage pour les personnes porteuses d'un handicap.** Pour ces personnes, les problèmes concernant la formation se posent dès le départ avec deux problématiques : l'accès à la formation d'une part, et l'adaptation de la formation à leur contexte particulier. Les outils numériques enrichis d'Intelligence Artificielle (logiciels de transcription instantanée de la parole, logiciels de lecture vocale, logiciels de reconnaissance de caractères, outils d'aide à la rédaction, plateformes d'enseignement à distance inclusives) peuvent apporter une aide significative aux dispositifs de formation et d'inclusion de tous les publics.

Les risques

L'introduction de systèmes basés sur l'IA dans de nombreux domaines présente des risques certains et pose des défis sociétaux, économiques et réglementaires importants exacerbés par l'émergence récente des grands modèles génératifs, en particulier les LLM (Large Language Model) et par l'ampleur et la rapidité sans précédent avec lesquelles les outils d'IA sont adoptés pour tout type d'usage.

Désinformation et manipulation des opinions

Le contenu généré par IA (images, textes, audios, vidéos...) peut être saisissant de réalisme. La création d'un contenu *deep fake*, ou *hypertrucage* en français, reposant sur la technologie d'IA est dorénavant à la portée de tous. Il peut être produit à bas coût, diffusé à grande échelle via Internet et les réseaux sociaux, et contribuer efficacement à des stratégies malveillantes de désinformation et de manipulation des opinions.

La *théorie du complot* ne date pas d'hier et la désinformation n'a attendu ni Internet, ni les réseaux sociaux pour se propager à très grande échelle. Les *protocoles des Sages de Sion*⁵ en témoignent. Néanmoins, les technologies modernes de création et de diffusion massive d'informations boostés à l'IA posent un risque accru impliquant une nécessaire régulation et des moyens pour atténuer le risque comme les mécanismes de marquage (tatouage) des productions créées artificiellement ou l'autorégulation des systèmes par les agrégateurs ou producteurs de contenu. Pour ces derniers, la question des critères d'appréciation de filtrage de l'information reste toutefois une gageure.

Des dispositifs institutionnels de lutte contre la désinformation existent en France (Viginum, Pharos), et de nombreuses structures françaises ou européenne œuvrent pour recenser, comprendre et déconstruire les contenus manipulateurs : EDMO⁶, CEDMO⁷, Disinfo⁸, Invid⁹, les organismes de presse¹⁰, des ONG associant des journalistes¹¹. Notons par ailleurs que l'IA, désormais au cœur de la désinformation sur les réseaux sociaux, peut aussi contribuer à la juguler. Certains outils de vérification de véracité d'informations reposent en effet sur des modèles d'IA. Citons par exemple le projet VERA¹² ou le Global Disinformation Index¹³.

Malgré ces initiatives, **le défi de résilience de notre société et de notre économie face aux désordres de l'information sur les réseaux sociaux**, qu'ils soient d'origine non intentionnelle ou le fait d'organisations ou de puissances étatiques en confrontation avec nos démocraties et nos valeurs, **reste entier. La capacité de la France et de l'Europe à relever ce défi renvoie à celui d'une autonomie technologique.** La transparence apportée par le DSA¹⁴ ne sera réellement utile que si des experts

⁵ Les Protocoles des Sages de Sion est un texte inventé de toutes pièces par la police secrète du tsar et publié pour la première fois en Russie en 1903. Ce faux se présente comme un plan de conquête du monde établi par les Juifs et les francs-maçons. Traduit en plusieurs langues et diffusé à l'échelle internationale dès sa parution, il devient un best-seller.

⁶ <https://edmo.eu>

⁷ <https://cedmohub.eu>

⁸ <https://euvsdisinfo.eu>

⁹ <https://www.invid-project.eu/>

¹⁰ <https://factuel.afp.com>, <https://www.lemonde.fr/les-decodeurs/>, https://www.bbc.com/news/reality_check

¹¹ <https://maldita.es/>

¹² <https://www.veraai.eu/home>

¹³ <https://www.disinformationindex.org/>

¹⁴ Le règlement européen DSA (Digital Services Act) fixe un ensemble de règles pour responsabiliser les plateformes numériques et lutter contre la diffusion de contenus illicites ou préjudiciables ou de produits illégaux

indépendants peuvent au minimum caractériser les effets et les risques des réseaux sociaux et des sites de recherche.

La meilleure arme contre le risque de désinformation reste avant tout l'esprit critique. Les dispositifs techniques comme **les tatouages numériques ne pourront se substituer à une formation des utilisateurs et une adaptation des contextes d'usage des logiciels.** Dans un monde où la frontière entre le vrai et le faux pourra être rendue plus floue, **l'importance de savoir identifier les sources d'information fiables devient une exigence incontournable.** Cette éducation à l'esprit critique, sans être en soi suffisante, est probablement nécessaire pour accompagner l'usage des IA génératives.

Biais et discriminations

L'IA est une technologie perméable aux biais cognitifs qui peut entraîner des discriminations injustes, en particulier dans les domaines du recrutement, du crédit, de la justice, de la santé et des prestations sociales où l'IA peut amplifier et renforcer les préjugés existants.

Il existe des biais plus récurrents que d'autres dans le domaine de l'IA, notamment :

- **Les biais historiques** qui se contentent de reproduire les inégalités du monde réel,
- **Les biais liés à la sélection ou à l'agrégation des données d'entraînement** qui ne reflètent pas toujours la complexité et la variabilité du monde réel,
- **Les biais algorithmiques** qui reflètent les biais des personnes impliquées dans le développement de l'algorithme et du système de décision,
- **Les biais de confirmation** qui nous poussent à croire une information lorsqu'elle corrobore nos croyances existantes.

Il convient donc, pour les administrations ou les entreprises responsables du développement de systèmes d'IA, de mettre en place des garde-fous sur l'ensemble de la chaîne de production afin :

- **D'évaluer les risques** liés aux variables d'entrées des algorithmes,
- **De maîtriser les risques de biais** dans les données d'apprentissage,
- **Et de superviser la production de résultats** des algorithmes, c'est-à-dire la justesse des prédictions effectuées.

Ces garde-fous impliquent des devoirs (via la régulation) mais aussi des investissements importants dans les moyens de contrôle, un équilibre difficile à trouver pour les jeunes pousses et qui ne doit pas brider l'innovation.

A noter qu'il existe aujourd'hui des outils créés pour lutter contre les biais de l'IA. On peut citer le What-if Tool de Google et l'AI Fairness 360 Open Source Toolkit d'IBM, qui permettent d'examiner et d'inspecter les modèles de Machine Learning pour détecter les biais.

Opacité des algorithmes et propriété intellectuelle

Les modèles de Machine Learning sont souvent considérés comme des boîtes noires. Les réseaux neuronaux utilisés dans l'apprentissage en profondeur sont parmi les plus difficiles à comprendre pour un être humain : une fois la phase d'apprentissage achevée, le raisonnement est distribué sur une multitude de neurones artificiels, chacun en charge d'une infime partie du traitement. Entre les données d'entrée et le résultat, le système est opaque même pour ses concepteurs. Comprendre ces *boîtes noires*, ou du moins rendre leur fonctionnement plus compréhensible, devient un enjeu majeur, en particulier si on vise à les utiliser pour des applications jugées critiques (santé, défense, ...). **C'est le défi que se fixe l'IA explicable (eXplainable AI ou XAI),** un ensemble de techniques et de méthodes

qui visent à **garantir que chaque décision prise au cours du processus de traitement peut être retracée et expliquée.**

En lien avec l'opacité de ses algorithmes, **l'IA soulève de nombreuses questions juridiques en regard des droits de propriété intellectuelle.** Les données d'entraînement peuvent inclure des informations, des créations, des signes ou des inventions protégés par des droits de propriété intellectuelle. Se pose alors la question de savoir si les actes de collecte des données, de leur intégration et de leur traitement dans l'outil d'IA constituent des actes de contrefaçon. En outre, l'IA est également en mesure de reproduire des voix et des expressions faciales à partir de quelques images et fichiers audio (hypertrucage). **L'utilisation de l'IA peut donc facilement conduire à une violation des droits de propriété intellectuelle et à des dérives allant jusqu'à l'usurpation d'identité,** des enjeux juridiques majeurs pour les régulations en cours que nous discuterons dans la partie suivante.

