LA FRANCE CONTRE LES ROBOTS ?

Automatisation et Intelligence Artificielle

LA FRANCE CONTRE LES ROBOTS ?

Automatisation et Intelligence Artificielle

Philippe Lecigne

« *Les ordinateurs ne pensent pas. Mais ils font parfois le lien mieux que nous* »
Henning Mankell, *le guerrier solitaire*

I. En guise d'introduction

Alors qu'au milieu du XIX^{e} siècle le chemin de fer est en passe de remplacer définitivement la diligence, nombre d'esprits éclairés, à l'instar du savant François Arago ne manquent pourtant pas de faire part de leurs craintes quant aux conséquences néfastes de la nouvelle invention, sur la santé des passagers par exemple.

Depuis cette époque, l'automobile et l'avion se sont imposés, et l'on voit ce dernier susciter l'ire des partisans du retour à un passé fantasmé. Dans l'industrie, il y a des années que la machine est devenue omniprésente, au point de remplacer l'homme dans bien des usages ; et le logiciel est devenu le compagnon, voire le substitut incontournable de notre vie professionnelle mais aussi privée, et son emprise ne cesse de s'étendre. Mark Andreesen, l'un des pionniers de l'internet a eu cette phrase terrible : « *software is eating the world* ».

Depuis quelques années, les derniers développements de l'informatique, qualifiés de manière un peu arbitraire d'« Intelligence Artificielle », semblent couronner une évolution vers un monde où l'homme serait cantonné à l'état de spectateur, voire de protagoniste inutile. Ils ravivent les craintes exprimées par Georges Bernanos en 1947 dans son essai sur *La France contre les robots* ou par Stanley Kubrick (*2001 : L'odyssée de l'espace*) vingt ans plus tard, en 1968.

Témoin de cette angoisse, le physicien Stephen Dawking n'hésitait pas à déclarer dans une interview à la BBC le 2 décembre 2014 que « *le développement de l'Intelligence Artificielle pourrait signifier la fin de l'humanité* ».

Sur un mode plus apocalyptique, voire franchement caricatural, Michel et Monique Pincon-Charlot, deux ex-sociologues proches de la gauche radicale, expliquaient récemment dans *L'Humanité* que « *le dérèglement climatique dont les capitalistes, qui ont pillé les ressources naturelles pour s'enrichir, sont les seuls responsables, constitue leur ultime arme pour éliminer la partie la plus pauvre de l'humanité devenue inutile à l'heure des robots et de l'automatisation généralisée. L'Intelligence Artificielle régnera alors sur une planète au service des riches survivants, une fois que les ouragans, tempêtes, inondations et incendies gigantesques auront fait le sale boulot* ».

Nous assistons en effet à un double mouvement au sein des économies développées :

-En premier lieu un phénomène quantitatif, à savoir un recours de plus en plus important aux robots, d'abord par la poursuite de la mécanisation de l'industrie avec l'implantation des robots industriels (notamment dans les secteurs automobile et électronique) puis le recours croissant à l'informatique dans un nombre croissant de secteurs économiques, des banques au commerce de détail.

-En second lieu, un phénomène plus qualitatif, avec le développement de solutions logicielles de plus en plus perfectionnées et surtout de plus en plus autonomes, au point que beaucoup envisagent qu'elles puissent un jour égaler, voire surpasser l'intelligence humaine. C'est ce que certains ont appelé – de manière un peu prématurée - l'« Intelligence Artificielle ».

Ce double mouvement d'automatisation pourrait-il un jour, comme dans les romans d'anticipation d'Isaac Asimov ou de Kurt Vonnegut, évincer les êtres humains, au point que l'on puisse parler de prise de pouvoir par les robots, et plus particulièrement des robots informatiques, les logiciels, qui, une fois programmés par l'homme, semblent fonctionner en totale autonomie ? D'autant que, depuis quelques années sont apparus des programmes réputés apprendre tout seuls (les anglo-saxons ne parlent-ils pas de « *machine learning* » ?) au risque d'échapper complétement à leurs créateurs, comme dans la légende du Golem.

Inquiet, l'homme du XXI^{e} siècle s'interroge : La machine va-t-elle nous dévorer ou somme-nous capables de la maîtriser comme nous ancêtres ont maîtrisé le feu ? Ou bien l'avenir sera-t-il à l'image des mondes décrits dans les romans d'anticipation ? D'autant que, pour une bonne part, parler d'anticipation est en partie inexact : les robots (mécaniques ou informatiques) sont déjà là, même s'ils n'occupent pas – pas encore ? – toute la place qu'on peut craindre – ou espérer - leur voir prendre dans nos vies.

Pour essayer d'y voir plus clair, il est nécessaire de préciser les termes du débat :

Le discours sur l'Intelligence Artificielle n'est-il pas surfait ? S'agit-il d'une véritable intelligence et est-elle en mesure de surpasser un jour l'intelligence humaine ? La première chose à faire est de bien comprendre de quoi nous parlons, et de faire la part entre ce qui aujourd'hui est avéré et ce qui relève de la pure spéculation. Ceci passe par un détour par la technique, un peu fastidieux mais indispensable. On découvrira à cette occasion que le terme d' « Intelligence » Artificielle n'est pas totalement dénué de réalité mais doit être significativement relativisé afin d'éviter de sombrer dans des spéculations dépourvues de sens.

Dans un second temps nous nous interrogerons sur l'étendue du phénomène : assiste-t-on à une machinisation » du monde ? Allons-nous voir les machines prendre le contrôle de nos vies ? Allons-nous perdre la course face aux machines ainsi que l'affirment certains auteurs comme Martin Ford (*The Rise of the Robots*) ?

Ensuite, nous chercherons à comprendre comme cette évolution peut impacter nos sociétés pour le meilleur ou pour le pire : l'automatisation détruit-elle l'emploi comme on l'en accuse souvent ? Renforce-t-elle les inégalités ou les discriminations ? Menace–t-elle nos libertés ? Quels sont les défis que l'automatisation, et l'Intelligence Artificielle en particulier, posent à nos sociétés ?

Enfin nous tenterons de poser quelques jalons, pour envisager quelles réponses nos sociétés peuvent mettre en place pour répondre aux défis de la technique et en faire bon usage. Face à l'avance prise dans ces domaines par les USA et de la Chine, l'Europe ne peut se contenter de subir et doit apprendre à « chevaucher le tigre ».

II. De quoi parle-t-on ? / Où en est-on ?

A. Quelques concepts

Jusque dans les années soixante-dix, automatisation rime avec mécanisation : si les robots humanoïdes, à l'instar de R2D2 et C3PO, peuplent les romans et films de science-fiction, ceux qui existent dans la vie réelle sont les robots industriels, destinés à la fabrication de produits, comme dans l'industrie automobile, où ils côtoient tant bien que mal les ouvriers sur les chaînes de production. Avec le développement de l'informatique, on entre dans l'ère de la gestion automatisée des tâches. Les machines physiques (les robots) et les « machines » logicielles » dont les derniers développements sont inspirés du raisonnement humain, débouchent sur ce que l'on n'a pas hésité à évoquer sous le nom d' « Intelligence Artificielle » (IA).

Tandis que l'informatique permet dans un premier temps d'augmenter la puissance de calcul, on assiste à une convergence mécanisation-informatique et au développement de la micro-électronique. La limite entre automatisation physique (robots) et automatisation logique va en s'estompant.

Mais surtout, depuis une vingtaine d'année, l'informatique a fusionné avec les télécommunications par le biais d'internet. Jusque-là cantonnée sur les disques durs des ordinateurs, elle est devenue omniprésente. On parle dès lors de « numérisation » de la société, et plus particulièrement du monde économique. La numérisation concerne aujourd'hui tous les flux de l'entreprise (logistique, gestion de production, gestion des clients, échanges avec les tiers).

Le stade ultime de la numérisation est l'Intelligence Artificielle, un concept pas toujours bien compris qui suscite beaucoup d'espoirs et beaucoup de craintes. L'IA va-t-elle permettre de démultiplier les capacités humaines ou bien prendre le pouvoir comme dans le film de Stanley Kubrick ?

Pourtant, le déploiement de machines pour suppléer la force humaine est quasiment aussi vieux que l'humanité. Avec des inflexions majeures : machine à vapeur, électricité, électronique (transistor). C'est justement avec le transistor, ancêtre des puces électroniques, que la machine va commencer à concurrencer non seulement les bras de l'homme mais aussi son cerveau.

Pour cerner les enjeux de l'intelligence artificielle, il est nécessaire de bien comprendre comment elle fonctionne et quels sont les concepts utilisés. L'Intelligence Artificielle n'est pas une mais plurielle. Pour cela, et pour mieux cerner l'enjeu des débats autour de l'IA (ou plutôt des IA), un petit détour par la technique s'avère donc indispensable.

B. Les Intelligences Artificielles

C'est en 1956 que quatre chercheurs américains, qui cherchent à croiser l'informatique naissante et les sciences cognitives, vont donner naissance à la notion d' « Intelligence Artificielle ». L'auteur du concept, Marvin Minsky n'hésitera pas à déclarer à l'époque que les ordinateurs auront dépassé les capacités de l'intelligence humaine dans un horizon de 20 ans.

Les découvertes de la fin des années cinquante (notamment le langage Lisp) déclenchent une vague d'optimisme dont témoigne la déclaration de Simon et Newell, qui affirment comme Minsky que « *les machines seront capables d'ici vingt ans de réaliser n'importe quelle tâche accessible à l'homme* » et que « *d'ici dix ans un ordinateur sera champion du monde d'échecs* ». En fait, les chemins de l'IA seront tout sauf linéaires. Les années soixante seront le premier âge d'or de la modélisation informatique. Mais celle-ci finit par se heurter à de nombreuses difficultés et semble alors dans l'impasse. Les années soixante-dix seront considérées comme le premier hiver de l'Intelligence Artificielle.

C. L'IA symboliste

La dynamique est relancée dans les années quatre-vingt avec la mise au point des « systèmes experts » qui visent à répondre à des problèmes précis à partir d'un enchaînement de règles. Mais ils ne fonctionnent que dans un cadre restreint, en raison de leur logique somme toute assez mécaniste.

La première approche, qui a dominé la recherche dans le domaine depuis le début des années cinquante jusque dans les années quatre-vingt a consisté à répliquer le raisonnement humain pour résoudre des problèmes. On pose un ensemble de règles, comme en mathématiques, et on en déduit une solution. C'est ce qu'on va appeler les systèmes experts. On parle alors d'intelligence artificielle « symbolique ».

Le principe de l'IA symbolique (systèmes experts), tel que théorisé par des chercheurs comme John McCarthy, repose sur un raisonnement déductif qui s'appuie sur la logique reposant sur des faits et des règles connus (problèmes mathématiques, problèmes d'optimisation).

Plus concrètement, un système expert est constitué d'une base de données couplée à un moteur de règles : » pour traiter de problèmes de raisonnement et de planification d'une manière logique et déductive pour répondre à des questions précises. Ainsi un banquier va entrer dans l'ordinateur les données concernant la situation financière de son client, les croiser avec les paramètres reflétant la politique de la banque et en déduire s'il peut accorder un prêt à son client, sur combien de temps et quels sera le montant des mensualités. Il utilise pour cela une suite de calculs plus ou moins complexes, appelée *algorithme*.

Le système expert présente l'avantage d'une très bonne traçabilité (puisque les règles sont explicites). L'IA symbolique repose sur des règles logiques et fait appel à un corpus de connaissances existantes pour prendre des décisions que l'on peut documenter. Qui plus est, certains acteurs comme IBM ont beaucoup œuvré pour la normalisation des moteurs de règles avec l'apparition des outils dits BRMS (*Business Rules Management Systems*).

Cette approche va s'avérer très pertinente dès lors qu'on se situe dans un univers bien balisé. On entre dans le système toutes les règles connues dans un secteur donné, par exemple la production industrielle et le système apporte la « bonne » réponse.

Les systèmes experts sophistiqués utilisent en plus des outils comme les arbres d'exploration (*Random Forest Deep Forest*) et des modèles de connaissance. Cette IA « symbolique fut le courant dominant jusqu'à la fin des années quatre-vingt. Or, une IA symboliste suppose des coûts de maintenance élevés (pour tenir à jour la base de connaissance et l'ensemble des règles) et montre ses limites lorsqu'on utilise des données moins conventionnelles, même si le recours à la logique floue a permis d'utiliser les systèmes experts avec des données non-booléennes (qui ne répondent pas strictement à la distinction vrai-faux).

Ces limites vont provoquer un certain désappointement face à la démarche symbolique. De fait, les années quatre-vingt-dix sont le second « hiver » de l'Intelligence Artificielle. Pourtant les chercheurs ne se découragent pas. Au cours de cette période la recherche continue à progresser : réseaux bayésiens à la croisée des de l'informatique et des statistiques, programmation par contrainte, etc. Mais surtout, on assiste à un renouveau d'intérêt pour les réseaux de neurones.

D. L'apprentissage automatique (machine learning)

Une autre approche va se développer d'abord de manière théorique (initialement au lendemain de la Seconde Guerre mondiale mais surtout dans les années quatre-vingt), puis, plus récemment, de

manière opérationnelle, celle des réseaux de neurones. On parlera alors d'IA « connexionniste » ou « convolutive ».