Protection de la vie privée

Les algorithmes et les données personnelles qu'ils utilisent sont déjà régulés en Europe par la réglementation RGPD, mais **les technologies d'IA et notamment celles qui reposent sur l'apprentissage automatique sont, par nature, opaques et intrusives, et peuvent porter atteinte aux droits et libertés des individus** par la violation de la vie privée et l'utilisation de quantités massives de données d'entraînement sans le consentement ou la compensation de leurs créateurs. Sans garde-fou, l'IA peut générer une société *orwellienne* susceptible de porter atteinte aux libertés fondamentales. Certaines applications chinoises actuelles peuvent donner un avant-goût de ces risques comme les applications de surveillance de foules et de vérification d'identité ou comme le système de crédit social qui vise à mettre en place un système national de réputation des citoyens et des entreprises à des fins de récompenses ou de pénalités.

Afin de se prémunir de ces risques en France et en Europe, le développement des systèmes d'IA doit prendre en compte les principes fondamentaux du RGPD, dont notamment la licéité, la loyauté et la proportionnalité du traitement.

Impact environnemental

Un des effets délétères les plus immédiatement mesurables des IA telles qu'elles sont développées en ce moment **est leur impact environnemental.** L'empreinte carbone des grands centres de données et des réseaux neuronaux nécessaires à la construction de ces systèmes d'IA est colossale. Il est essentiel d'en tenir compte et d'**encourager autant que possible les solutions alternatives qui utilisent moins de données, moins de puissance de calcul et de promouvoir le développement des technologies qui visent à réduire drastiquement la consommation d'énergie des IA.** Il ne faut sûrement pas vouloir mettre de l'IA partout et pour tout, un projet d'IA est justifié avant tout par les cas d'usage et la maîtrise des impacts.

Emploi et précarité du travail

La *collaboration* avec l'IA inquiète de nombreux travailleurs qui s'interrogent sur la pérennité de leur emploi et sur l'évolution de leurs conditions de travail. Gartner¹⁵ prévoit ainsi que l'IA conversationnelle réduira les coûts de main-d'œuvre des centres de contact de 80 milliards de dollars en 2026. La banque Goldman Sachs estime que 300 millions d'emplois dans le monde seront remplacés par l'IA dans les années à venir. L'Organisation de coopération et de développement économiques (OCDE) prévoit un impact considérable de l'IA sur l'emploi avec 27 % des professions profondément transformées par l'IA

¹⁵ <https://www.gartner.com/en/newsroom/press-releases/2022-08-31-gartner-predicts-conversational-ai-will-reduce-contact>

et près de 10 % des métiers menacés de disparition au cours de la prochaine décennie. Selon cet organisme, quasiment tous les secteurs d'activité et toutes les professions seront affectés par l'IA. Les emplois peu et moyennement qualifiés seront les plus menacés tandis que les professions hautement qualifiées, bien que particulièrement exposées aux progrès récents de l'IA, seront a priori plus résilientes.

Ce qui est certain, c'est que nous allons assister, dans tous les secteurs d'activité, à une disparition progressive de nombreuses tâches répétitives ou faisant appel à de la déduction à partir de règles, qu'elles soient manuelles ou intellectuelles, mais en parallèle à l'émergence de nouvelles tâches nécessitant des compétences spécifiques. **Si les prévisions de l'impact de l'IA sur l'emploi peuvent être perçues comme alarmistes, il apparaît raisonnable d'adopter une vision schumpétérienne et considérer que l'IA, comme cela a été le cas pour les révolutions précédentes, opérera avant tout un déplacement des professions et des emplois avec un enjeu d'élévation et d'adaptation des qualifications qui impliquera des actions très importantes en matière de gestion du changement, d'éducation et de formation.** De nouvelles compétences seront nécessaires, tandis que d'autres deviendront obsolètes. Les gouvernements devront encourager les employeurs à fournir davantage de formation à tous les travailleurs quel que soit leur niveau de qualification, à intégrer les compétences en IA dans l'éducation et à soutenir la diversité de la main-d'œuvre en IA. Luc Julia, l'un des inventeurs du système de commande vocale Siri d'Apple, considère que « toutes ces technologies ont pour but de nous assister dans des tâches ponctuelles, souvent répétitives et fortement codifiées. Elles nous fournissent une aide qui vient amplifier notre humanité, et augmenter nos capacités intellectuelles, mais elles ne peuvent en aucun cas nous remplacer ».

Cette perspective optimiste de transformation du travail doit être relativisée par le développement massif d'une nouvelle forme de travail précaire nécessaire au bon fonctionnement des IA, ce qu'on appelle le *digital labor*, des millions de travailleurs invisibles et précarisés qui annotent, forment et corrigent les systèmes d'IA. Ils se trouvent en grande majorité dans des pays en développement où les salaires sont dérisoires mais le phénomène est global. La France compterait par exemple plus de 250 000 *travailleurs du clic*. **Offrir un cadre protecteur à tous les travailleurs de l'IA oubliés du droit et en particulier, réinscrire les nouvelles formes de travail dans notre système de protection sociale, sont des impératifs.**

Sûreté et sécurité

Les Intelligences Artificielles sont fragiles : elles peuvent se tromper et être trompées. Comme tous les outils, elles peuvent être utilisées à bon ou à mauvais escient mais avec un impact plus important que ce qu'aucun outil nous a permis de faire jusqu'à présent. Citons parmi les risques portant sur la sûreté ou la sécurité :

- **Le manque de fiabilité des systèmes** à fort impact sur les individus ou sur l'environnement vulnérables aux pannes critiques, aux attaques physiques et aux cyberattaques.
- **L'utilisation malveillante** susceptible de constituer une menace pour la vie privée des individus ou la sécurité des États : surveillance des populations, menaces nucléaires, biologiques, menaces portant sur les infrastructures critiques ou la sécurité énergétique.
- **La perte de contrôle humain ou la confiance excessive** envers un système mal configuré avec des conséquences potentiellement désastreuses comme la voiture autonome ou les systèmes d'arme léthale autonome.

Concentration du pouvoir et plateformes

L'IA est source de pouvoir : elle deviendra vraisemblablement la couche de contrôle de l'ensemble des systèmes numériques de demain et jouera un rôle déterminant dans les questions militaires. Le risque de concentration de ce pouvoir autour d'une poignée d'entreprises et de deux énormes puissances, les Etats-Unis et la Chine, est réel. **Si les États européens ne parviennent pas à définir des règles en commun et à déployer leurs propres moyens et compétences en matière d'IA, la domination économique des GAFAM sera inéluctable.** Ces derniers continueront d'influencer les organismes de régulation et d'orienter les décisions en leur faveur. Ils fixeront leurs propres règles du jeu et comme le stipule Naomi Klein « s'empareront unilatéralement de la somme totale des connaissances humaines qui existent sous forme numérique, et les enfermeront dans des produits propriétaires ».

Une étude menée par France Stratégie¹⁶ met en garde contre un effet de *plateformisation* similaire à celui constaté dans le secteur du tourisme ou du transport avec l'émergence d'entreprises comme Booking, Uber ou Airbnb. Cette situation pourrait conduire à **une concentration encore plus importante de la valeur économique entre les mains des GAFAM** avec pour conséquence de moindres recettes fiscales pour les États et un risque de perte de souveraineté.

Perception culturelle

On parle trop peu des décalages culturels et de leurs effets. Les Japonais, par exemple, font plus confiance à un robot qu'à une personne pour s'occuper des personnes âgées, ce qui nous paraît, en France, inconcevable. Les conséquences sont importantes, pour plusieurs raisons :

- L'idée que l'homme se fait du robot et de l'IA influence les progrès dans ces domaines,
- Si l'opinion publique française ou européenne venait à consommer un divorce avec l'IA (à l'instar de l'Italie qui a banni ChatGPT), elle devrait se passer de nombreux usages, et probablement, à terme, subir des évolutions dictées selon d'autres sensibilités et d'autres valeurs, sans avoir un mot à dire,
- L'impact économique de l'IA dépend fortement de la motivation des utilisateurs à l'adopter.

D'où vient cette méfiance en France ? D'une sensibilité aigüe aux excès du néo-libéralisme, d'un ancrage contre les révolutions technologiques (néo luddisme) ou à des craintes, certaines légitimes, d'autres savamment entretenues par des mouvances idéologiques, complotistes ou même politiques ? Sans doute un peu de tout cela.

Afin d'éviter le risque de tomber dans des comportements de type réflexe ou repli malthusien, il faut veiller à ce que le dialogue et le débat public sur les questions de l'IA soit vivant et bien nourri. Les médias publics apportent une contribution essentielle, à entretenir et même à développer : mais est-ce suffisant ?

¹⁶ <https://www.strategie.gouv.fr/publications/anticiper-impacts-economiques-sociaux-de-lintelligence-artificielle>

Les enjeux réglementaires

Les risques induits par l'IA soulèvent de nombreuses questions en matière de réglementation. **Aux États-Unis, l'approche est celle de la *light touch regulation*.** Cette approche, encouragée par les GAFAM, a eu tendance à s'imposer. Elle pose un problème en Europe, aggravé par la nature extraterritoriale de certaines législations américaines. **L'Union Européenne, en perte continue de souveraineté dans le domaine du numérique, a choisi d'adopter une démarche plus *musclée*** qui se traduit par de multiples réglementations progressives et coercitives sur le numérique et l'IA et qui repose sur l'idée d'une **IA de confiance**. Cette stratégie, guidée par les principes de sécurité et les droits fondamentaux de l'Union Européenne, **fait débat car elle tend à imposer de nombreuses contraintes (financières, procédurales) aux entreprises qui innovent dans le domaine** et tentent de pénétrer un marché déjà largement dominé par les géants américains ou chinois.

L'IA de confiance

Un groupe d'experts mandaté par la Commission européenne pour la conseiller sur sa stratégie en matière d'IA a publié en 2019 des lignes directrices éthiques pour une IA digne de confiance et dégagé sept exigences qui constituent le socle de toute nouvelle réglementation ou mesure de régulation :

- **Action humaine et contrôle humain** : en dernier lieu, les décisions importantes doivent être confirmées par des humains,
- **Robustesse technique et sécurité** : les systèmes d'IA doivent fonctionner de manière fiable et stable,
- **Respect des droits fondamentaux** : ne pas porter atteinte aux droits fondamentaux tels que la liberté d'expression et la vie privée,
- **Diversité, non-discrimination et équité** : garantir une équité sociale et économique,
- **Transparence** (traçabilité, explicabilité, communication) : les décisions doivent être explicables pour les utilisateurs et les parties prenantes,
- **Bien-être sociétal et environnemental** : respecter les droits des personnes et protéger l'environnement, tout en promouvant le bien-être social,
- **Responsabilité et redevabilité** : les entreprises et les personnes responsables de la conception et de l'utilisation de systèmes d'IA doivent être tenues responsables de leurs actions et décisions.

La lutte contre l'extraterritorialité

L'Union Européenne continue de lutter par la régulation ou la justice contre la doctrine américaine d'extraterritorialité ou du moins tente de s'en accommoder.

En 1978, le Congrès américain vote le **FISA (Foreign Intelligence Surveillance Act)**, base légale des surveillances physiques et électroniques des États-Unis, ainsi que de la collecte d'informations sur des puissances étrangères soit directement, soit par l'échange d'informations avec d'autres puissances étrangères. En 2008, un amendement est ajouté à cette loi, le **FISA Amendments Act of 2008**. Il contient notamment la **section 702**, qui autorise l'administration américaine à collecter, utiliser et partager des données personnelles étrangères stockées sur des serveurs américains. Seule restriction : les personnes ciblées ne doivent pas être américaines.

Alors que la loi FISA couvre les activités de surveillance à l'intérieur des États-Unis, le gouvernement américain peut également mener des activités de surveillance à l'extérieur des États-Unis en vertu de l'**Executive Order 12333**, un décret exécutif signé par le président américain Ronald Reagan en 1981.

Ce décret énonce les principes généraux qui guident l'activité des services de renseignement, notamment la collecte de données étrangères, la protection de la vie privée et des libertés civiles, ainsi que la conduite d'opérations secrètes à l'étranger.

Le **Cloud Act** adopté en mars 2018, est une loi extraterritoriale américaine qui permet aux administrations des États-Unis, disposant d'un mandat et de l'autorisation d'un juge, d'accéder aux données hébergées dans les serveurs informatiques situés dans d'autres pays, au nom de la protection de la sécurité publique aux États-Unis et de la lutte contre les infractions les plus graves dont les crimes et le terrorisme.

Le 16 juillet 2020, le **Privacy Shield** (bouclier de protection des données) qui permet le transfert de données entre l'Union européenne et les opérateurs américains adhérant à ses principes de protection des données sans autre formalité, est invalidé par la Cour de justice de l'Union européenne par l'arrêt **Schrems II**.

Par une décision du 10 juillet 2023, la Commission européenne constate finalement que les États-Unis assurent un niveau de protection des données personnelles équivalent à celui de l'Union européenne. Les transferts de données personnelles depuis l'Union Européenne vers certains organismes états-uniens peuvent à nouveau s'effectuer librement, sans encadrement spécifique. Selon de nombreux commentateurs, il s'agit d'une concession de l'Union Européenne : le prétendu nouveau cadre transatlantique de protection des données personnelles est en grande partie une copie du Privacy Shield qui a échoué. Malgré les efforts de relations publiques de la Commission européenne, la législation américaine et l'approche adoptée par l'Union Européenne n'ont guère changé. Le problème fondamental de la FISA 70217 n'a pas été abordé par les États-Unis, qui considèrent toujours que seuls les ressortissants américains peuvent bénéficier de droits constitutionnels.

Les évolutions de la réglementation européenne

Afin de soutenir les stratégies nationales des états membres, L'Europe s'est attachée à définir des règles normatives communes encadrant l'usage du numérique et de l'IA et permettant le développement d'une économie européenne de la donnée, notamment à la faveur du développement du marché des données industrielles. :

- Le règlement **RGPD (Règlement Général sur la Protection des Données)** encadre le traitement des données de manière égalitaire sur tout le territoire de l'Union Européenne. Il est entré en application le 25 mai 2018.
- Le règlement **Data Governance Act (DGA)** publié le 25 novembre 2020 tend à favoriser le partage des données personnelles et industrielles par la mise en place de structures d'intermédiation. Le règlement vise également à renforcer la confiance dans le partage des données (relatives à la santé, de mobilité, environnementales, agricoles...) en rendant ce dernier plus sûr, facile et conforme à la législation sur la protection des données.
- Le règlement **Data Act**, adopté en juillet 2023, tend à maximiser la valeur des données dans l'économie en facilitant la disponibilité et le partage des données entre les entreprises, les consommateurs et les organismes publics. Complétant le DGA, il précise qui peut créer de la valeur à partir des données et dans quelles conditions. Il comporte une obligation de partage des données aux utilisateurs pour les détenteurs de données, à savoir les fabricants de produits connectés et fournisseurs de services connexes.

¹⁷ Foreign Intelligence Surveillance Act de 1978 décrivant les procédures des surveillances physiques et électronique, ainsi que la collecte d'information sur des puissances étrangères soit directement, soit par l'échange d'informations avec d'autres puissances étrangères.

- Le règlement **DMA (Digital Markets Act)** entend prévenir les abus de position dominante des géants du numérique, que sont en particulier les GAFAM (Google, Apple, Facebook, Amazon et Microsoft), et offrir un plus grand choix aux consommateurs européens. Il est progressivement applicable depuis le 2 mai 2023.
- Le règlement **DSA (Digital Services Act)**, prévoit de lutter contre les contenus et les produits illégaux en ligne (haine, désinformation, contrefaçons...). L'objectif est de faire de l'Internet un espace plus sûr pour les utilisateurs. Avec ce texte, « ce qui est illégal dans le monde physique le sera aussi en ligne ». Les obligations prévues par ce texte doivent entrer en application le 17 février 2024. Les très grandes plateformes en ligne et les très grands moteurs de recherche sont concernés depuis le 25 août 2023.
- Le **Cybersecurity Act**, entré en application le 28 juin 2021, instaure un cadre européen de certifications de cybersécurité à l'échelle de l'Union pour les produits, services et processus digitaux. L'une des questions posées pour ce cadre concerne le critère d'immunité aux lois extraterritoriales du schéma européen pour le *cloud* (*European Cybersecurity Certification Scheme for Cloud Services*, EUCS).
- La régulation portant sur l'IA, l'**IA Act** :
 - o Les principes et objectifs concernant l'IA ont été définis par la Commission dès 2021.
 - o Le Parlement européen a adopté sa position concernant le règlement sur l'IA le 14 juin 2023 avec un objectif d'application du règlement en 2026.
 - o Le 8 décembre 2023, après d'intenses négociations entre les États membres, le Parlement européen a voté en faveur d'un premier accord de réglementation, l'IA Act.