Dès 1943, Pitts et McCulloch avaient émis l'hypothèse selon laquelle les neurones du cerveau humain ont un fonctionnement binaire (inerte-activé) analogue à celui de l'informatique. Leurs travaux sont repris dans les années cinquante par Rozenblatt qui modélise le premier réseau de neurones artificiels, le Preceptron.

La Machine de Rozenblatt

Dans la machine de Rozenblatt (Preceptron, 1960), un neurone artificiel est un petit circuit (puce électronique) relié par des fils à d'autres neurones. Aujourd'hui dans les réseaux de neurones, il n'y a plus de neurones physiques (circuits), mais des simulations numériques. Le neurone artificiel fusionne les processus de calcul et de mémorisation, contrairement à la logique informatique classique. Le neurone teste l'efficacité des connexions (leur « poids »). Il « apprend » à mémoriser les connexions les plus efficaces. Frank Rozenblatt a perfectionné sa « machine » en organisant plusieurs couches de neurones artificiels, posant les bases de l'apprentissage automatique.

Alors que l'IA symboliste visait à enseigner à la machine des règles mises au point par un cerveau humain, la démarche suivie par l'IA connexionniste tente de reconstituer le fonctionnement du cerveau, qui ne fonctionne pas en appliquant des règles mais en les inventant au fur et à mesure. On peut définir l'IA connexionniste (basée sur des réseaux de neurones) comme un raisonnement inductif qui réalise des prévisions et des généralisations à partir d'observations. Ces méthodes sont à la fois empiriques (on part des données pour en dégager des règles) et probabilistes.

Un neurone artificiel est un objet logiciel (c'est la simulation par informatique des propriétés d'un neurone) qui récupère des valeurs numériques en entrée, associées à un poids, et combine ces valeurs pour générer une valeur de sortie à l'aide d'une fonction.

Les réseaux de neurones artificiels, dont l'architecture logique est inspirée du fonctionnement du cerveau humain, constituent la base des systèmes d'apprentissage automatique (*machine learning*). Celui-ci utilise, non plus des moteurs de règles, mais des outils probabilistes (une propriété est d'autant plus vraie que sa probabilité s'approche de 100%), qui vont traiter les données de manière itérative et inductive. Le *machine learning* va trouver son application dans des domaines difficilement structurables a priori comme la reconnaissance vocale ou la reconnaissance d'image. Ce processus d'apprentissage à partir des données va donner naissance à l'illusion selon laquelle la machine apprend toute seule. Ce qui n'est que partiellement vrai.

Comme le note un rapport de France stratégie (2018), l'apprentissage machine est « *une technologie déterministe et contrôlée, au sens où le programmateur de l'IA choisit l'architecture logicielle qu'il veut utiliser (type de réseaux de neurones, nombre de couches, etc.), la méthode d'apprentissage (algorithme d'initialisation et de mise à jour des poids associés à chaque neurone) et les données d'apprentissage utilisées* ». Ce n'est donc pas une application totalement autonome qui contrôlerait elle-même son développement.

Avec le *machine learning*, on passe d'une logique de programmation (on définit des règles a priori) à une logique d'apprentissage à partir des données. La machine n'a pas besoin d'un jeu d'instruction détaillée. L'algorithme apprend à corriger ses erreurs : il apprend quelle est la décision à adopter (reconnaître une image de chat par exemple) et créé un modèle. Le *machine learning* simple, à base de raisonnements linéaires, est bien adapté pour un nombre pas trop élevé de données et si l'on cherche seulement un résultat mathématique.

Le *machine learning* vise donc à faire émerger des règles à partir de données d'entraînement dans une approche de probabilités et en ayant recours à des corrélations entre données et évènements. Alors que l'IA symboliste donne un résultat certain, dans l'IA connexionniste, on obtient une probabilité : il y a 95% de chances pour que cette image représente un chat. À partir de moment où la marge d'erreur est égale ou inférieure à celle du jugement humain, on peut considérer que l'algorithme a bien fait son travail, mais la machine ne sait pas reconnaître qu'elle s'est trompée. Il y a toujours au départ des données (d'où le risque de biais si les données utilisées ne sont pas pertinentes ou présentent des lacunes). Cette méthode est très dépendante de la qualité des données et de leur exhaustivité. Ainsi, l'algorithme d'Apple Card qui accordait des prêts plus importants aux hommes qu'aux femmes était biaisé car il utilisait des données historiques où les femmes demandant des prêts moins importants.

On distingue deux catégories de machine *learning* :

- l'apprentissage supervisé, construit à partir d'exemples. C'est le plus utilisé. L'apprentissage supervisé repose sur l'entraînement d'un réseau de neurones avec des données qui ont été au préalable « étiquetées » ou « labellisées » (on associe les données d'entrée à une donnée de résultat souhaité). Par exemple on donne à l'algorithme des exemples de chiens en disant que ce sont des chiens. L'algorithme définit lui-même des règles de manière itérative à partir de cas validés (données d'apprentissage). Dans ce cas on dit que l'apprentissage des réseaux de neurones est supervisé (il utilise des labels descriptifs des objets d'une base de référence) et automatique (le poids des neurones est ajusté automatiquement).

-l'apprentissage non supervisé, qui utilise des données non labellisées. Contrairement à l'apprentissage supervisé qui part de règles de labellisation a priori, l'apprentissage non supervisé fait émerger sa propre segmentation, ses propres catégories. On donne à la machine des images sans leur attribuer de catégorie. C'est la machine qui crée les catégories par regroupement à partir des caractéristiques (*patterns*) qu'elle identifie.

À côté des méthodes d'apprentissage classique, supervisé ou non, est donc apparue une troisième méthode dite « renforcée » qui repose sur l'entraînement des données avec un nombre moins important de données d'entraînement, et où on associe les données à un résultat espéré. L'apprentissage par renforcement est fondé sur une approche « essai-erreur ». La machine avance en recevant des « récompenses » chaque fois qu'elle trouve une « bonne » réponse. C'est donc un apprentissage supervisé plus fin.

E. Le deep learning

L'étape suivante est l'apprentissage profond (*deep learning*), une version plus sophistiquée du *machine learning* qui fait appel à un grand nombre de couches de neurones (d'où le qualificatif deep »). Entre la couche d'entrée (dédiée à la perception de l'information) et la couche de sortie (résultats), on trouve plusieurs dizaines de couches dites « cachées », au sein desquelles l'algorithme fait son apprentissage. Dans le *deep learning*, les neurones de réseaux convolutionnels ne sont pas seulement organisés en couches, ils sont structurés par des couches de traitement indépendantes. Comme on ne sait pas bien ce qui s'y passe, on parle de « boites noires ».

Pour réduire le risque d'erreur en début d'apprentissage a été inventée en 2003 la technique dite de « rétropropagation du gradient », qui améliore la performance des calculs mais pas leur manque de transparence. Les modèles de *deep learning* utilisent des méthodes statistiques pour faire des prévisions probabilistes à partir de données d'évènements passés (par exemple : en médecine, ou en prévention de la délinquance). En 2012, Geoffrey Hinton, de l'Université de Toronto, annonce que son équipe est parvenue à faire fonctionner un réseau de neurones multicouches pour la reconnaissance d'images avec un très faible taux d'erreur. Le *deep learning* est né. L'entreprise qu'il vient de créer, DNN research, est d'ailleurs rachetée par Google en 2013. Depuis, grâce aux

réseaux convolutionnels, le taux d'erreurs en reconnaissance faciale est passé en dessous de celui des humains.

Cette révolution est la résultante de plusieurs conditions :
- Une masse considérable de données pertinentes : ainsi la base de données ImageNet comprend plus de 10 millions de visuels étiquetés à la main.
- Un algorithme fort (notamment grâce à l'apparition en 2006 des réseaux DBN qui permettent de gérer l'apprentissage de chaque couche de neurones),
- Une augmentation très sensible de la puissance de calcul des machines et, corrélativement, la baisse du coût de traitement.

Mais cette révolution a aussi ses limites. La pertinence de ces outils suppose un domaine d'action bien circonscrit (on parle ici d' « IA faible ») et un objectif bien défini. On est encore assez loin de l'universalité de l'intelligence humaine.

La mise en œuvre d'un système de *machine learning*, se fait en trois phases utilisant chacune des jeux de données spécifiques : les données d'entraînement et les données de test, que l'algorithme va utiliser pour s'entraîner à reconnaître un chat, par exemple, puis les données de production, sur lesquelles l'algorithme une fois entraîné va travailler. Dans le *machine learning* simple et le *deep learning* supervisé, les données d'entraînement et de test sont labellisées, le plus souvent par des humains.

La première phase est appelée phase d'entraînement. Pour les données d'entraînement, dans la reconnaissance d'image par exemple, on utilise un jeu de 1000 images de chats qui seront « labelisées » (étiquetées) à la main pour en décrire les contenus, via Amazon Mechanical Turk (AMT), filiale d'Amazon, par exemple, ou par un logiciel. Le but est d'obtenir un taux de reconnaissance des images de chat avec une probabilité de l'ordre de 95%, voire plus. C'est pourquoi on parle d'approche « probabiliste ». Plus on a de données labellisées, moins on a besoin d'introduire des règles, et plus l'algorithme est autonome. À l'inverse, les données non labellisées n'ont pas de valeur.

La banque de données ImageNet, lancée aux États-Unis en 2010, comprend aujourd'hui 10 millions d'images. Elle a été dépassée par Google qui a constitué une base de données de 100 millions d'images, lui conférant un avantage décisif pour l'entraînement de ses algorithmes. Google et Facebook ont rendu public beaucoup de leurs algorithmes mais se sont gardés d'en faire autant avec leurs données, ce qui leur donne une longueur d'avance sur la concurrence. L'enjeu, en effet, c'est l'exhaustivité la plus grande des données, pour améliorer la pertinence et éviter les biais statistiques.

On voit qu'il y a bien au départ une intervention humaine pour guider l'algorithme. AMT (ou une autre société du même type) n'est pas l'employeur des « petites mains » qui labellisent les données. C'est une plateforme d'intermédiation, rémunérée à la commission, entre les micro-travailleurs et les donneurs d'ordres. Les micro-travailleurs d'Amazon Mechanical Turk sont en majorité américains. Mais quand le jour se couche en Californie, ce sont les Indiens qui prennent la relève. En 2017, le revenu horaire médian d'un micro-travailleur d'AMT s'élevait à environ 2$. Mais cela ne peut pas dire qu'il peut espérer gagner 16 dollars par jour, tant les séquences de travail sont erratiques. Typiquement le micro-travailleur freelance travaille 28h par semaine dont 6 h à titre gratuit (connexion, réglages, tâches administratives). Il gagnera donc moins de 50 dollars par semaine.

Nous pouvons d'ailleurs tous être des micro-travailleurs à notre insu : Les « captcha » (diminutif de *Completely Automated Public Turing test to tell Computers and Humans Apart*) où l'on demande aux êtres humains de recopier une série de lettres et/ou de chiffres, ou de cocher les cases qui représentent un arbre, une voiture ou un feu rouge, sont des exemples de labellisation.

Les micro-tâches peuvent être très diverses : faire des captures d'écran, annoter une vidéo, corriger des valeurs dans une base de données, relier deux produits dans un catalogue, retranscrire en texte un ticket de caisse, décrire une photo en trois mots, supprimer les doublons et éliminer les erreurs qui pourraient générer des « faux positifs ». Chaque photo mise en ligne sur Flickr, Pinterest ou Instagram va venir entraîner des bases de données, à l'insu de ceux qui l'ont mise en ligne.
Plus la base est importante, mieux le système sera entraîné, mais aussi plus il consommera de puissance machine (d'où le recours au *cloud*) et plus il lui faudra de temps. La base d'entraînement doit être qualifiée, ce qui suppose un gros travail préalable.

Facebook a recours à des bénévoles, mais aussi à des micro-travailleurs pour noter des contenus (textes, photos, vidéos) comme inappropriés ou mensongers, qui vont alimenter des bases de données de *Fake News* pour entraîner les algorithmes à détecter automatiquement les fausses nouvelles.

Le *deep learning* est très exigeant en ressources machine, dans ces phases d'entraînement qui servent à l'optimisation des algorithmes. Il n'y aurait pas eu de réseaux de neurones puissants (*deep learning*) sans la capacité de calcul fournie par les cartes graphiques (GPU), utilisée initialement pour améliorer les jeux vidéo. L'analyse d'image et de vidéo est un des domaines qui ont le plus profité du *Deep Learning*. Une carte graphique GPU contient des milliers de processeurs indépendants qui peuvent effectuer un très grand nombre de traitement parallèles, soit des milliards de transistors contre quelques dizaines pour une carte CPU classique.

Vient ensuite la deuxième étape, dite de test, qui a pour but d'affiner la fiabilité de l'algorithme. Pour les données de test, on envoie par exemple 200 images au hasard et c'est aux étiqueteurs de leur attribuer l'étiquette « chat ». Les données de test également labellisées servent à valider le travail d'entraînement. Il faut qu'il y ait homogénéité des classifications entre données d'entraînement et données de test. Les tests servent à ramener le taux d'erreur en dessous de celui d'un humain.

Les micro-travailleurs interviennent donc également pendant cette phase de validation, qui vise à contrôler les résultats obtenus à l'issue de la phase d'entraînement, rectifier les erreurs et suggérer des solutions pour corriger les « faux négatifs » (résultats positifs non décelés à l'entraînement).

Mais ils ne connaissent pas forcément (contrairement aux experts utilisés dans l'IA symboliste pour construire un système expert) le contexte et sont susceptibles de commettre des erreurs ou d'omettre des points importants Dans le domaine industriel, par exemple, la labellisation repose sur des compétences métiers, ce qui coûte plus cher. C'est pourquoi certains grands groupes comme Microsoft préfèrent avoir recours à des travailleurs plus qualifiés opérant sur une plateforme dédiée appelée UHRS (*Universal Human Relevance System*).

Ce n'est que lors de la troisième étape, dite d'inférence ou de production, que l'algorithme va se mettre au travail de manière relativement autonome. Les données de production, non labellisées, mais pas anonymisées sont celles qui vont faire l'objet du traitement. C'est à ce stade que l'algorithme devient relativement autonome, puisqu'il s'entraîne lui-même.

On voit donc que contrairement à une idée reçue, dans le *machine learning*, la machine n'apprend pas toute seule. Au contraire, la labellisation des données repose dans bien des cas sur la connaissance scientifique des spécialistes humains, comme pour les moteurs de règles.

F. L'histoire n'est pas terminée *:*

Depuis 2012, de nouveaux perfectionnements ont été apportés aux méthodes d'apprentissage automatique. Les plus prometteurs sont les réseaux de neurones génératifs dits « antagonistes »,

apparus en 2014, qui génèrent du contenu, comme la prévision de l'image suivante dans une série vidéo. Les *General Adversarial Networks* (GAN) permettent de générer de nouvelles données quand on a au départ un jeu de données limitées. La machine peut générer la suite d'une séquence d'images. Elle utilise deux réseaux, dits « antagonistes » : le générateur et le discriminateur. Le rôle du générateur est de défier le discriminateur. Dans les GAN donc, deux algorithmes s'affrontent : l'un s'efforce de reconnaître des images de chats, tandis que l'autre lui fournit de fausses images pour tenter de le tromper, et le forcer ainsi à corriger ses erreurs. Ils permettent par exemple de générer *ex nihilo* des images de synthèse, à l'image du « portrait d'Edmond de Bellamy », vendu 400.000 $ chez Christie's, entièrement créé par IA, ou de manière plus inquiétante, la création de fausses photos de célébrités (*deepfakes*).

Notons aussi l'apparition en 2016 des réseaux DNC (*Differential Neural Computers*) mis au point par Deepmind, une filiale britannique de Google, qui permettent de mieux structurer le *deep learning*, avec des outils de raisonnement (symbolistes), qui concourent à une meilleure explicabilité.

On pourra enfin citer les réseaux de neurones ultra-profonds, les réseaux en fractales (développés par une équipe de Microsoft en Chine) et enfin les réseaux en capsule (2017), mis au point par l'équipe de Geoffrey Hinton. Google travaille de son côté sur des réseaux neuronaux complexes, composés de plusieurs sous-réseaux, ce qui permettrait de traiter une quantité encore plus grande de paramètres.

G. L'apprentissage non supervisé, nouvelle frontière de l'IA

Quand les réseaux de neurones parviennent à générer des résultats par eux-mêmes, sans labellisation préalable des données, on parle d'apprentissage non supervisé. Dans l'apprentissage non supervisé, les premières couches ne peuvent pas se guider à l'aide d'exemples. Elles vont donc chercher sans savoir ce qu'elles cherchent, mais vont faire apparaître leur propre classification. La machine « découvre » de nouvelles classifications, alors que dans l'apprentissage supervisé, la classification est faite a priori par la labellisation des données d'entraînement. Au final, c'est le réseau qui va décider quelles sont les formes les plus intéressantes à repérer pour classer les images. Le modèle est donc libre d'évoluer vers n'importe quel résultat final. En pratique, on utilise plutôt ces méthodes en complément de l'apprentissage supervisé.

Un des enjeux majeurs pour l'adoption en masse de l'Intelligence Artificielle est le croisement des approches symboliste et connexionniste (utilisation du *deep learning* pour alimenter des moteurs de règles, intégration de briques de raisonnement dans le *deep learning*). Cela a permis de résoudre certains problèmes spécifiques à chaque approche « pure ». De même, l'apprentissage supervisé peut être rendu plus performant en utilisant des arbres de décision (qui permettent d'extraire des règles logiques de cause à effet à partir de données non structurées), des réseaux bayésiens (pour mixer de règles a priori comme dans les systèmes experts avec les résultats d'apprentissage) ou des machines à vecteur de support.

Les approches symboliste ou connexionniste devraient être dépassées par l'approche dite d'« adaptation contextuelle », qui a pour but de faciliter l'explicabilité (pourquoi la machine donne la réponse A ?) et l'interaction avec les humains. C'est ce que l'agence américaine en charge de l'innovation militaire, la DARPA appelle la « troisième vague de l'Intelligence Artificielle ». Par exemple, l'automate Pandai doit pouvoir converser avec les humains en langage naturel (le langage de tous les jours utilisé par les humains). On entrevoit aussi des perspectives prometteuses dans des domaines aussi variés que la voiture autonome ou l'imagerie endoscopique. Mais, au-delà des aspects techniques, la convergence des deux approches doit permettre de résoudre les problèmes d'explicabilité des algorithmes et la réduction des biais, qui sont actuellement, on le verra par la suite, deux questions fondamentales qui conditionnent l'acceptation ou non des outils d'Intelligence Artificielle dans nos sociétés.

III. Entre réalisations impressionnantes et déconvenues cinglantes

Pour tenter de comprendre les enjeux du déploiement de l'Intelligence Artificielle, penchons-nous sur quelques secteurs : l'industrie, la médecine, le monde de la finance et la prise en compte du vieillissement des populations dans les pays développés. Ceci remet en question l'utopie de la machine omnisciente. L'homme a encore de beaux jours devant lui …

A. Vers l'industrie 4.0 ?

Le concept d'Industrie 4.0 a vu le jour en 2011 en Allemagne, en réaction aux deux phénomènes compris comme des menaces pour l'industrie allemande, la concurrence internationale (Chine) et l'irruption des GAFAM (robots, données clients, Google Car).

Il repose sur trois convictions :

- L'avenir de l'Allemagne passe par l'industrie,
- La priorité est le maintien du leadership mondial des biens d'équipement,
- L'anticipation technologique et son impact sur les process de production sont cruciaux.

Il ne s'agit rien moins que d'organiser la convergence des flux industriels et informationnels qui partent des clients sous forme de commande ferme ou de besoins exprimés vers l'amont des chaînes de valeur industrielles, en s'appuyant sur l'ensemble des technologies IP (internet mobile, *cloud computing*, objets connectés, etc.). Au passage, l'information est restituée – dans l'idéal en temps réel - à chacun des acteurs, qui peut ainsi les analyser dans le cadre d'une démarche d'amélioration continue. Chaque « nœud » met ainsi à jour ses process de conception, d'industrialisation, de maintenance. Dans l'idéal, Industrie 4.0 fait converger les systèmes d'information partagés dans l'entreprise donneuse d'ordre et entre ses sous-traitants. Elle associe robots industriels, capteurs connectés par internet, logiciels de gestion de communication pour gérer les flux et outils d'analyse des données.

1. La robotique industrielle, un engouement inégalement réparti

D'après les statistiques de l'International Federation of Robotics, le parc robotique industriel atteindrait plus de 3 millions de robots installés fin 2020 (dont 60% en Asie), affichant une progression de plus de 10% par an. Aux États-Unis, les dépôts de brevets concernant l'automatisation représentaient en 2018 50% du total des brevets industriels contre 25% en 1975. L'industrie automobile et le secteur électronique représentent à eux deux les deux tiers de la demande mondiale de robots. À l'inverse, le taux d'automatisation de la chaîne logistique, et notamment des entrepôts, est encore très faible (moins de 5% aux USA par exemple).

La Corée du Sud et l'Allemagne ont recours de manière massive aux robots, mais la Chine, dont la force reposait jusqu'à récemment sur la présence d'une main d'œuvre bon marché, semble vouloir rattraper son retard et rejoindre le peloton de tête mondial. En 2016, Foxconn, le fournisseur d'Apple fortement implanté en Chine (il s'agit en fait d'un groupe taïwanais) a annoncé un plan de robotisation massive de ses usines pour remplacer progressivement la majorité de ses 1,2 million d'employés. Foxconn visait à automatiser 30% de sa production à horizon 2020. Dans les faits, ce plan n'a touché que 5% de l'effectif et n'a plus progressé depuis. De même, la robotisation de l'usine Tesla dans le Nevada a été un échec retentissant.

En effet, les robots manquent actuellement de dextérité, c'est-à-dire de capacité à reproduire la palette des comportements corporels des humains, et notamment la préhension. Et sur une chaîne de production, ils sont incapables de détecter des défauts d'assemblage non prévus, comme un boulon vissé légèrement de travers.

La situation pourrait évoluer, car des améliorations ont vu le jour récemment :

Tout d'abord, l'introduction de l'Intelligence Artificielle (apprentissage renforcé) dans la robotique a donné naissance aux robots relationnels, qui interagissent avec leur environnement, et permis d'augmenter sensiblement les facultés de préhension, c'est-à-dire la capacité à assembler des objets d'une manière qui ressemble à celle des humains (assemblage de chaise IKEA réalisé à Singapour en 2018).

Par ailleurs, alors que les robots industriels traditionnels sont de grosses machines situées à l'écart des humains à l'intérieur des usines, on assiste au développement d'une nouvelle génération de machines, les robots collaboratifs, qui viennent compléter le travail des opérateurs dans la manutention d'objets lourds, grâce à l'utilisation de bras robotiques ou d'exosquelettes : Les exosquelettes comme ceux qui sont utilisés dans une quinzaine d'usines Ford permettent de diminuer sensiblement la fréquence (et donc le coût) des accidents du travail et les TMS (Troubles musculosquelettiques). Les « cobots » facilitent les interventions dans les espaces confinés d'accès difficile, que ce soit pour la production ou la maintenance industrielle, mais aussi pour les interventions en environnement difficile (en cas d'accident dans une usine chimique par exemple). Mais c'est encore un marché émergent : selon l'IFR, les « cobots » représentent moins de 4% des ventes mondiales.

2. L'internet des objets et le grand défi de la connectivité

Apparu au tournant des années 2000, l'internet des objets (en anglais : *Internet of Things* ou IOT) consiste à connecter non plus seulement des ordinateurs entre eux, mais à munir des étiquettes ou des capteurs de puces informatiques simples (les puces RFID par exemple) pour leur permettre d'envoyer des informations à un serveur via un réseau internet à bas débit. Les projets de type Industrie 4.0 (appelée Industrie du Futur en France) reposent justement sur un flux continu de données d'un bout à l'autre de la chaîne de valeur, données qui peuvent concerner aussi bien des quantités de pièces en stock que des mesures physiques concernant un équipement industriel. Pour l'internet des objets, l'interopérabilité des systèmes de communication à bas débit entre les différents acteurs et les différents systèmes (Lora, Sigfox, 2G) est un enjeu majeur. L'introduction de la 5G devrait améliorer la circulation des données générées par les capteurs industriels pour les applications critiques qui demandent une information en temps réel (et donc une faible latence), comme la voiture autonome.

Certains problèmes n'ont pas toujours trouvé de solution parfaitement satisfaisante à ce jour :
D'une part l'internet des objets entraîne un développement de l'économie de plateforme, pilotée par les grands donneurs d'ordre, avec un risque de domination de la chaîne de valeur par un acteur qui centralise les données. Peut-on au contraire envisager un partage équitable des informations entre donneurs d'ordres et sous-traitants ? Et comment s'opère dès lors la propriété des données partagées ?
L'autre problème est l'ampleur des risques de piratage des données, mais aussi la possible compromission des algorithmes de calcul par des tiers malveillants. Le piratage de la centrale nucléaire de Natanz en Iran en 2010 montre qu'on aurait tort de sous-estimer ces risques. D'où l'intérêt croissant pour les systèmes dit d' *« Edge AI* » où les données sont collectées et traitées en local, sans besoin de connexion internet afin d'assurer la sécurité des IOT industriels.

3. Réalités et mirages de la data driven corporation

Après la robotique industrielle et la connectivité, l'Intelligence Artificielle constitue le troisième pilier de l'industrie du futur, puisqu'il faut bien analyser toutes les données collectées. Mais le paradoxe est que ces données sont souvent très spécifiques à des métiers. Ce qui rend parfois difficile pour les acteurs européens d'atteindre la masse critique en termes de données d'entraînement. Il n'en va pas de même aux États-Unis et en Chine, où quantités de données sont aux mains d'un nombre limité d'acteurs, publics ou privés, ce qui permet d'atteindre la masse

critique. D'où l'émergence de propositions de mutualisation des données entre entreprises européennes. Faute de quoi la « *data driven corporation* » risque de n'être qu'un mythe.