L'IA Act

L'IA Act vient compléter les règlements précédents avec **un ensemble de règles harmonisées qui concernent la conception, le développement et l'utilisation de certains systèmes d'IA.**

L'approche de la proposition de règlement est fondée sur les risques et distingue **quatre niveaux pour les systèmes d'IA : risque inacceptable, risque élevé, risque limité, risque minime.** La classification des risques repose sur la finalité assignée au système d'IA, conformément à la législation existante de l'UE en matière de sécurité des produits. Ce cadre juridique, récemment adopté au Conseil et au Parlement européen, s'appliquera pour tout système d'IA mis sur le marché dans l'Union ou dont l'utilisation a une incidence sur des personnes situées dans l'Union.

Concernant les systèmes à risque élevé, des normes de qualité et de sécurité s'appliquent, telles que la traçabilité de l'utilisation de la technologie, la transparence vis-à-vis des utilisateurs ainsi que la nécessité d'un contrôle humain. Ils doivent par ailleurs « atteindre un niveau approprié d'exactitude, de robustesse et de cybersécurité ». Des contrôles ex ante et ex post sont également prévus. Le fournisseur devra enregistrer officiellement son système d'intelligence artificielle dans une base de données de l'UE après une évaluation de sa conformité. Parmi les systèmes à risque élevé, on trouve en particulier ceux destinés aux infrastructures critiques, à l'éducation, aux ressources humaines et au maintien de l'ordre.

Les sanctions pourront aller jusqu'à 30 millions d'euros ou 6 % du chiffre d'affaires en cas de non-respect des règles relatives aux pratiques prohibées ou à l'usage des données.

Concernant les applications comportant un risque faible ou minimal d'atteinte aux droits fondamentaux, les fournisseurs sont encouragés à appliquer, sur la base du volontariat, des codes de conduite facultatifs.

Pour une régulation efficace

Comme le souligne Joelle Toledano (économiste, membre du Conseil national du numérique et de l'Académie des technologies) dans un article du Monde de janvier 2023 : « sans régulation efficace, la politique de souveraineté numérique européenne échouera ». Il convient de s'interroger justement sur l'*efficacité* de la réglementation récemment adoptée par l'UE (IA Act). **L'évolution de l'IA est tellement rapide que la réglementation risque d'être toujours en retard, avec en arrière-plan le risque de mettre en opposition la régulation et les capacités d'innovation des entreprises.**

Une note d'information du CIGREF¹⁸ de juillet 2023 résume bien le risque de positionner la régulation comme condition préalable à l'innovation : « quel que soit leur positionnement, toutes les organisations sont unanimes pour dire que **le plus gros risque est de passer à côté ou d'être en retard sur la transformation induite par les IA génératives. Les risques et la sécurité sont à gérer en parallèle et non en préalable à la réflexion sur les opportunités** ».

Un système plus souple de régulation et de contrôle réactif gouverné par des instances adaptatives et agiles mérite donc d'être étudié. Il doit répondre au principe d'*agentivité* des technologies, dès lors que ces dernières viennent à se diffuser partout auprès de tous les publics. **Leurs effets de long terme sont imprévisibles, et nécessitent la mise en place d'instances de surveillance et de réaction si nécessaire.**

Il convient enfin de faire preuve de pragmatisme et de capitaliser sur les systèmes de gestion des risques en vigueur. Dans les secteurs sensibles comme celui de la santé par exemple, où les décisions peuvent affecter l'intégrité physique ou les droits humains, des règles éthiques et des protocoles scientifiques rigoureux existent déjà avant d'accepter qu'un nouveau système entre en fonction ou soit commercialisé. Ces règles et protocoles doivent, bien entendu, être renforcées et adaptées aux spécificités de l'IA, notamment quand elle crée des applications entièrement nouvelles comme la conduite autonome ou les systèmes de conversation, afin de garantir que ces applications soient utilisées de manière responsable et sûre.

Vers une convergence des régulations ?

Les positions américaines et européennes restent encore éloignées mais des efforts de convergence se dessinent. La question de l'accès aux données reste un irritant très sensible. Aux États-Unis, le débat sur la responsabilité des plateformes numériques a évolué ces dernières années à la suite de plusieurs affaires à fort retentissement médiatique : l'affaire de la société Cambridge Analytica¹⁹, l'audition au Sénat américain de l'ex-ingénieure de Facebook Mme Frances Haugen²⁰, et les procès Gonzales versus Google ou Twitter versus Taamneh où les plaignants ont réclamé que Google et Twitter soient tenus responsables des recommandations formulées par leurs algorithmes.

En Europe, la convergence des systèmes nationaux a appelé une intervention forte à l'échelon de l'Union européenne (le *Digital Service Act*, le *Digital Market Act*, l'*IA Act*). Aux États-Unis, le président Joe Biden a publié le 30 octobre dernier un document sur les règles que devront suivre les États-Unis en matière d'Intelligence Artificielle intitulé : « Executive Order on the Safe, Secure, and Trustworthy

¹⁸ CIGREF : Note d'information et d'actualité - Recommandations au sujet des IA génératives de juillet 2023
<https://www.cigref.fr/wp/wp-content/uploads/2023/07/Cigref-Note-dinformation-et-dactualite-Recommandations-au-sujet-des-IA-generatives-Juillet-2023.pdf>

¹⁹ La Société Cambridge Analytica aurait utilisé les données personnelles de 87 millions de personnes, données recueillies et exploitées à leurs insu, dans le cadre de la campagne en 2016 de Donald Trump.

²⁰ En octobre 2021, Mme Frances Haugen, lanceuse d'alerte, a expliqué certaines pratiques internes à la société Facebook posant de graves problèmes éthiques.

Development and Use of Artificial Intelligence ». Ce document, pragmatique, orienté vers l'action et le déploiement rapide des directives, appelle en particulier à la coopération internationale pour qu'un consensus se fasse sur une grande partie des idées défendues. En outre, il met en évidence **la nécessité pour l'Europe de décliner rapidement et de manière très opérationnelle** son processus de régulation.

Certains commentateurs en Europe ont jugé cet Executive Order insuffisant car non coercitif. Mais il s'agit bien d'une initiative structurante, comme l'indique le volume des référentiels élaborés par le NIST en amont, et les dispositions déjà prises par de multiples départements d'État pour le mettre en œuvre chacun dans son domaine.

Le 1^{er} novembre 2023, Le Royaume-Uni a organisé à Londres (Bletchley Park) **le premier évènement international sur les risques de l'IA**, qui a réuni vingt-huit pays ainsi que de nombreux acteurs ayant participé à l'essor de l'IA ces dix dernières années. Malgré une atmosphère de coopération positive, ce rassemblement a donné lieu à d'épineux débats entre ceux (les Français) qui craignent de voir l'innovation bridée par la régulation et la constitution de monopoles disqualifiant la concurrence et ceux (les Anglais et les Américains) qui dénoncent les risques de *déferlement* susceptibles de déstabiliser le monde et les états-nations, tels que dénoncés par Mustafa Suleyman dans son dernier ouvrage²¹ sur l'IA.

La France et l'Europe peuvent, et doivent, se montrer pionnières dans l'élaboration d'un cadre juridique de l'IA solide et pertinent tout en faisant preuve de pragmatisme, d'efficacité et d'ouverture aux propositions de convergence transnationales.

²¹ <https://www.fayard.fr/livre/la-deferlante-9782213726687>

Les capacités de mise en œuvre en France et en Europe

L'IA comme source de puissance en fait un facteur de compétition entre les États-Unis et la Chine et contraint l'Europe à s'engager sur le même chemin même si elle le fait avec des moyens et des visées différents. L'action de l'Europe est indispensable pour la bonne exécution de la stratégie française en matière d'IA : l'Europe intervient de manière substantielle pour réguler les usages, fixer les règles pour la gestion des données, financer la recherche et favoriser l'émergence de champions industriels. Dans ce contexte, quelles sont les capacités réelles de développement et de mise en œuvre d'une industrie autour de l'IA en France et en Europe ?

La domination américaine et chinoise

Les acteurs américains dominent dans tous les segments de l'IA (investissement dans les start-ups, financement de la recherche et du développement, ...) avec une forte concurrence de la Chine.