En maintenance prédictive, par exemple, les entreprises disposent rarement de séries de données sur les pannes passées. C'est pourquoi une entreprise comme Renault a mis au point dès 2016 un système de collecte de toutes les données disponibles sur 22 de ses sites (données qu'il a fallu au passage homogénéiser pour pouvoir les traiter efficacement) avant de se lancer dans un grand projet d'intelligence artificielle avec Google en 2020. D'où l'intérêt pour les Européens de rechercher la mutualisation. En fait, ces démarches de mutualisation ne fonctionnent qu'à l'échelle nationale, comme en témoigne le consortium piloté par EDF, Total et Thales dédié à l'optimisation des systèmes critiques.

L'alternative, dès lorsqu'on utilise des données très spécialisés (« *small data* »), c'est le « *Few-shot learning* » (algorithme s'entraînant avec peu de données), Ainsi, en maintenance prédictive, on utilise plutôt des moteurs de règles spécifiques au métier, car les outils de *machine learning* génériques génèrent beaucoup de faux positifs. C'est d'ailleurs un risque inhérent aux techniques d'apprentissage automatique, qu'on retrouve dans d'autres domaines – on le verra un peu plus loin - avec les systèmes dits de « police prédictive » par exemple.

4. Et la France ?

La France compte 150 robots pour 100.000 salariés contre 330 en Allemagne et plus de 800 en Corée. L'écart de robotisation avec l'Allemagne s'expliquerait aux deux-tiers par la faible taille des entreprises françaises (pas de *Mittelstand*, donc moins de capacité d'investissement et moins de volume pour amortir les machines) et le reste par la part moindre de l'industrie. Mais notre pays est aussi en retard sur la numérisation de la chaîne logistique (beaucoup de commandes manuelles avec émission de papier plutôt qu'avec des chaînes de commande informatisées de type EDI). Il est vrai que la législation comptable ne favorise pas forcément la dématérialisation pour les PME.

Alors que la sidérurgie avait été pionnière dans l'utilisation des systèmes experts dès le début des années 80, avec le système SACHEM dédié à la gestion des hauts-fourneaux, l'Industrie du Futur est à peine mentionnée dans le rapport Villani. L'industrie est clairement l'angle mort de la stratégie française en matière d'intelligence artificielle. Pourtant, le *smart manufacturing* constitue une réponse à la stagnation des gains de productivité. Il se heurte aussi à des rigidités organisationnelles au sein des entreprises, alors qu'il est dépendant de l'alignement des process (SMED, capacité à produire en petites séries) sur les informations captées et transmises par les objets connectés.

B. Les applications médicales : Watson, robots médicaux et diagnostic automatique

1. Du chirurgien augmenté …

Les robots médicaux ont fait leur apparition dans les années 80. Ce sont principalement des robots interventionnels comme Da Vinci ou Cyberknife. Commandés par des médecins dans le cadre d'opérations de téléchirurgie, ils ne sont pas autonomes. Ils peuvent s'avérer précieux pour les opérations dites 'mini-invasives » qui nécessitent une grande précision. En revanche, ils ont parfois du mal à gérer la résistance tissulaire et peuvent donc causer des lésions s'ils ne sont pas maniés avec la plus grande précision. Intuitive Surgical (robots Da Vinci et Zeus) revendique plus de cinq millions d'interventions à ce jour. Récemment, le robot STAR a réussi à suturer un intestin grêle de porc avec des points plus homogènes qu'un chirurgien humain. La diffusion de ces semi-automates se heurte néanmoins à leur coût élevé : un grand hôpital métropolitain ne peut se payer plus de deux ou trois exemplaires.

2. …au diagnostic automatisé

Le premier système expert médical, Mycin (1974), était dédié au diagnostic et au traitement des infections sanguines. Il a connu un succès mitigé en raison de difficultés à maintenir la base de connaissances et de la rigidité de son moteur de règles.

C'est pourtant de manière assez logique que la deuxième vague d'Intelligence Artificielle médicale, fondée sur la reconnaissance de caractéristiques pathologiques, a permis des avancées significatives dans le domaine du diagnostic médical. Ainsi, en s'appuyant sur une base de données de 10.000 clichés de lésions dermatologiques, une équipe de chercheurs français, allemands et américains a récemment réussi à l'aide d'un système expert (associé à des outils d'apprentissage automatique pour la reconnaissance d'images), à diagnostiquer des cancers de la peau avec une précision de 95%, alors que, sur cet échantillon, le degré de pertinence des dermatologues travaillant sans assistance informatique était de l'ordre de 90%. Google de son côté a annoncé début 2020 avoir développé un système d'IA pour détecter les cancers du sein avec plus de précision que les radiologues, c'est-à-dire présentant moins de faux positifs ou négatifs qu'avec les humains. En France, Cardiologs propose la détection par algorithme des pathologies cardiaques (à partir de l'interprétation d'électrocardiogrammes). Et l'Institut Gustave Roussy a mis au point à partir de 65 variables un outil de simulation de l'évolution de la maladie sur des patients atteints du Covid 19, en entraînant l'algorithme sur un millier de patients.

Le diagnostic pourrait ne pas être le seul domaine prometteur : le MIT prétend avoir découvert un nouvel antibiotique grâce à l'apprentissage automatique : en fait l'IA a permis de démontrer que l'halicine, utilisée dans le traitement du diabète, pouvait s'avérer très efficace contre le bacille e-colli, sans que le bacille devienne résistant, confirmant les thèses du professeur Raoult sur la réutilisation des molécules.

Comme on l'a vu, le déploiement des outils d'analyse et de modélisation informatiques repose sur un grand nombre de données. Celles-ci sont utiles pour l'entraînement des outils de diagnostic, mais aussi en matière de choix de santé publique. Les entrepôts de données clinique, comme le *Health Data Hub* lancé en France en octobre 2018, doivent permettre une optimisation des décisions en matière de besoins de santé. Il doit être progressivement complété par les données issues du Dossier Médical Partagé. Celui-ci, qui a connu bien des vicissitudes, est désormais créé à l'initiative des patients. Il contient des documents et non des données, d'où une grande difficulté à faire des traitements statistiques.

En médecine comme dans d'autres domaines, l'outil logiciel permet de compléter les compétences des professionnels. C'est l'intelligence augmentée, c'est-à-dire l'analyse des résultats issus de l'IA par des humains experts dans leur domaine, qui donne des résultats probants. Ainsi, pour le diagnostic du cancer, les outils d'apprentissage automatique utilisés seuls ont un taux d'erreur de 7%, alors que si on associe des cardiologues à l'analyse, le taux d'erreur tombe à 0,5%.

Le recours à la machine est comme la langue d'Esope. Il contient le pire et le meilleur, comme le montre l'exemple suivant : en Chine, un algorithme vise à évaluer les chances de rémission de malades qui sont dans le coma. Mal appliqué, il peut conduire à un tri arbitraire et à l'abandon de soins. Les promoteurs du logiciel affirment qu'en fait lorsque les médecins ont émis un avis négatif, l'algorithme permet de découvrir ceux qui auraient quand même une chance de s'en sortir, alors qu'ils étaient condamnés par la médecine, et qu'ils peuvent ainsi sauver des vies. L'étude publiée sur *The Lancet* en mai 2020 à propos de l'usage de l'hydrochloroquine illustre ce qui se passe quand on utilise des données mal qualifiées. Si la labellisation d'images est une tâche basique, labelliser des images médicales requiert beaucoup d'expertise.

Enfin, la focalisation sur les symptômes, mise à jour à parti d'analyses de corrélations statistiques ou visuelles, risque d'occulter l'enjeu majeur de la médecine qu'est, depuis Claude Bernard, la recherche des causes, qui seule peut permettre la mise en place de politique de prévention.

Watson à la conquête de l'intelligence artificielle médicale

Depuis 2011, IBM a investi plus de 5 milliards de dollars dans sa division Intelligence Artificielle, connue sous le nom de Watson. Watson s'est fait connaître à cette époque par sa victoire au jeu télévisé *Jeopardy !* où il avait fait preuve de ses capacités impressionnantes à comprendre le langage naturel, celui que nous utilisons tous les jours, et pas simplement à faire un tri dans des mots-clés.

Après un temps d'hésitation, IBM a décidé que c'était dans le domaine médical, où abondent les données non structurées, que Watson pourrait faire ses preuves sur le plan commercial. Six ans plus tard, les résultats sont plus que mitigés :

-Le partenariat avec le Centre Anderson de traitement contre le cancer de l'Université du Texas, qui visait à émettre des recommandations thérapeutiques à partir de l'analyse automatique des dossiers des patients, s'est avéré un échec doublé d'un gouffre financier pour le client, qui y a mis fin au bout de deux ans.

-Le programme de coaching médical lancé en partenariat avec l'équipementier sportif Under Armour n'a jamais vu le jour, de même que le programme d'accompagnement des patients atteints de diabète associant IBM et Medtronic, qui avait pourtant reçu l'aval de la FDA en 2018.

L'argument à la base de discours d'IBM sur la capacité de la machine à stocker et traiter une masse d'informations qu'aucun médecin n'était en mesure d'absorber, et à faire le lien, grâce aux outils d'analyse du langage naturel, avec les dossiers des patients, s'est heurté au mur des réalités. Le domaine auquel Watson s'est attaqué était bien plus compliqué que prévu. À ce jour, les seules réussites probantes de Watson se situent dans le domaine du diagnostic, et notamment dans l'analyse génomique, où les données sont au contraire bien structurées.

C. Finance : la prudence s'impose

Alors que la crise de 2008 a montré le peu de confiance que l'on peut faire dans les modèles mathématiques informatisés, le secteur bancaire est l'un de ceux qui ont adopté l'Intelligence Artificielle avec le plus d'enthousiasme.

L'automatisation du trading montre que des emplois même très bien payés peuvent être automatisés. Dans les salles de marché, le trading algorithmique représente aujourd'hui plus de 90% des ordres boursiers. Ce qui ne va pas sans quelques déboires, comme l'illustre le « Flash Crash » du 6 mai 2010 : les algorithmes ont amplifié l'erreur d'un trader au point qu'en une demi-journée, le Dow Jones a perdu 9% en l'espace de 10 minutes. De son côté, le fonds indiciel *AI Powered Equity ETF* côté au NYSE, qui utilise Watson d'IBM pour guider ses choix d'investissement, n'a pas fait de merveilles Pour le *Financial Stability Board*, organisme de coordination financière créé par le G20 après la crise de 2008, l'évolution de l'IA peut accélérer les risques systémiques dans le monde bancaire.

IBM a conclu des accords avec certaines banques françaises pour automatiser les relations avec les clients ou le choix des produits financiers à leur proposer avec Watson, comme Orange Bank (pour affiner ses offres de crédit à la consommation) ou le Crédit Mutuel (pour aider les chargés de clientèle à s'orienter dans la documentation technique de la banque). La Société Générale a mené des expériences en banque privée avec Watson, mais l'essai a tourné court. Au Royaume-Uni, l'assureur en ligne Admiral avait établi un score pour les jeunes conducteurs en fonction du contenu de leurs posts sur Facebook. L'autorité de régulation bancaire française, l'ACPR, a pointé le risque que le recours au cloud pour l'IA bancaire multiplie les points d'entrée pour les criminels.

Mais surtout, de manière certes moins visible, les grands établissements bancaires ont déployé des outils de RPA (Robotic Process Automation) dans leurs back-offices : à l'aide de l'apprentissage automatique, le logiciel « apprend » les processus de traitement pour les répliquer sans aide humaine). Les effectifs dans le domaine de la banque et de l'assurance pourraient ne pas baisser

autant que certains le prédisent, mais l'évolution vers la massification des traitements et de la relation commerciale et donc leur automatisation - ne devrait pas ralentir dans un secteur où la rentabilité est de plus en plus contrainte par la règlementation et la persistance de taux d'intérêt très bas.

L'automatisation comme remède au vieillissement ?

Dans un rapport de mars 2018 intitulé *Intelligence et travail,* France Stratégie constate qu'*« il est difficile aujourd'hui de savoir ce qu'il adviendra d'ici quinze ou vingt ans, tant du point de vue des avancées technologiques que de leur diffusion et de leur appropriation, notamment dans le monde du travail* ».

Le rapport a choisi de se concentrer sur trois secteurs : la santé (imagerie médicale, ingénierie du génome), la banque de détail (maintenance prédictive dans les usines, montée en puissance des fintech) et les transports (voiture autonome).

Dans le secteur des transports, on pourrait assister à « une « tri polarisation » des emplois, entre ceux qui progressent vers les postes de supervision, ceux qui sont réorientés vers l'accueil voyageur et ceux qui disparaissent du fait de l'apparition de véhicules sans chauffeurs » (notamment dans les transports routiers).

Dans la banques, seraient menacés les métiers de back office, ceux de l'évaluation des risques et une partie de l'accueil client. Les banques de détail ont réduit leurs nombre d'agences, pas tant du fait de l'automatisation, mais surtout en raison de la baisse des taux qui a stoppé la course aux dépôts. Dans les métiers de marché de capitaux, la majorité des transactions boursières sont maintenant réalisées par des algorithmes de *trading.* Les banques sont également contraintes par la réglementation (Bâle III) qui limite la recherche de rentabilité fondée sur les opérations plus risquées. Ceci plaide en faveur d'un scénario de rupture, où à tout le moins d'une inflexion forte.

On notera qu'en France, la population dans ces deux secteurs est vieillissante. Or Aghion (2018) a montré que ce sont les pays à population vieillissante (Japon, Corée, Allemagne) qui ont le plus robotisé leur appareil productif.

Dans la santé, l'aide au diagnostic automatisé a fortement progressé, en cardiologie par exemple. Ceci intervient dans un contexte de démographie médicale déclinante, tandis que le déploiement de l'e-santé reste balbutiant dans notre pays. On notera que vingt ans après l'introduction de la carte vitale, nombre de médecins continuent à recourir à des feuilles de soin manuelles …

On notera que les pays fortement robotisés (Japon, Corée du Sud, Allemagne) à la population vieillissante présentent un faible taux de chômage. Comme l'a montré Jean-Marie Bouissou (*Les leçons du Japon*, 2019), il existe au Japon une foule d'emplois souvent occupés par des seniors, qui n'existent peu ou pas chez nous. Ce sont des emplois faiblement qualifiés qui servent surtout à produire du « liant » dans la société.

D. Vers la domination de la machine ? Le mythe de l'Intelligence Artificielle Générale

Les perspectives ouvertes par les réseaux de neurones et l'apprentissage automatique ont relancé le débat sur la possibilité pour les machines de raisonner sans intervention humaine. Ainsi, Geoffrey Hinton affirmait par exemple dès 2016 qu'il était inutile de former des radiologues car l'Intelligence Artificielle ferait mieux qu'eux leur travail. En effet, dès qu'il y a de gros volumes de données à traiter, la machine l'emporte sur l'intelligence humaine, qui est freinée par sa capacité de mémoire. Si l'Intelligence Artificielle a pratiquement rattrapé l'homme ou l'a parfois dépassé dans le domaine du calcul spécialisé, rien n'empêche alors de penser que cette supériorité va s'exercer peu à peu dans tous les domaines. Partout, l'ordinateur ferait mieux que l'homme qui n'aurait d'autre choix que de s'effacer au profit de la machine, pour son plus grand bien.

1. La thèse de la Singularité

Dès les années cinquante, des experts aussi avertis d'Herbert Simon ou Marvin Minsky pensent que les ordinateurs parviendront à concurrencer l'intelligence humaine à horizon de vingt ans. Cette thèse est reprise par John Irving Good au cours de la décennie suivante. Il affirme que si l'Intelligence Artificielle parvient à dépasser l'intelligence humaine, les logiciels seront donc naturellement supérieurs à l'humain dans la conception même de logiciels. Autrement dit, l'Intelligence Artificielle s'engendrera elle-même. C'est exactement ce qu'affirme Stephen Dawking dans l'interview à la BBC citée dans notre introduction : « *Une fois que les humains auront développé une Intelligence Artificielle, celle-ci pourrait s'amplifier par elle-même et se reconfigurer à un rythme grandissant sans cesse* ».

C'est avec Raymond Kurzweil, directeur technique de Google, que ces thèses vont connaître leur apogée. Empruntant à un auteur de science-fiction, Vernor Vinge, le thème de la « Singularité », Kurzweil explique qu'avec la loi de Moore, qui prévoit une accélération infinie de la puissance de calcul des ordinateurs, la course de vitesse entre l'homme et la machine ne peut que se terminer par la défaite du premier. La Singularité ou Intelligente Artificielle Générale, qui est l'état où la machine a définitivement détrôné l'intelligence humaine, pourrait se produire selon lui à l'horizon 2045.

Il n'est pas neutre de remarquer que c'est en Asie – on verra pourquoi par la suite - que les thèses de Kurzweil ont rencontré le plus de succès.

2. Des performances mitigées

Comme on l'a vu précédemment, en dépit de réussites indéniables, l'Intelligence Artificielle a clairement des limites. Comme le remarque Yann Le Cun, pourtant l'une des figures de l'optimisme technologique, l'apprentissage automatique se heurte aujourd'hui à des problèmes de mémoire, qui seront peut-être résolus par l'apparition de nouveaux méga-ordinateurs surpuissants (les calculateurs hexaflopiques) et par les progrès de l'informatique quantique. Mais il y a plus.

Les systèmes les plus performants d'analyse du langage naturel revendiquent une pertinence d'environ 90%, contre plus de 95% chez les humains (test de Winograd). En réalité, elle dépasse rarement 80%, soit bien en deçà des capacités humaines. Nous constatons tous les jours que les agents conversationnels (*chatbots*) sont très limités (ce sont en fait des systèmes experts assez simples) et qu'ils n'ont pas de mémoire, donc pas de capacité d'apprentissage. Quant aux configurateurs, qu'on trouve notamment sur Netflix ou Amazon pour orienter nos choix d'achats, ils ont faits de gros progrès, mais ont encore du chemin à faire. Un observateur facétieux notait qu'ayant commandé une lunette de toilette sur Amazon, l'algorithme de recommandation du site lui en proposait de nouveaux modèles à chaque nouvelle connexion…Qui plus est, l'Intelligence

Artificielle est très spécialisée : en médecine, par exemple, un outil d'apprentissage automatique dédié à une pathologie sera totalement inutile pour une autre, parce que les données d'entraînement sont différentes, et parce que les règles devront être totalement reprogrammées. Les mésaventures d'IBM dans ce domaine s'expliquent en partie par cela.

3. L'Intelligence Artificielle ne « réfléchit » pas

Dans son roman *8200*, le journaliste franco-israélien Dov Alfon – devenu depuis rédacteur en chef de *Libération* - met en scène des agents chinois qui espionnent les conversations d'un policier israélien devant l'ambassade de France. Leur logiciel de traduction automatique hébreu-chinois leur apprend que celui-ci se cache à Créteil-sur-mer. Tout citoyen raisonnablement instruit sait pourtant que Créteil est fort éloigné des côtes. En fait, le policier se cache chez sa mère. Mais le logiciel ne « comprend » pas ce qu'il traduit.
En fait les algorithmes sont, pour reprendre l'expression de Jean-Louis Dessalles des « idiots savants ». Watson a gagné au jeu *Jeopardy* en compilant une masse colossale de données sur internet, mais il est incapable de « comprendre » les questions. Il peut repérer des occurrences dans un texte sans rien comprendre au sens du texte. Dès lors, conclut Dessalles, on peut dire que « *les Intelligence Artificielles actuelles sont très intelligentes et très stupides à la fois. Elles sont stupides car lorsqu'elles prennent une décision intelligentes, elles n'ont aucun moyen de le savoir* ».

4. L'Intelligence Artificielle ne réplique pas vraiment le cerveau humain

En 1943, le neuropsychiatre Warren McCulloch et le mathématicien Walter Pitts ont proposé une modélisation du cerveau humain qui a directement inspiré les réseaux de neurones. Mais ceux-ci ne sont qu'une réplication très simplifiée de notre cerveau, dont l'arborescence et la complexité sont loin d'avoir été véritablement reconstituées. Nous n'avons d'ailleurs pas encore réussi à ce jour à formaliser une théorie générale de l'intelligence.

Des chercheurs de l'École Polytechnique Fédérale de Lausanne sont parvenus en 2015, dans le cadre du projet européen BlueBrain, à simuler le comportement d'une partie du cerveau du rat. Mais pour simuler celui du cerveau humain, il faudrait une puissance de calcul presque dix fois supérieure à celle des plus gros ordinateurs existants, si tant est qu'on parvienne à répondre à certaines questions sur la biologie du cerveau encore en suspens.

Le cerveau humain a en outre une capacité de traitement massivement parallèle, c'est-à-dire de traitement simultané d'un grand nombre d'informations (jusqu'à 60 processus en même temps), capacité qui égale les plus grosses « fermes de calcul » informatiques. Le plus puissant ordinateur neuromorphique du monde, le Spinnaker, mis au point par l'Université de Manchester, ne parvient à simuler qu'un milliard de neurones, soit cent fois moins que le cerveau humain. Contrairement aux réseaux de neurones qui ont besoin d'une importante quantité de données pour s'entraîner, le cerveau humain parvient à bâtir des raisonnements à partir d'un nombre limité de données. Il a donc une capacité dont l'Intelligence Artificielle est dépourvue.

Il est peut-être après tout illusoire de penser que l'Intelligence Artificielle doive s'inspirer de l'intelligence humaine, même si c'est la piste qui est suivie aujourd'hui. L'automobile va plus vite que le cheval, mais elle ne fonctionne pas de la même manière. Les réseaux de neurones ne sont donc pas forcément la forme ultime de l'Intelligence Artificielle (il est donc vain de croire que c'est là que se trouve le chemin vers la Singularité). Certains parlent déjà d'un hiver de l'IA connexionniste…
« *La question de la Singularité technologique,* concluait avec humour Jean-Louis Dessalles, *ressemble à celle de la surpopulation sur Mars : on ne peut exclure que le problème se pose un jour, mais ce n'est pas pour demain* »

IV. Un parcours encore semé d'obstacles

A. L'emploi et l'inégalité

1. Un monde sans emploi ?

En septembre 2013, l'année qui suit l'annonce des découvertes de Geoffrey Hinton sur les réseaux de neurones, deux professeurs d'économie de l'Université d'Oxford, Carl-Benedikt Frey et Michael Osborne, publient un article de 70 pages qui connaît un grand retentissement. À partir d'une grille de 903 métiers-types, ils concluent que 47% des emplois aux États-Unis risquent de disparaître du fait de l'automatisation, à l'horizon d'une vingtaine d'année. L'année suivante, un cabinet de conseil européen (Roland Berger) a renchéri en estimant ce risque à 42% pour la France.
L'analyse de Frey et Osborne est la suivante : à partir du moment où plus de 70% des tâches qui correspondent à un métier sont automatisables, celui-ci est en risque. Les taches routinières, qu'elles soient effectuées par des cols bleus ou des cols blancs, sont automatisables. Les professions considérées comme « nobles » comme les avocats, seraient, elles aussi, menacées, car elles comportent en réalité beaucoup de tâches automatisables (comme la recherche de textes juridiques).

Le choc des robots
Restrepo et Acemoglu (2017, 2018) ont documenté le « choc des robots » à partir du milieu des années 90, qui a touché principalement les hommes travaillant dans l'industrie. Aux États-Unis, pour chaque nouveau robot industriel, sur la période 1990-2007, 6 emplois locaux ont disparu, directement dans l'industrie, mais aussi dans l'environnement local : construction, distribution (secteurs intensifs en main d'œuvre). Les robots industriels auraient ainsi éliminé environ 500.000 emplois dans l'industrie sur 20 ans.
Pour Chiacchio, dans son étude de 6 pays européens sur la période 1995-2007, chaque robot industriel a détruit 3 emplois.
Même en Allemagne, Dauth (2018) considère que la robotisation a entraîné une baisse des emplois industriels compensée par les services. Sur la période 1994-2014, chaque robot aurait entraîné en Allemagne la perte de 2 emplois industriels, compensés par 2 emplois, souvent moins rémunérés, créés dans le secteur des services.
Il y avait environ 2.8 millions de robots en service dans le monde fin 2019. Ce chiffre pourrait passer à 20 millions en 2030, dont les deux tiers en Chine. Dans ce cas, les robots industriels pourraient déplacer le même nombre d'emplois manufacturiers dans le monde à cette échéance.

Un sondage de l'institut Pew Research de 2018 mené dans les principaux pays développés montre qu'en moyenne 75% des personnes interrogées pensent que la robotisation et l'automatisation viendront détruire une grande partie des tâches actuellement accomplies par les humains. Et ce dans un contexte où, jusqu'à la crise du coronavirus, le niveau de chômage avait fortement baissé dans des pays comme les États-Unis, le Royaume-Uni ou l'Allemagne. Pour ne rien arranger, selon un rapport pour le World Economic Forum de Davos de 2019, plus d'un milliard de personnes seraient menacées par l'automatisation, dont seulement 25% pourraient faire l'objet de programmes de requalification.

À l'inverse, Frey et Osborne distinguent trois 3 éléments de résistance à l'automatisation : la perception et l'évolution au sein d'environnement non structurés et à structure complexe (atelier de garagiste, plombier), les tâches demandant de la créativité et les tâches reposant sur l'intelligence sociale (empathie, vente).

Pourtant, Levy et Murnane (2003) considéraient l'approche par métier trop globale pour être pertinente : il faut selon eux raisonner en termes de tâches concrètes et non en termes de

qualification globale des personnes. Un emploi est une somme de tâches : l'analyse par les tâches permet de mieux comprendre les besoins d'évolution des compétences. Des personnes qui ont le même niveau d'éducation peuvent en réalité accomplir des tâches très différentes. L'analyse par les tâches permet d'obtenir un meilleur degré de granularité. C'est cette approche qui est retenue par l'OCDE (2016) qui considère que seuls 9% des emplois seraient réellement automatisables. En France, cela signifierait, toutes choses égales par ailleurs, un doublement des demandeurs d'emploi de catégorie A.

On peut trouver une autre raison de dresser un tableau nuancé des relations entre automation et emploi dans un article récent de trois chercheurs de la Brookings Institution (Muro, Maxim, Whiton, 2019) : Certes l'automatisation se substitue dans un certain nombre de cas au travail (détruit des tâches), mais la machine ne peut pas tout faire. Les machines se substituent aux tâches et non aux emplois. Certaines tâches sont mieux faites par les machines, d'autres par les humains. Dans les autres cas, l'automatisation complète le travail humain. Cela rend le travail humain résiduel plus important, entraînant des gains de productivité.