Le secteur privé américain est très dynamique et investit massivement dans cette technologie, comme le montre le graphique ci-après.

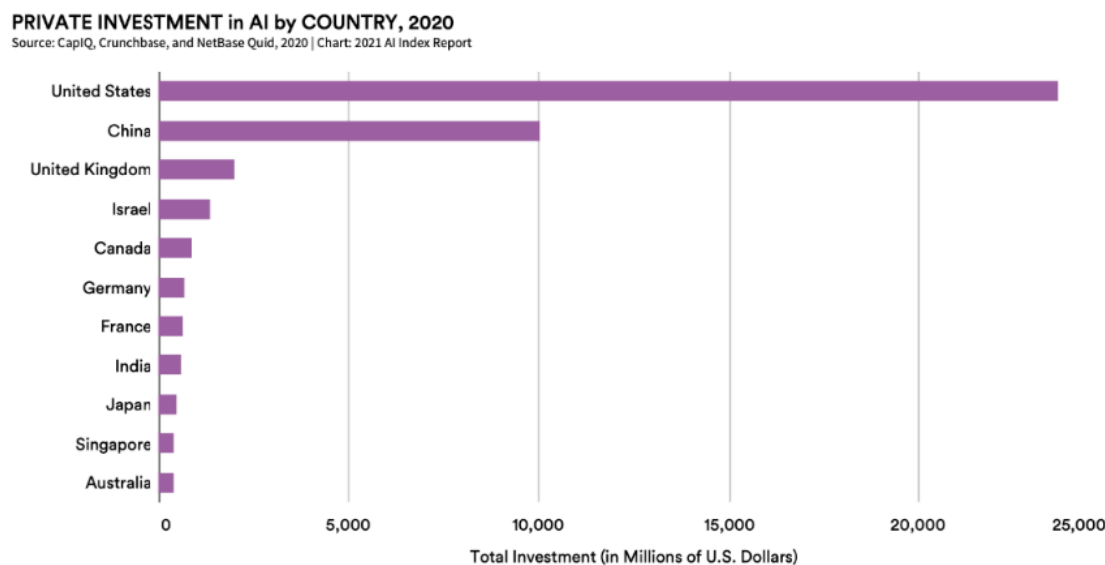


Figure 2 - Investissement du secteur privé en IA par pays (2020)

Les GAFAM dominant largement le marché avec leurs homologues chinois, les BATX (Baidu, Alibaba, Tencent, Xiaomi) et Huawei.

Concernant les dépôts de brevets en IA, les États-Unis et la Chine sont au même niveau (sept entreprises dans le top 20 pour chaque pays). Néanmoins, si on prend en compte que la plupart des grandes entreprises chinoises sont largement subventionnées et accompagnées par l'État, que les publications de recherche des GAFAM sont davantage citées et qu'ils exportent bien plus leurs solutions dans le reste du monde que leurs homologues chinois, **l'écosystème américain de la tech reste le plus dynamique et le plus influent.**

Concernant les montants globaux investis, le rapport *Private Equity Investment in Artificial Intelligence* de l'OCDE donne aussi quelques ordres de grandeur : Les États-Unis investissent 70 fois plus que la France dans les jeunes pousses, ce qui, ramené à l'échelle des PIB respectifs, représente presque sept fois plus d'investissement.

Total investments from 2012 to 2020 by industry and country of start-up, in USD millions

	Mobility and autonomous vehicles	Healthcare, drugs and biotechnology	Business processes and support services	IT infrastructure and hosting	Media, social platforms, marketing	Financial and insurance services	Digital security	Robots, sensors, IT hardware	Logistics, wholesale and retail	Consumer products	Construction and air conditioning	Education and training	Other activities	Total
United States	52 423	21 866	19 510	18 029	13 027	12 076	10 534	5 880	6 627	5 081	2 193	1 643	4 624	173 513
United Kingdom	163	1 599	1 611	1 259	1 275	1 960	650	788	101	43	7	54	58	9 567
Israel	217	906	302	1 129	517	120	925	1 294	151	22	40		304	5 926
Germany	122	505	718	455	746	1 276	167	187	30	5	14	63	386	4 675
Canada	30	709	364	796	745	405	56	451	108	39		25	495	4 224
Japan	814	272	509	653	577	497	60	234	40	58	20	100	110	3 944
France	94	402	196	451	673	265	93	56	166	81	65	18	25	2 585

Source : OCDE - *Private Equity Investment in Artificial Intelligence rapport*

Figure 3 - Total des investissements des start-up, par pays et par secteur, 2012 et 2020, en M\$

Le montant des budgets publics de soutien à l'innovation en Europe, relativement faible par rapport à celui des Etats-Unis, n'est certainement pas le seul responsable de cette situation : l'écart gigantesque et qui perdure entre la surface financière des grands fonds d'investissement états-uniens et celle des acteurs européens crée un effet d'aspiration très puissant des startups les plus fructueuses, pendant leur deuxième ou troisième tour d'investissement, et l'Europe est en panne pour endiguer le phénomène.

En France, les racines du problème sont connues et pour certaines profondes : complexité administrative, système d'exonérations et niches fiscales protéiformes et globalement peu lisibles, poids du système de retraite sur la disponibilité de capitaux, instabilités fiscales et juridiques génératrices de risques, système de taxation de plus-values mouvant... Les enjeux de la révolution de l'IA susciteront-ils un sursaut qualitatif ambitieux ?

En tout état de cause, la position dominante américaine, concurrencée par la Chine, est une forte contrainte que les autres pays doivent nécessairement prendre en compte.

Le soutien de l'Europe aux stratégies nationales

Au-delà de définir des règles normatives communes, l'Union Européenne intervient pour soutenir les stratégies nationales, dont celle de la France, en instituant divers programmes budgétaires de soutien à la recherche et en posant de nouvelles règles pour favoriser l'émergence de champions européens industriels. Compte tenu du contexte de domination américaine et chinoise dans l'IA, ce soutien est indispensable.

La Commission a publié en 2021 un nouveau plan coordonné 2021 sur l'IA. Il définit les domaines de coopération et encourage les États membres à développer des plans stratégiques nationaux. Les axes de travail sont déclinés autour de quatre thèmes principaux : **poser les conditions permettant le développement de l'IA, faire de l'Europe le lieu de l'excellence du laboratoire au marché, assurer que l'IA est une force positive pour la société, construire des stratégies de leadership dans les secteurs à fort impact.**

Les programmes budgétaires de soutien à la recherche

Les deux principaux programmes sont les programmes Horizon Europe et Pour une Europe numérique (Digital Europe). **La Commission prévoit d'investir 1 Md€ par an dans l'Intelligence Artificielle par ces**

deux programmes. Citons également **le programme européen pour la santé EU4Health** et l'investissement dans le cadre de **l'entreprise commune EuroHPC pour un supercalculateur européen.** Ce supercalculateur exaflopique haut de gamme sera hébergé par le centre de calcul de la commission française des énergies alternatives et de l'énergie atomique à Bruyères-le-Châtel (France) et exploité par le consortium Jules Verne. Il représente un investissement commun partagé entre la France, les Pays-Bas et l'Union Européenne d'environ 540 millions d'euros. L'Union Européenne contribuera à hauteur de 50 % au total des coûts du programme.

La construction d'un écosystème *AI-friendly*

Ce serait une erreur de concentrer les moyens de l'Europe sur l'improbable émergence d'un clone des GAFAM ou d'un géant de l'IA générative. L'Europe peut et doit capitaliser sur ses atouts et ses grands acteurs industriels afin de construire un écosystème *AI-friendly* incluant entre autres : les technologies de composants à très basse consommation, les architectures électroniques ouvertes, évolutives, modulables et neuronales, l'Edge Computing, l'IA embarquée, les processeurs pour les calculateurs exascale, les supercalculateurs, les calculateurs quantiques et les plateformes de partage et de recherche ouvertes.