Par ailleurs, selon nos trois auteurs, l'automatisation est créatrice de richesse : la redistribution des gains de productivité (notamment par la baisse des prix) génère une hausse de la demande et donc des emplois (sauf si recours aux importations) et la combinaison capital-travail induit l'innovation (les guichets automatiques n'ont pas réduit l'emploi bancaire).

Enfin, la faisabilité technologique n'induit pas forcément un déploiement destructeur d'emplois. L''implantation de bornes dans les *fast foods* a permis de mieux traiter l'afflux de consommateurs en heures de pointe mais n'a que peu réduit le nombre de salariés. Et ce n'est qu'avec la baisse des taux que l'emploi bancaire a été substantiellement réduit.

Pour les auteurs, si l'on regarde les trente dernières années, le bilan de l'automatisation a été légèrement positif en termes d'emploi : les gains ont été concentrés sur les deux extrêmes de la pyramide sociale, au détriment de la classe moyenne. Du moins jusqu'à présent. À l'avenir, il n'est pas d'emploi qui ne sera affecté par l'automatisation. Environ 25% des emplois aux USA seraient fortement touchés (emplois comportant plus de 70% des tâches présentant un risque de substitution). Ce seront les tâches prédictibles qui seront affectées : administration, production, transport, les premiers touchés occupant les emplois relativement peu qualifiés dans les zones industrielles traditionnelles, les zones rurales et les petites villes. À l'inverse les 40% situés aux deux extrêmes de la pyramide sociale ne seraient pratiquement pas touchés.

Dans cette configuration, les hommes (présents dans les transports, la production, la construction et l'installation) sont plus vulnérables que les femmes (services à la personne, éducation). L'automatisation physique est préjudiciable aux métiers qui demandent de la force. C'est pourquoi la robotisation de l'industrie, comme l'a noté Ameçoglu, s'est faite au détriment de l'emploi masculin. La seconde vague pourrait toucher l'emploi féminin (métiers de service peu qualifiés à l'exception des services à la personne).

Les études empiriques tendent à montrer que, sur longue période, la mécanisation – et aujourd'hui l'automatisation -, est plutôt créatrice d'emploi. À court terme, elle a deux effets (Autor, 2018) :

- un effet de substitution, avec dans un premier temps une destruction d'emplois
- un effet de déplacement, avec la création d'emplois dans de nouveaux secteurs, une hausse de l'emploi dans les entreprises innovantes au détriment de leurs concurrentes, une baisse de l'emploi dans les catégories intermédiaires au profit des extrémités de la pyramide sociale.

Dans un long rapport publié en 2017, le Conseil d'orientation pour l'emploi cite Gergory, Salomon et Zierahn (2016) qui ont tenté de mesurer sur une période de sept ans (1999-2016) l'impact de l'automatisation dans l'Union européenne. Leurs conclusions sont plutôt positives sur l'emploi. De

même, entre 1980 et 2007, l'emploi a progressé de plus de 17% aux USA et la moitié de la croissance s'explique par l'émergence de nouveaux emplois.
L'automatisation des tâches routinières aurait un impact positif sur l'emploi, que ce soit dans les entreprises concernées, les plus productives chassant leurs concurrents (mais pas forcément au niveau du secteur, comme en témoigne la baisse des effectifs dans l'agriculture puis dans l'industrie) ou par la croissance de nouveaux secteurs. Les effets de compensation seraient deux fois supérieurs aux effets de substitution.

L'innovation crée en effet des emplois de différentes manières : construction de nouvelles machines (matérielle, mais aussi logicielles), réduction des coûts de production qui permet l'augmentation des ventes (en fonction de l'élasticité de la demande), amélioration des marges qui renforce la capacité d'investissement et enfin création de nouveaux produits et services (innovation produit).

On néglige d'ailleurs souvent l'apport de l'automatisation dans l'innovation de produit, qui est créatrice d'emploi. L'étude de Harrisson (2014) sur les entreprises européennes, et notamment françaises, tend à montrer un effet positif de l'innovation de procédé sur l'emploi, mais pas de manière très significative. En revanche, l'innovation de produit serait très positive pour l'emploi. L'innovation de procédé crée de l'emploi dans les entreprises qui y ont recours, mais en détruit chez leurs concurrents. Du coup l'impact sectoriel est limité.

2. Un peu d'histoire

Née de la machine à vapeur, alimentée au charbon et de la « mule » à tisser le coton, la révolution industrielle avait permis à des travailleurs peu qualifiés, souvent des paysans chassés de la campagne par le mouvement des *enclosures*, de pratiquer des tâches autrefois réservées à des artisans qualifiés, d'abord avec des salaires très bas. Ils constituent le premier prolétariat industriel, dont les conditions de vie sont décrites par Engels dans *La situation de la classe ouvrière en Angleterre en 1844*.

La seconde révolution industrielle (à partir de 1870) a vu le train remplacer la diligence, le moteur électrique la vapeur, et les grues les manutentionnaires. Elle a créé des métiers d'ingénieurs, de mécaniciens, des réparateurs, des conducteurs de camions, d'employés de bureau. À cette époque, constate Carl Benedikt Frey (*The technology trap*, 2019), on avait assisté à une polarisation de l'emploi (augmentation de l'emploi peu qualifié ou très qualifié) au détriment de l'emploi moyennement qualifié. Si la révolution industrielle a apporté la prospérité sur le moyen terme, elle s'est traduite à court terme par l'appauvrissement des classes moyennes, la captation des fruits du travail par les plus riches. Mais dans un deuxième temps la standardisation des tâches complexes tend à les rendre accessibles aux salariés les moins qualifiés.

Au tournant du siècle, tandis que l'industrie a besoin d'ouvriers plus qualifiés, se met en place le compromis « fordiste » (production de masse, permettant la disponibilité de produits à prix abordables et redistribution d'une partie des gains de productivité sous forme de salaires plus élevés pour solvabiliser la demande). Parallèlement, le développement de l'éducation secondaire, favorisé par l'interdiction du travail des enfants, a permis de réduire les inégalités et le fossé entre travailleurs qualifiés et peu qualifiés.

Après la Seconde Guerre mondiale, le compromis fordiste a fonctionné à plein. L'existence d'un mouvement syndical puissant a engendré une pression à la hausse sur les salaires, ce qui ne posait pas de problème tant que l'économie générait des gains de productivité élevés (plus de 4% par an en Europe pendant les Trente Glorieuses). Aux États-Unis, dans l'industrie automobile, les syndicats (UAW) avaient renoncé à la grève en échange d'augmentations salariales et d'avantages sociaux (Traité de Détroit en 1950).

Ces gains de productivité, étaient liés, comme le montrent les travaux de Solow, au progrès technologique et à la formation des salariés pour favoriser leur montée en compétence. C'est ainsi que la formation continue s'est développée à partir des années soixante. Les changements technologiques ont donc eu des effets importants sur le marché de l'emploi mais ils ont été tempérés par des politiques publiques (éducation, protection sociale) et un contrat social fort entre employeurs et employés (ce que les économistes appellent le « Fordisme »).

Alors pourquoi pendant deux siècles la part du travail est-elle restée constante et pourquoi les salaires ont-ils augmenté ? Parce que la technologie a créé parallèlement des emplois consommateurs de main d'œuvre. La mécanisation de l'agriculture a été compensée par la montée en puissance de l'industrie (et des services). Comme l'a montré Baumol en 1967, l'emploi s'est déplacé des secteurs les plus productifs vers les moins productifs. Les craintes, exprimées par des économistes comme Keynes et Leontieff ne se sont pas concrétisées. Leontieff affirmait (*Technological Advance, Economic Growth, and the Distribution of Income*, 1983) que la machine remplacera l'homme comme elle a remplacé le cheval. Mais l'homme dispose d'un avantage compétitif par rapport à la machine dans une vaste gamme de tâches, ce qui n'est pas le cas du cheval. Tant que cet avantage est suffisamment important, l'homme trouve d'autres tâches et l'emploi n'est pas menacé.

Le renversement de tendance s'est opéré au milieu des années 70, non pas tant à cause du choc pétrolier de 1973 que du ralentissement notable de la hausse des gains de productivité des deux côtés de l'Atlantique.

De fait, les craintes qui s'expriment aujourd'hui sur la baisse de l'emploi étaient déjà présentes, on l'a vu, chez un économiste comme Leontief dans les années 80. Et de fait, celles-ci n'étaient pas justifiées : selon le cabinet McKinsey, entre 1970 et 2017, le déploiement des ordinateurs personnels (PC) a créé 16 millions d'emplois aux États-Unis et en aurait détruite environ 4 millions (*jobs lost jobs gained*, 2018), soit un gain net de plus de 10 millions d'emplois.

Les délocalisations et le recours à la sous-traitance ont constitué une autre alternative. On pouvait trouver des travailleurs à bas coût, en Chine par exemple, qu'il n'était pas nécessaire de former au-delà du strict nécessaire. D'autant qu'en Chine l'existence d'un réservoir de main d'œuvre dans l'Ouest du pays a longtemps permis d'avoir sous la main des salariés payés moins de 300$ par mois.

Comme le montrent Autor, Dorn et Hanson (2015, 2016), ce sont les délocalisations et la concurrence de pays comme la Chine, plus que l'automatisation, qui ont détruit l'emploi dans les pays anciennement industrialisés. Alors que la Chine affiche des ambitions très importantes en la matière, **il se pourrait bien que le choix ne soit pas d'accepter ou de refuser les robots mais de savoir si les robots de demain seront produits chez eux ou chez nous.** L'automatisation semble conduire en revanche à un phénomène qui n'est pas moins inquiétant, à savoir la disparition progressive de la classe moyenne.

3. Polarisation sociale : vers la fin de la classe moyenne ?

L'autre question porte en effet sur la structure de l'emploi : Pour Ameçoglu et Restrepo (2016), l'automation fait disparaitre certaines tâches routinières et complexifie les tâches restantes, qui deviennent plus résilientes face à l'automatisation. Dans le même temps, les taches complexes se banalisent et deviennent plus accessibles aux travailleurs non qualifiés.

Au cours des deux révolutions industrielles et des Trente Glorieuses, l'emploi s'est en effet déplacé des secteurs les moins productifs (l'agriculture) vers les secteurs les plus productifs (l'industrie). La situation a commencé à évoluer de manière moins nette au cours des cinquante dernières années. Les salaires ont continué à croître régulièrement pour les diplômés. Ceux-ci ont pu monter en

qualification. Et certains emplois d'opérateur dans l'industrie ont été enrichis par la nécessité de piloter des machines à commande numérique et parce que les écrans ont fait leur entrée dans les ateliers. L'opérateur gère ses données pour piloter la production ou le temps d'utilisation des machines. Ils sont devenus des techniciens.

Mais nombre de ceux qui avaient un niveau inférieur au bac sont passés, du fait de la désindustrialisation ou de leur remplacement par des robots, d'emplois moyennement qualifiés à des emplois peu qualifiés. Certains emplois moyennement qualifiés ont été remplacés par des emplois non-qualifiés dans le même secteur. C'est ainsi qu'avec le GPS il faut aujourd'hui moins de qualification pour être chauffeur routier ou chauffeur de VTC. Aux États-Unis, les emplois qualifiés sont passés de 60% de l'emploi en 1979 à 46% en 2012. C'est bien le processus de disparition, d'aspiration par le vide » (*hollowing out*) de la classe moyenne. Selon Watcher, Song et Manchester (2017), toujours aux USA, les personnes ayant plus de trois ans d'ancienneté qui ont perdu leur emploi dans les années 80, n'en ont retrouvé qu'avec une perte de salaire de l'ordre de 30%. 20 ans plus tard, le différentiel de salaire avec leurs collègues ayant gardé leur emploi se situait encore entre 15 et 20%.

Katz et Margo (2014) ont analysé la répartition par qualification de l'emploi industriel aux États-Unis sur la période 1920-2010 :

- L'emploi hautement qualifié ne cesse de progresser, passant de 5 à 30% du total,
- L'emploi moyennement qualifié reste à peu près stable, passant de 35% à 30%,
- L'emploi faiblement qualifié baisse, passant de 60% à 40%.

En 2018, les salaires en termes réels des personnes ayant un niveau inférieur au brevet étaient de 10 à 20% inférieurs à ceux de 1980. On constate des deux côtés de l'Atlantique que les personnes ayant perdu leur emploi ont une plus forte morbidité et une espérance de vie plus courte.

Si l'on s'en tient au seul emploi industriel, il y a bien polarisation, mais moins spectaculaire que si l'on raisonne sur l'emploi total. L'emploi industriel a sensiblement baissé sur la période, au profit des services, qui ont vu une croissance des emplois très qualifiés de cadres supérieurs et d'experts, d'une part, et d'emplois faiblement qualifiés dans les services à la personne ou l'entretien. C'est bien au niveau de la tertiarisation que se fait la polarisation.

Pendant la même période, on a assisté à l'accroissement des inégalités géographiques. L'emploi qualifié a prospéré dans les métropoles et dans certains territoires (l'économie résidentielle sur la côte atlantique en France). Les métropoles comporteraient une part moindre d'emplois facilement automatisables. Tandis que les vieilles régions industrielles souffraient (sauf en Allemagne qui, grâce au *Mittelstand,* a compensé le déclin de la sidérurgie).

Il faut en fait bien distinguer deux types de mutation technologiques en sens opposé : d'un côté, l'automatisation des tâches existantes diminue l'emploi et exerce une pression à la baisse sur les salaires. Autor, Levy et Murnane (2003), de même qu'Ameçoglu et Restrepo (2017, 2018) notent que l'automatisation des tâches routinières implique le développement des inégalités. Mais les mêmes auteurs constatent aussi que l'introduction de nouvelles tâches pour lesquelles l'homme est mieux placé que la machine a un effet positif sur l'emploi et les salaires.

Ironiquement, l'automatisation conduira à ce que les emplois de demain soient plus « humains » car ils se concentreront plus sur les domaines où l'homme dispose d'une valeur ajoutée, mais les conséquences en termes de rémunération sont très différentes selon que ces emplois sont considérés comme très qualifié ou pas. Un aide-soignant en EHPAD n'est pas rémunéré de la même manière qu'un spécialiste du traitement des données.

Aux USA entre 1980 et 2016, les profils les moins rémunérés (entre les déciles 1 et 6) ont vu l'emploi diminuer. C'est là que les salaires ont le moins augmenté en termes nominaux. Les emplois où s'étaient réfugiés ceux qui avaient perdu leur emploi précédent sont à leur tour menacés, sauf sur les soins à la personne). On le voit dans la grande distribution, où, avec l'essor du e-commerce, il y a un déversement des emplois vers la logistique (ex : Amazon), donc vers des emplois moins qualifiés. De même, les travailleurs chassé de l'industrie se sont retrouvés dans les métiers de service peu qualifiés et à faible salaire. Aux USA, dans le secteur du *fast-food*, l'âge moyen est de 35 ans). Il y a une sorte de double peine pour les moins qualifiés : une baisse du nombre d'emplois d'une part et les emplois qui restent sont peu rémunérés d'autre part.

Parmi les travaux des économistes qui ont tenté de comprendre ces phénomènes, on retrouve deux types d'analyses, celles qui portent sur l'évolution de la productivité du travail et celles qui privilégient la fin du compromis fordien comme élément de rupture.

4. Une réallocation régressive ?

Alors que Stansbury et Summers (2017) affirment que la croissance de la productivité et des salaires sont corrélés positivement, Acemoglu et Restrepo (2018) constatent pourtant qu'aux États-Unis, sur la période 1970-2007, les secteurs qui enregistrent une hausse significative de la productivité globale des facteurs subissent *aussi* une baisse de l'emploi et de la part des salaires dans la valeur ajoutée. À l'inverse, pour Aghion et Bergeaud (2019), la baisse de la part de salaires dans la valeur ajoutée est liée à la plus forte croissance relative des secteurs à bas salaires.

Aux USA, la croissance de la productivité a commencé à s'essouffler à partir de la fin des années soixante-dix. On constate justement que, depuis les 30 dernières années, il y a eu une accélération de l'automatisation et une baisse du rythme de création de nouveaux emplois dans l'industrie (qui représente moins de 8% de l'emploi total contre 13.5% en France)

5. La fin du compromis fordien

Jusque dans les années soixante-dix, les pays occidentaux avaient connu une inégalité modérée et une classe moyenne robuste. Mais les choses sont changé : aux USA, la productivité du travail a crû de 75% entre 1973 et 2016, alors que le revenu salarial médian n'augmentait que de 12%.
Dans le processus de réallocation des emplois, les salaires n'augmentent pas autant que la productivité. Les observations empiriques (Ameçoglu et Restrepo, 2018) font ressortir une baisse de la part des salaires dans la valeur ajoutée.

Dans son ouvrage sur *La société en réseau,* Manuel Castells note que l'économie de l'information introduit une dichotomie entre « gestionnaires informés » et main d'œuvre prolétarisée. De fait, la distribution des salaires entre hauts et bas salaires a totalement divergé. Ceux qui étaient dans les 5 plus bas déciles ont vu leurs revenus stagner. Aux USA, 90% des personnes nées en 1940 gagnaient à l'âge adulte plus que leurs parents. Ce ratio est tombé à 50% pour les personnes nées en1980.

Comme l'a très bien vu Pierre Veltz, raisonner uniquement sur l'automatisation ne permet pas de disposer d'une vue d'ensemble. Il faut penser le triangle mondialisation-financiarisation-numérisation. Face au ralentissement des gains de productivité, les dirigeants et actionnaires des grandes entreprises ont répondu par l'automatisation et les délocalisations. L'implantation de robots industriels entraîne une baisse de la demande de travail humain et une baisse de l'emploi et des salaires. Cela a touché les salaires les moins élevés et les personnes les moins éduquées. Et contrairement à une idée reçue, l'Europe est la région où l'automatisation est la plus utilisée pour faire baisser les coûts.

	Europe	Asie	Amérique du Nord
Baisse des coûts	68%	62%	64%
Développement des revenus	32%	38%	36%

Source : McKinsey 2019

Bengtsson, Rubolino et Waldenström (2020) considèrent de leur côté, en citant notamment le cas du *Trade Union Act* de 1984 au Royaume-Uni, que c'est la désyndicalisation qui a été un des éléments majeurs de l'inflexion de la baisse des salaires dans la valeur ajoutée.

Globalement, on peut dire que les organisations syndicales ont, des deux côtés de l'Atlantique, eu du mal à articuler une réponse à l'automatisation. Elles ont eu à faire face à un patronat conquérant, motivé par la perspective des *stock-options* et elles ont aussi été victimes de leur bureaucratisation, selon une logique décrite il y a plus d'un siècle par Roberto Michels au sein de la sociale démocratie allemande.

Les syndicats ont joué les « insiders » au détriment des « outsiders » précaires. C'est le cas en Allemagne où la vigueur de l'industrie a permis de maintenir un emploi industriel conséquent et bien rémunéré, mais le prix des réformes Harz a été un accroissement sensible des inégalités, qui n'a pas été compensé comme en France par les transferts sociaux

6. Tout ça pour ça ?

Acemoglu et Restrepo, dans un article qui tranche avec l'optimisme technologique ambiant (*The wrong kind of AI*, 2019) viennent renouveler le débat autour de l'automatisation et de l'emploi, qu'ils ont d'ailleurs fortement contribué à structurer : investissons-nous dans le bon type d'Intelligence Artificielle (et plus largement dans le « bon » type d'automatisation), celle qui améliore la productivité et amène la prospérité ? Il n'y a en effet de gain au niveau collectif que si l'automatisation génère des gains de productivité significatifs qui seront ensuite redistribués en partie aux salariés. Or ce n'est pas toujours le cas, tant s'en faut.

La « révolution » numérique ne s'est pas traduite par un rebond sensible des gains de productivité alors qu'elle a contribué à réduire les coûts. Par ailleurs les innovations qui augmentent la productivité ne réduisent pas forcément l'emploi et celles qui réduisent l'emploi n'améliorent pas forcément la productivité. C'est ce qu'ils appellent les innovations « so-so » (« couci-couça »).

Si les gains engendrés par l'automatisation n'ont pas toujours été à la hauteur des espérances, c'est notamment parce que l'automatisation de substitution n'améliorerait que faiblement la productivité : la machine fait un peu mieux que l'homme, mais pas beaucoup. Si la productivité du travail dans les tâches qui vont être automatisées est déjà importante (ou que les salaires dans le secteurs sont très bas), le gain de productivité lié à l'automatisation est faible. Ce peut être le cas dans la relation client, où la qualité des « *chatbots* » est très inégale. De même, lorsque l'intelligence artificielle s'attaque à des compétences où l'homme avait un avantage distinctif, comme la reconnaissance vocale ou visuelle, il y a substitution sans qu'on constate de bond significatif dans la productivité.

Le fétichisme technologique correspond à un problème bien identifié par les consultants en management, que les Américains appellent « *Better mouse trap syndroma* ». Si on adopte une technologie sans que celle-ci fasse vraiment mieux que les humains, on a dépensé de l'argent pour pas grand-chose. Il y a malheureusement dans la Silicon Valley une idéologie de l'automatisation au détriment du travail humain ; l'idée que la machine est (ou sera) forcément plus efficiente que l'homme a fait bien des dégâts.

On peut identifier plusieurs causes à la source de ce type d'égarement :
Tout d'abord on constate que, traditionnellement, les entreprises investissent plus dans l'automatisation en période haute du cycle économique, lorsque les tensions sur le marché du travail se traduisent par une hausse des salaires et que remplacer des humains par des machine semble un bonne affaire. C'est moins vrai sur une longue période, quand l'alternance des cycles économiques rend moins évidente la rentabilité des investissements en machines (et en logiciels). D'autant que ceux-ci constituent un coût fixe, alors qu'il est possible de licencier des salariés en temps de crise et d'embaucher des intérimaires pour faire face aux pointes d'activité. Jaimovitch et Siu considèrent à l'inverse que ce sont les phases de crise économique qui sont propices à l'automatisation des tâches routinières, ce qui expliquerait en partie que les reprises ne sont pas si créatrices d'emploi qu'on aurait pu l'espérer (*jobless recoveries*), d'autant que les consommateurs, pour préserver leur pouvoir d'achat écorné par la crise, se reportent sur des produits moins riches en main d'œuvre, donc plus « automatisés ».

Ensuite, la persistance de taux d'intérêts réels faibles (et c'est le cas depuis 2012 en Europe) et la déductibilité des intérêts d'emprunt encouragent l'investissement en capital au détriment du travail. Cela est vrai aussi en période de récession, alors que la pression à la baisse sur les salaires rendrait le choix de l'automatisation moins pertinent. Le problème est que si le coût du capital à long terme est inférieur au coût du travail, il n'y a pas suffisamment d'incitation à créer de l'emploi. Acemoglu et Restrepo militent d'ailleurs pour une refonte de la législation fiscale.

Enfin Rodrik (2020) vient compléter cette analyse. Pour lui, les choix technologiques ne sont pas politiquement neutres. Ils sont trop souvent guidés par une idéologie du salarié « jetable ». Au contraire, les entreprises doivent créer des emplois pour leurs salariés tels qu'ils sont, et qui sont disponibles, et non pas pour les salariés qu'elles rêveraient d'embaucher. C'est à la technologie de s'adapter. L'impératif du changement perpétuel, de l'adaptation permanente, débouche sur un « trotskisme managérial » générateur de souffrance et de misère.

7. Et en France ?

En France, les analyses de l'INSEE et de la DARES (Ministère du Travail) font ressortir un tableau assez différent de celui des États-Unis. On assiste bien à une forte croissance des emplois de cadres et de service à la personne d'une part et à une baisse continue d'emplois d'ouvriers peu qualifiés d'autre part, qui représentent l'essentiel de la baisse des effectifs de l'industrie au cours des trente dernières années. Ceci explique la stagnation de l'emploi masculin sur la période.

Mais la grande différence avec les États-Unis est la croissance régulière des professions intermédiaires (agents de maîtrise, enseignants, infirmières, etc.) dont la moitié sont titulaires du baccalauréat. Ce serait donc la relative robustesse du système d'enseignement secondaire qui aurait empêché jusqu'à présent l'effondrement de la classe moyenne.

On constate en revanche, comme outre-Atlantique, une forte croissance des métiers de service à la personne peu qualifiés, principalement occupé par des femmes. L'exercice prospectif mené par la DARES et France Stratégie en 2015 (*Les métiers en 2022*) laisserait apparaître une poursuite de la dynamique de l'emploi observée ces dernières années, avec une forte croissance des emplois qualifiés (cadres, informaticiens), les professions intermédiaires restant dynamiques, avec des besoins de qualification plus importantes dans le bâtiment et la logistique.

L'autre tendance est le développement des métiers du secteur médico-social qui continueraient à augmenter sensiblement, sous le double effet de la croissance des effectifs et du remplacement des départs en retraite. Les catégories qui connaitraient la plus forte croissance nette seraient les cadres administratifs et financiers, les aides à domiciles et les aides-soignants.

Enfin on assisterait à la poursuite de la baisse des emplois peu qualifiés non seulement dans l'industrie mais aussi dans les services, avec une érosion des métiers administratifs les moins

qualifiés. Ce tableau plutôt rassurant ne doit pas faire oublier la persistance d'un chômage de masse : aux États-Unis, l'ajustement s'est fait sur les salaires (effet prix) alors que dans notre pays, il s'est fait sur la population employée (effet volume).

Si nous ne disposons pas pour la France d'analyse aussi poussée que celles d'Acemoglu et Restrepo ou d'Autor, il est possible de s'appuyer sur les analyses de l'INSEE pour tenter d'appréhender comment l'automatisation (en l'occurrence l'informatisation, puisque hors automobile, le parc robotique français reste assez peu développé) a fait évoluer la structure de l'emploi.