Le projet de Cloud souverain européen est une illustration des tâtonnements de l'Europe sur la stratégie à adopter en matière d'infrastructure numérique. Faisant suite à plusieurs échecs de mise en place d'un cloud souverain (CloudWatt, Numergy), la France s'est associée à l'Allemagne pour lancer en 2019 le **projet d'infrastructure sécurisée de données européennes de nouvelle génération GAIA-X.** Influencé par les intérêts d'entreprises extra-européennes, les GAFAM en particulier, ce projet n'a malheureusement abouti pour l'instant à aucune réalisation concrète, il s'est contenté de définir un corpus de normes pour la mise en commun des données d'entreprises européennes. Une déclaration commune des 27 États membres a, en outre, été signée le 15 octobre 2020, afin de lancer une alliance européenne pour les données industrielles et le cloud. Enfin, **un projet important d'intérêt européen commun (PIIEC) sur les infrastructures et services cloud de nouvelle génération,** initié par la France et l'Allemagne et réunissant 11 États membres, a été pré-notifié en avril 2022. Quel que soit le destin de ces initiatives, la question de savoir si l'Europe pourra un jour gagner son indépendance en matière de Cloud et tourner le dos aux Clouds américains reste ouverte tellement le retard est conséquent. Alors que les enjeux de confiance, de sécurité et de résilience sont plus que jamais d'actualité, **les décideurs politiques et les instances de soutien à l'innovation ont intérêt à soutenir des initiatives moins ambitieuses qui visent à répondre à des besoins spécifiques pour les services publics ou les acteurs critiques de l'économie et qui capitalisent sur les briques de confiance qui ont émergé en France et Europe ces 10 dernières années.** La création de l'offre Cloud NumSpot²² par Docapost, avec Dassault Systèmes, Bouygues télécom et la Banque des Territoires, est à ce titre une démarche pertinente.

Le contrôle des acquisitions étrangères dans l'UE.

Le cadre européen pour **le filtrage des investissements directs étrangers (IDE) est devenu opérationnel** le 11 octobre 2020. Le 25 mars 2020, la Commission avait publié des orientations à

²² <https://numspot.com/>

l'intention des États membres, appelant notamment tous les États membres à mettre en place un mécanisme de filtrage.

La stratégie française

Le Gouvernement français a lancé **une stratégie nationale pour l'intelligence artificielle (IA)**, en 2018 qui s'articule en deux phases.

La première phase (2018-2022) consiste à doter la France de capacités de recherche compétitives. Cette première étape a été financée à hauteur de 1,85 milliard d'euros. Elle a notamment financé la création et le développement d'un réseau d'instituts interdisciplinaires d'intelligence artificielle, la mise en place de chaires d'excellence et de programmes doctoraux, ainsi que le déploiement du supercalculateur Jean Zay²³.

La seconde phase consiste à diffuser des technologies d'intelligence artificielle au sein de l'économie et à soutenir le développement et l'innovation dans des domaines prioritaires comme l'IA embarquée, l'IA de confiance, l'IA frugale et l'IA générative. Cette seconde phase est dotée de 1,5 milliard d'euros dans le cadre de France 2030²⁴. Cette nouvelle phase s'articule autour de trois piliers stratégiques : le soutien à l'offre deep tech, la formation et l'attraction des talents, le rapprochement de l'offre et de la demande de solutions en IA.

Faisant suite à la vision apportée par la mission Villani²⁵, la stratégie nationale pour l'intelligence artificielle a jeté les bases d'une structuration de long terme de l'écosystème d'IA, à tous les stades du développement technologique : recherche, développements et innovations, applicatifs, mise sur le marché et diffusion intersectorielle, soutien et encadrement du déploiement.

Le 19 septembre 2023 le Comité de l'intelligence artificielle générative²⁶ a été lancé. Ce Comité réunit des acteurs de différents secteurs (culturel, économique, technologique, de recherche), il vise à éclairer les décisions du gouvernement et de faire de la France un pays à la pointe de la révolution de l'IA.

Les capacités de mise en œuvre

La prise de conscience des enjeux de l'IA a eu lieu avec un certain retard en Europe alors que les Etats-Unis et la Chine ont réalisé des investissements considérables dans la technologie dès 2015. **Malgré la qualité de sa recherche et ses talents de niveau mondial, la France et l'Europe se retrouvent coincées entre ces deux superpuissances dont les visions de l'IA sont très différentes, chacune déterminée à distancer l'autre.**

L'Europe manque d'acteurs industriels majeurs du numérique et de l'électronique et **les investissements en IA ne sont pas, à date, suffisants pour rattraper le retard.** A l'instar de ce qui s'est passé avec les moteurs de recherche, il est peu probable que l'Europe soit capable de créer des outils généraux compétitifs dans le domaine de l'IA générative. Un des principaux problèmes pour le développement de l'IA de ce type en Europe réside dans sa capacité à disposer, de manière souveraine, de la puissance de calcul nécessaire et des microprocesseurs adaptés (les fameux processeurs graphiques GPU). Les GAFAM ont réalisé des investissements colossaux dans les infrastructures d'IA afin de rester à la pointe et de continuer d'occuper une position hégémonique sur le marché des infrastructures et des services numériques. **Ces niveaux d'investissement ne sont pourtant pas**

²³ <https://www.cnrs.fr/fr/presse/jean-zay-le-supercalculateur-le-plus-puissant-de-france-pour-la-recherche>

²⁴ <https://www.gouvernement.fr/france-2030>

²⁵ <https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>

²⁶ <https://www.economie.gouv.fr/comite-intelligence-artificielle-generative>

forcément inaccessibles à l'échelle européenne, mais ils nécessitent un vrai effort et une volonté politique. L'écosystème numérique français a des capacités financières non négligeables, et si on ajoute celles des autres pays européens, des moyens importants peuvent être mobilisés. **Il convient pour cela de réorienter progressivement les dépenses des Européens vers des solutions européennes, parce que les investissements dans l'offre ne suffiront pas.** Cela suppose entre autres une révision des modalités de la commande publique au niveau européen. En outre, dans un contexte de fragmentation du web mondial en une multitude de sous-espaces régionaux (effet de splinternet), l'Europe pourrait se retrouver seule avec les Etats-Unis avec le risque de colonisation numérique par ces derniers. **Elle a donc intérêt à nourrir un dialogue et à coopérer sur les questions de souveraineté numérique et d'autonomie stratégique avec les États (Inde, Brésil, Malaisie, Nigéria, ...) qui restent partagés à l'idée de se subordonner aux deux leaders mondiaux.**

Cet état de fait n'exonère pas pour autant les entreprises françaises d'avancer lorsqu'elles le peuvent, d'autant que pas mal d'entre elles disposent déjà aujourd'hui de grosses capacités en termes d'infrastructures et d'importants soutiens financiers. Récemment, Xavier Niel, patron du groupe Iliad (Free, 42, Station F), et Rodolphe Saadé, PDG du transporteur maritime CMA CGM, soutenus par l'ancien patron de Google Eric Schmidt, ont annoncé le lancement d'un laboratoire à but non lucratif dédié à la recherche en intelligence artificielle. Baptisé Kyutai, ce centre basé à Paris aura une approche open source et sera doté d'un premier fonds de 300 millions d'euros. La filiale du groupe La Poste, Docaposte, entend aussi prendre la vague de l'intelligence artificielle générative en lançant une offre d'IA générative souveraine (robot conversationnel) avec trois entreprises françaises : LightOn, Aleia et enfin, NumSpot dont elle est l'une des initiatrices. Enfin, le jeune pousse française Mistral.ai (créée en avril 2023) vient de rejoindre la licorne européenne (allemande) Aleph Alpha. La dynamique est donc en marche.

Au-delà de l'enjeu des infrastructures et de la fourniture des services fondations de l'IA, **l'Europe ne doit pas manquer l'opportunité de l'utilisation de l'IA dans tous les secteurs d'activité.** De multiples chaires à forte composante en IA ont été créées en France, associant des académiques et des grands donneurs d'ordre. SAP se positionne de manière audacieuse et pertinente pour accompagner des porteurs de projets d'applications de l'IA dans de nombreux domaines. La plupart des grandes sociétés de services (SSII) françaises proposent déjà des services pour les projets d'IA associés au cloud et comptent se développer fortement sur ce marché porteur. En outre, **l'Europe n'a pas à rougir en termes de talents, d'idées et de concepts novateurs et peut aussi tirer parti de la taille importante de son marché intérieur.** Elle doit continuer de faire course en tête sur la mise au point de systèmes d'IA dignes de confiance afin de mieux répondre aux préoccupations des populations et des entreprises. Il s'agit d'un pari ambitieux, potentiellement porteur sur le long terme mais non exempt de risques car l'IA Act pourrait faire pression sur certaines dynamiques d'innovation et brider les investissements privés.

On peut donc espérer que l'Europe ne subira pas de déclassement dans le domaine de l'IA, mais c'est tout sauf une certitude, pour au moins deux raisons :

- Comme pour Internet, les leaders auront de bonnes chances de remporter la majorité des marchés, et actuellement ils ne sont pas européens, même si le nombre de licornes en IA ou qui y recourent massivement y est en croissance forte,
- Le marché de Etats-Unis est deux fois et demi plus important que le marché européen, et l'opinion publique européenne est plus réservée, voire méfiante, qu'ailleurs : on peut donc s'attendre à ce que la dynamique d'innovation sur les marchés grand public soit bien plus faible en Europe.

L'imprégnation réussie de tous les pans de l'économie par l'IA représente une fenêtre d'opportunité pour la France et l'Europe, ou au contraire, un ensemble de risques tels qu'elles seront déclassées, à une échelle de temps que l'on peut estimer, au rythme actuel entre 5 et 10 ans. En d'autres termes on peut considérer qu'en 2035 au plus tard, les jeux seront faits en grande partie.

Les mesures à prendre, les orientations stratégiques

Compte tenu des développements précédents, nous, progressistes pour la social-démocratie, reconnaissons que la technologie d'IA est un sujet politique et pas seulement par ses implications éthiques ou sociales. Nous souhaitons promouvoir les développements et les applications de l'IA sûrs et dignes de confiance, respectueux des droits fondamentaux et de la vie privée, bénéfiques à l'innovation, à l'emploi et à l'inclusion, et susceptibles de hisser la France et l'Europe au rang des leaders mondiaux du domaine sur les plans juridique et économique.

Les mesures et les orientations suivantes nous paraissent essentielles :

Pour une IA sûre et digne de confiance

Nous pensons que **les humains doivent rester aux postes de contrôle** et qu'ils ne doivent jamais laisser les systèmes d'IA les remplacer dans la prise de décision, mais toujours les seconder. En outre, nous souhaitons promouvoir **une innovation responsable** qui puisse garantir l'avènement d'une **IA sûre et digne de confiance**.

Afin de ne pas brider l'innovation, **le principe de responsabilité doit prévaloir à celui de précaution**. En ce sens, les systèmes d'IA doivent pouvoir expliquer leurs résultats, garantir la traçabilité de leurs données et de leurs traitements et doivent faire l'objet d'évaluations robustes, fiables, reproductibles et standardisées avant d'être utilisés. **Nous soutenons et encourageons donc le développement des techniques d'Explainable Artificial Intelligence (XAI)**, un ensemble de processus et de méthodes qui permettent aux utilisateurs de comprendre et de se fier aux résultats et aux informations générés par les algorithmes de Machine Learning de l'IA. Nous soutenons, en amont de tout développement de système d'IA complexe à forts enjeux, ou destiné à soutenir voire prendre des décisions critiques, la conduite systématique d'une analyse indépendante et impartiale des rôles et responsabilités des parties prenantes. En outre, nous demandons aux systèmes d'IA de faire leur meilleur effort pour **rendre possible l'identification, l'authentification et le traçage des données** qu'ils utilisent afin de **lutter contre les biais notamment de genre, ethniques et culturels** encore trop présents ou encore afin de **limiter la production de contenu illégal ou potentiellement dangereux** (matériel d'abus sexuel sur des enfants, hypertrucage). Nous encourageons, à ce titre, **la systématisation des dispositifs techniques comme les tatouages numériques**, qui permettent de résoudre le problème de l'authentification et la détection des contenus artificiels.

Nous exigeons des systèmes d'IA qu'ils soient robustes et résilients et préconisons la mise en œuvre de normes et de procédures permettant aux développeurs de tester et d'identifier les failles et vulnérabilités potentielles, en particulier quand les systèmes d'IA sont embarqués dans le contrôle ou la sécurisation d'infrastructures et d'applications critiques. Les risques de cybersécurité spécifiques à l'IA doivent faire l'objet d'une attention particulière.

Nous encourageons les régulations, les standards, les normes ou les systèmes qui visent à **lutter contre tout utilisation malveillante de l'IA, contre la perte de contrôle humain ou la confiance excessive**. Dans cette perspective, les fournisseurs de services de l'IA en cloud devraient avoir l'obligation d'informer systématiquement les autorités compétentes dès lors qu'une personne, une entreprise ou une organisation effectue une transaction dans le cloud pour former un grand modèle d'IA avec des capacités qui pourraient être utilisées dans des activités malveillantes.

Nous préconisons **une évaluation complète des outils d'IA dont les résultats pourraient représenter des menaces nucléaires, chimiques, biologiques, ou porter atteinte aux infrastructures critiques et à la sécurité énergétique**.

Pour la promotion de l'innovation

Afin de promouvoir l'innovation, nous soutenons les politiques d'investissements dans l'éducation, la formation, la R&D, et dans les capacités souveraines et encourageons les régulations qui visent à lutter contre la collusion illégale et le monopole de certaines entreprises extra-européennes sur les actifs et technologies clés.

Nous pensons qu'il est nécessaire de disposer d'une **politique industrielle européenne ambitieuse dans le domaine de l'IA** et plus largement du numérique.

Cette politique portera l'objectif de **réorienter progressivement les dépenses** des Européens **vers des solutions conformes aux spécificités européennes**, parce que les investissements dans l'offre technologique ne suffiront pas. L'atteinte de cet objectif passera par les actions suivantes :

- **Renoncer à poursuivre des chimères.** L'Europe est fortement distancée sur les Clouds généralistes. Elle a intérêt à développer des offres adaptées à des besoins jusque-là moins bien satisfaits, d'ambition réaliste, en agrégeant des acteurs locaux en écosystème. Il n'est pas certain, par ailleurs, qu'elle rattrapera les leaders états-uniens dans l'IA générative. Elle a intérêt à miser sur des capacités complémentaires qui satisferont des acteurs publics ou semi-publics, et sur d'autres formes d'IA, plus légères, déportées, moins énergivores et plus éthiques.
- **Réviser les modalités ou adapter l'exécution de la commande publique** au niveau européen afin de privilégier les solutions européennes. La commande publique et la régulation doivent agir de concert pour motiver l'hébergement des applications d'IA qui manipulent des données sensibles au sein d'infrastructures souveraines. Il s'agit en premier lieu de répondre à des enjeux évidents de sécurité, mais aussi de **donner l'opportunité à des acteurs souverains de se développer sur le marché porteur des plateformes d'IA en cloud**. Dans cette logique, il nous apparaît nécessaire de migrer sans tarder les applications de santé publique comme le Health Data Hub, solution actuellement hébergé dans le cloud AZURE de Microsoft, vers des solutions d'hébergement souveraines.
- **Intensifier la création de référentiels (normes et standards) permettant de traduire et de structurer les attentes légitimes de l'Europe** par la mise en place de projets collaboratifs et l'instauration d'un dialogue de confiance avec les acteurs industriels et de l'innovation.
- **Lever tous les obstacles à l'innovation** en acceptant, de manière limitée et contrôlée par la puissance publique, la levée temporaire de certaines contraintes réglementaires afin de ne pas brider les créateurs européens dans leurs expérimentations.
- **Lutter contre toute forme de collusion illégale ou de monopole** sans aucunement tourner le dos à des leaders étrangers parce qu'ils sont étrangers
- **Investir massivement dans la R&D et dans les projets d'intérêt commun européens.**

La politique industrielle européenne dans le domaine de l'IA devra également **favoriser la mise en oeuvre des briques constitutives d'un écosystème AI-friendly**. Quelques exemples de briques sont proposés ci-après :

- Le développement de schémas amplifiés de **soutien à l'innovation selon des principes HITL (Human In The Loop) / SITL (Society In The Loop)**²⁷ pour répondre aux enjeux de confiance, de performance et d'efficacité des systèmes d'IA.

²⁷ <https://medium.com/mit-media-lab/society-in-the-loop-54ffd71cd802>

- **L'orientation des moyens vers des projets différenciants** qui pourraient tirer parti de la richesse et de la diversité des données des administrations et entreprises européennes et qui porteraient **des objectifs bénéfiques pour la société** : villes intelligentes, économie circulaire, santé publique, réseaux sociaux de taille humaine, gestion des déchets, applications de lutte contre le réchauffement climatique et les inégalités par exemple.
- **Le soutien aux développements de solutions de robotique collaborative et de jumeaux numériques** dans la médecine, l'industrie et l'agriculture.
- **L'attraction, la formation et la rétention des talents en IA** européens et étrangers

Pour le soutien à l'éducation, à la formation et à l'emploi dans un souci d'excellence, d'équité et d'inclusion

Le développement et l'utilisation responsables de l'IA nécessitent **un engagement à soutenir les travailleurs grâce à l'éducation et à la formation professionnelle et à comprendre l'impact de l'IA sur la main-d'œuvre et les droits des travailleurs.**

Constatant le caractère inéluctable de la diffusion large des systèmes d'IA, constatant, en outre, les risques que cela pose au regard de la cohésion épistémique de nos démocraties, en termes d'accroissement des inégalités et d'exclusion de certaines populations, **nous faisons de la formation une de nos exigences les plus importantes.** Elle est de la co-responsabilité des entreprises, des autorités publiques, des communautés et des individus. Elle commence à l'école et doit être continue.

Nous préconisons une adaptation rapide du dispositif d'éducation et de formation.

A l'école, la mise en place d'un rapport exigeant avec les outils d'IA accessibles au grand public doit être stimulé. Cela peut être envisagé dès le plus jeune âge, à la condition que soient mis en place tous les éléments permettant un contrôle strict d'éventuels effets ou risques sur le développement neurologique, psycho-affectif et social des enfants.

Parallèlement les élèves, devront être mis en garde contre les déviations potentielles des outils d'IA (biais cognitifs, plagiat, hypertrucages et désinformation), ceci pouvant aller de pair avec le développement de l'esprit critique individuel.

Afin de lutter contre le risque de déclassement compétitif de la France, Il conviendra par ailleurs de remettre au goût du jour une formation d'excellence dans les disciplines scientifiques et un temps d'apprentissage approprié.

L'enseignement supérieur devra intégrer la composante IA de manière systématique dans quasiment tous les cursus (scientifiques, ingénierie, sciences humaines). Les formations scientifiques françaises, notamment dans les grandes écoles, procurent un très bon niveau en mathématiques et en informatique pour aborder les spécialisations en IA. Le système français n'a aucunement à rougir du nombre de chercheurs reconnus en IA, même si plusieurs d'entre eux continuent de rejoindre les grands groupes américains.

Les cadres et ingénieurs concepteurs de systèmes et d'applications d'IA devront acquérir une connaissance approfondie des bases théoriques et développer de solides compétences sur les outils d'IA et leur manipulation. Le développement de formations en cursus double, sciences *dures* et sciences sociales, comme pratiquée par Dauphine et certaines écoles d'ingénieurs doit être à ce titre encouragé.

Les utilisateurs professionnels devront, quant à eux, avoir une connaissance suffisante des opportunités et des limites des outils d'IA qu'ils seront amenés à utiliser au quotidien (aide au diagnostic médical par exemple).

Enfin, les enseignants et les formateurs tireront avantageusement parti des outils d'IA pour **moderniser et adapter les techniques de formation et d'apprentissage afin d'inclure tous les publics**.

Les impacts de l'IA sur la structure et la dynamique du marché du travail devront être précisément évalués afin d'anticiper les reconversions certainement nécessaires des intervenants des secteurs fortement automatisables par l'IA vers d'autres secteurs comme ceux des services à la personne où la composante humaine restera essentielle.

En termes de droits et de protection sociale, **il conviendra d'offrir un cadre protecteur aux travailleurs du clic (digital labour) et une juste rétribution des tâches humaines nécessaire à l'IA**.

Pour le respect des droits fondamentaux et de la vie privée

Les politiques en matière d'IA doivent être cohérentes avec la promotion de l'équité et des droits civils.

Nous souhaitons **la mise en œuvre effective de régulations, d'autorités et de systèmes de contrôles qui puissent garantir que les technologies d'IA ne favorisent pas les préjugés et la discrimination** dans un large éventail de domaines : justice pénale (détermination, libération conditionnelle, libération sous caution, gestion des prisons), programmes de prestations sociales, marchés financiers du logement et de la consommation, location immobilière, systèmes d'information de R.H.

Nous demandons **l'accès aux données nous concernant, la capacité d'en définir l'usage et des garanties sur le caractère légal et sécurisé de la collecte, l'utilisation et la conservation des données**.

Nous estimons que **la CNIL est parfaitement indiquée pour assurer la supervision globale de la mise en œuvre de l'IA Act**. Elle sera ainsi en prise directe avec les enjeux de protection des données personnelles et de préservation des libertés individuelles.

Nous reconnaissons que **l'utilisation de l'IA peut facilement conduire à une violation des droits de propriété intellectuelle** et à des dérives allant jusqu'à l'usurpation d'identité. Nous préconisons **la mise en œuvre d'un cadre juridique approprié pour traiter ces risques** et la mise en œuvre systématique de techniques pour mieux les gérer (tatouages numériques par exemple).

Nous reconnaissons le potentiel de l'IA pour **améliorer l'efficacité d'action des forces de l'ordre** à partir du moment où ces dernières **agissent dans le respect de la protection de la vie privée, des droits civils et des libertés civiles**.

Pour la sobriété d'usage et le respect de l'environnement

Nous souhaitons promouvoir les politiques et les projets qui visent à **minimiser l'impact environnemental de l'IA**.

Nous encourageons les efforts de recherche dans les solutions qui minimisent les données, la puissance de calcul, la consommation d'énergie des IA.

Nous encourageons **la sobriété d'usage**, car l'IA n'a pas à être utilisé partout et pour tout, et l'instauration de la primauté du raisonnement humain sur l'IA, non seulement de manière ad hoc pour tous les systèmes d'IA créateurs d'impacts sur les individus, les groupes, le vivant, l'environnement, mais aussi de manière systémique pour assurer la robustesse de cette primauté sur le long terme.

Nous reconnaissons le potentiel de l'IA pour optimiser la **consommation énergétique des systèmes complexes** (réseaux électriques, réseaux de communication, villes ...) et pour **aider à la lutte contre le réchauffement climatique**, sous réserve que son développement reçoive un encadrement permettant d'assurer une mise en œuvre proportionnée alignée sur les impératifs planétaires et de justice sociale.

Pour une régulation adaptative et une gouvernance d'implémentation efficace

Demander l'équilibre entre la liberté d'innovation et la protection des humains et de leurs sociétés n'est pas suffisant. Nous souhaitons **la mise en place de dispositifs réglementaires adaptés à la vitesse d'évolution des IA et des actions concrètes de mise en œuvre des réglementations par la puissance publique.**

Dans cette perspective, nous préconisons une **régulation plus souple, évolutive, proactive et réactive** :

- Nous partons du principe que même en présence de *bacs à sable réglementaires*, il faut pouvoir tester dans des conditions réelles, notamment en utilisant spontanément des jeux de données de test.
- Nous proposons la création d'une **entité administrative européenne indépendante dotée d'une mission de régulation réactive** pour tenir compte de la dynamique technologique afin d'adapter les obligations et contraintes de l'AI Act au contexte du moment.

Nous encourageons **les partenariats internationaux pour le développement d'un cadre de gestion global des risques de l'IA et l'établissement de normes techniques mondiales** afin de libérer tout le potentiel positif de l'IA et de promouvoir une approche commune face aux défis.

Nous pensons qu'il faut également **faire preuve de pragmatisme et capitaliser sur les systèmes de gestion des risques en vigueur dans les secteurs à risque** (santé, armée, aviation civile, ...).

Enfin, à l'instar de l'*Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence* signé par le président américain Joe Biden en octobre 2023, nous pensons qu'**une déclinaison très opérationnelle et objectivée dans le temps des grandes orientations juridiques et politiques de l'IA** (IA Act, stratégie nationale française pour l'IA) au sein de l'ensemble des organismes publics français et européens de régulation, de normalisation, de financement ou de recherche, **est un impératif à court-terme pour rester dans la course à l'innovation et à la compétitivité.**

Groupe de travail IA

Rapporteur : **Louis Cougouille** (Ingénieur)

Autres rédacteurs et contributeurs : **Frédéric Tatout** (Ingénieur), **Francis Pisani** (Auteur, Journaliste), **Jean-Claude Laroche** (Ingénieur, Président du CIGREF), **Philippe Picard** (Ingénieur), **Xavier Labouze** (Auteur, Maître de Conférence en Mathématiques)

Relecteurs : **Jacques-Roger Machart**, **Rémi Thomas**