L'impact de l'informatique est sensiblement différent selon que l'on étudie les secteurs traditionnels (pauvres en R&D) et les secteurs innovants (riches en R&D). Dans les secteurs traditionnels, l'informatisation se traduit par de forts gains de productivité et une forte baisse de l'emploi, alors que dans les secteurs innovants, on ne constate pas de baisse d'effectifs. Les emplois y sont moyennement ou fortement qualifiés. Pak et Poissonnier (2016) confirment les thèses d'Autor sur la résilience des emplois non-routiniers face à l'automatisation. En France, l'informatisation détruit de l'emploi peu qualifié et développe l'emploi qualifié.

Il semble que les gains de productivité soient mieux répartis en France, ce qui se traduirait par une moindre déformation de la structure des emplois et des rémunérations. Entre 1994 et 2015, la part des salaires dans la valeur ajoutée est restée stable aux environs de 58% en France, alors qu'elle a baissé de 3% en Allemagne et aux Etats-Unis.

Dans leur étude portant sur l'évolution de l'emploi en France sur la période comprise entre 1982 et 2010, Mathilde Pak et Aurélien Poissonnier concluent à augmentation des emplois qualifiés, portée par le développement du secteur manufacturier de haute technologie (aéronautique, ferroviaire, défense et R&D).

Alors que plusieurs économies développées font apparaître une polarisation de l'emploi au détriment des personnes moyennement qualifiées, la tendance dans notre pays est différente : il y aurait eu une montée de l'emploi qualifié, le développement des services se faisant au détriment de l'agriculture et de l'industrie manufacturière traditionnelle. Globalement, ce sont les métiers qualifiés qui ont porté la croissance de l'emploi. Entre 1982 et 2010, la part des emplois hautement qualifiés est d'ailleurs passée de 8% à 17%. Les emplois de services peu qualifiés ont remplacé d'autres emplois non qualifiés, qui étaient principalement situés dans l'industrie. Comme le note le think tank *La Fabrique de l'industrie*, dans les régions désindustrialisées, les ouvrières sont devenues caissières de supermarché.

Au cours de ces dernières années, on observe néanmoins une certaine tendance à la polarisation de l'emploi, avec le développement des emplois peu qualifiés dans les services à la personne (très majoritairement féminin) et une politique fiscale favorable au maintien des emplois dont la rémunération est inférieure à 2,5 fois le SMIC (instauration du CICE en 2015, transformé en allègement de cotisations en 2019).

Contrairement à ce qu'a observé David Autor aux États-Unis, il semble que le commerce extérieur a eu un effet positif sur l'emploi et la production, avec, il est vrai, de fortes disparités régionales. Si l'on en croit Pak et Poissonnier, ces disparités reflètent en réalité une dichotomie entre entreprises innovantes (dépenses de R&D significatives) et entreprises plus traditionnelles. À ce titre, ils remarquent qu'il vaut mieux occuper un emploi faiblement qualifié dans une entreprise industrielle innovante, car on y profite d'un effet positif sur l'emploi et les salaires, que dans un secteur à faible rentabilité.

Si la productivité du travail a augmenté de 52% sur la période, elle a été inégalement répartie : pour un million d'euros de chiffre d'affaires, le coût de production est passé au cours de la période de 1 à 0,3 (-70%) dans les entreprises innovantes et à 0.8 (-20%) dans les services non-échangeables. On touche probablement là à une faiblesse du modèle français : le différentiel de dépenses de R&D entre la France et l'Allemagne est aujourd'hui de l'ordre de 20 milliards d'euros par an,

principalement au niveau des entreprises. Ceci s'explique principalement par la baisse de la part de l'industrie en France.

B. Les libertés publiques

1. La rué vers la collecte des données personnelle

L'affaire *Cambridge Analytica*, dans laquelle le cabinet de conseil au service du Parti Républicain avait acheté une base de données de profils élaborés à partir de la fonction *SocialGraph* de Facebook a causé un énorme scandale. Mais c'est oublier qu'en 2008, la campagne Obama avait déjà utilisé des profils d'électeurs extrêmement détaillés car il est possible aux États Unis de se procurer des bases de données sur les citoyens et de croiser les données (âge, sexe, statut marital, profession, niveau de revenu, marque de la voiture, abonnement à des magazines, type de loisirs, etc.) qui lui avaient permis de cibler très précisément les électeurs. Il est vrai que la collecte des données est une chose, et que les algorithmes en sont une autre. Les algorithmes permettent de traiter une masse importante de données, et c'est ce qui a rendu les campagnes menés par *Cambridge Analytica* si efficaces.

2. Vers un « capitalisme de surveillance » ?

Contrairement à ce qui se passe en Europe, les États-Unis se distinguent par le manque de garde-fous juridiques en matière de collecte et d'analyse des données personnelles des citoyens, faits sans leur consentement, par des entreprises privées. L'industrie (privée) du « *background check* », qui sert par exemple à vérifier votre solvabilité, vos problèmes passés avec l'alcool ou la drogue, le bon paiement de vos loyers ou vos démêlés, même mineurs avec la police, est florissante. D'autant que le « droit à l'oubli » n'existe pas aux États-Unis.

L'entrée d'acteurs issus du numérique sur ce marché a encore renforcé cette tendance. Ils ont à la fois accès à une masse de données bien plus considérable et disposent d'outils d'analyse algorithmique extrêmement puissants. Comme l'a bien résumé un commentateur : « Foursquare sait où vous êtes, Flickr sait ce que vous voyez, Facebook sait ce que vous faites, Amazon sait ce que vous achetez, Google sait ce que vous pensez. »

Et les exemples d'abus possibles sont légions :

- Lorsque Facebook met en place un outil de détection des pensées suicidaires après l'affaire Katelyn Nicole Davis, pour alerter les services d'écoute et de prévention, on se rapproche du monde de *1984*.
- L'application Donteat.at permettant de recenser les restaurants qui ne respectent pas les règles d'hygiène.
- Facebook a mis au point un algorithme de calcul de solvabilité : qui va l'utiliser ?
- Les assistants vocaux (Siri, Google Home, Echo) peuvent écouter les conversations à l'insu des utilisateurs sous le prétexte, justement, d'entraîner les données ainsi recueillies. Google s'est défendu en plaidant que « *si vous faites quelque chose que vous souhaitez cacher, alors ne le faites pas* », faisant écho aux propos du philosophe Jean Bodin, qui affirmait dès le XVI^{e} siècle, qu' « *il n'y a que le méchant qui a intérêt au secret, le bon citoyen n'a rien à cacher* ».
- Le système de surveillance domestique, Ring Alarm d'Amazon, qui allie sonnette connectée et caméras de surveillance, permet aux services de police d'identifier toutes les personnes qui sonnent à votre porte.

Et c'est d'autant plus inquiétant que ces outils, utilisés sans contrôle, ne sont pas forcément, tant s'en faut, plus pertinent que les méthodes traditionnelles.

Ainsi, aux États-Unis, nombre de procès ont opposé des candidats à la location à des sociétés de vérification d'antécédents de paiement de loyer. Au Michigan, le remplacement par le logiciel MIDAS des vérifications des fraudes aux allocations chômage, qui étaient auparavant faites par les agents de l'agence locale pour l'emploi, a fait grimper le taux d'erreur de 44% à … 93% !
Et si l'on cite volontiers le système de reconnaissance faciale mis au point par Baidu pour les clients chinois de KFC, qui propose un menu personnalisé en de l'âge, du sexe ou de l'expression du visage, on oublie que nos pays ne sont pas épargnés. Ainsi, en Europe, 80% des applications mobiles gratuites collectent des données à l'insu des utilisateurs. On voit même des acteurs publics comme la RATP procéder ainsi. D'autant que la plupart des utilisateurs ne savent pas configurer les options de confidentialité de leur smartphone. Il existe des logiciels destinés à l'espionnage des smartphones comme *Stealthgenie* ou *Ikeymonitor*, qui peuvent être installés par des personnes malveillantes (concurrent ou… conjoint).

Pour éviter les dérives observées aux États-Unis ou en Chine, l'Union européenne a mis en place en 2016 le Règlement Général pour la Protection des Données (RGPD), qui est applicable depuis le 25 mai 2018. Le RGPD vise à encadrer la collecte et la conservation des données personnelles. Il s'applique dès lors que les données d'un résident européen sont collectées. Il est venu consacrer des dispositions existant en droit français telles que l'interdiction d'interconnexion des bases de données, le droit à l'oubli, la protection de la vie privée et la limitation de la durée de stockage des données.

Il a été complété par la Directive 2016/680 du même jour, transposée en droit français par la loi 2016-1321 du 7 octobre 2016, qui élargit les pouvoirs de sanction de la CNIL et autorise l'ouverture des données publiques (*open data*), ainsi que le droit à la communication des algorithmes utilisés par les administrations pour les personnes concernées (confirmé par un arrêt du Conseil d'État en date de 3 avril 2020).

Le RGPD est opposable aux entreprises situées hors de l'Union. Il autorisait cependant le transfert de données hors de l'UE, si le pays destinataire des données dispose d'une réglementation sur la vie privée « au moins équivalente » à celle de l'Union européenne ou que les entreprises offrent des garanties suffisantes (« *Privacy Shield* »). Cette possibilité offerte aux entreprises américaines a pris fin en juillet 2020 avec l'abolition de ce dispositif, qui permettait ces transferts sous condition.

3. Les dangers de la surveillance de masse

Face aux problèmes de violence et de terrorisme, la reconnaissance faciale est devenue la solution-miracle. Elle est utilisée de manière croissante par la police des grandes métropoles américaines, couplée à des algorithmes de prévision de la récidive. Comme on l'a vu dans le cas d'Amazon, les techniques de surveillance de masse associent souvent les structures publiques et des partenaires privés.

Le programme PRISM mis en place par la NSA depuis 2002 permet de collecter les données des clients de Microsoft, Yahoo, Google, Facebok, Paltalk, Youtube, Skype, AOL et Apple, directement sur les serveurs de ces entreprises pour toute personne « *of interest* » sur la base d'une « suspicion raisonnable ». Plus récemment, toujours aux USA le projet MAVEN a pour objet le traitement, par des outils d'intelligence artificielle, de l'ensemble des informations collectée par les drones d'observation, avec des partenaires comme Google (qui s'est retiré depuis), Amazon et Palantir.

En Russie, a été mis en place le système SOBM, complété par la loi Yarovaya entrée en vigueur en 2018, qui demande aux opérateurs télécom de conserver les informations sur leurs clients pendant 6 à 36 mois (contre 12 mois en France jusqu'à l'arrêt de la Cour de Justice de l'Union européenne d'octobre 2020).

L'exploitation des métadonnées, utilisée par exemple par l'administration Obama dans le cadre de la répression du mouvement *Occupy Wall Street* (septembre-novembre 2011), permet de savoir qui a appelé qui, quand, depuis quel endroit et à quelle fréquence.

En Chine, dans plusieurs grandes villes, dans le cadre du système de Crédit Social, la photo des personnes qui traversent en dehors des passages réservés apparaît immédiatement sur un écran mural géant. Dans certains cas, l'écran affiche la carte d'identité et l'adresse du « délinquant ». Plus récemment, en Chine et en Corée du Sud notamment, les autorités ont eu recours à la géolocalisation des personnes contaminées par la Covid-19 à partir des données des opérateurs pour lutter contre le coronavirus.

Partout dans le monde, le recours massif des outils algorithmiques par les gouvernements, couplés à l'interconnexion de multiples bases de données, suscite la peur de voir les données personnelles de la population utilisées à mauvais escient.

Or il faut reconnaître que les moyens de surveillance de masse n'ont pas empêché les attentats terroristes, mais risquent d'incriminer à tort des innocents. Richelieu n'affirmait-il pas « *Qu'on me donne six lignes écrites de la main du plus honnête homme et j'y trouverai de qui le faire pendre* » ?

Le cas délicat de la « Smart city »

Depuis quelques années, cabinets de conseil, éditeurs de logiciels spécialisés et prestataires informatiques ont fait le siège des collectivités locales pour leur proposer des projets de numérisation des services publics.
Le concept de « *smart city* » ou ville intelligente est certes alléchant. Il ne s'agit rien de moins que d'améliorer la qualité des services tout en diminuant les coûts. La numérisation permettrait par exemple de réduire les embouteillages et la pollution en optimisant le trafic urbain. Pour les élèves et les étudiants, la même carte permet l'usage de la cantine et des transports publics.
Dans une large mesure, la *smart city* est l'extension de la notion de gestion technique centralisée, visant à piloter depuis un point unique les différents équipements et services pour « mieux vivre dans la ville ». Et c'est cette centralisation qui peut poser problème du point de vue des libertés publiques.
À cela s'ajoutent des applications plus orientées vers la sécurité ou la gestion des foules, reposant sur des formes atténués de reconnaissance (reconnaissance d'une forme plutôt que d'un visage) par des circuits de vidéo-surveillance et faisant appel à des outils d'apprentissage automatique. Dans un contexte déjà marqué par la controverse sur les compteurs de télé-relevés d'électricité Linky, la *smart city* a suscité une certaine méfiance de la population.
Si le législateur français a instauré certains garde-fous (les données collectées sont en France la propriété des collectivités, par exemple), il n'en va pas de même aux États-Unis. À New York, le maire Bill De Blasio s'était engagé en 2017 à communiquer à une commission de citoyens toutes les informations nécessaires sur les algorithmes mis en place dans le cadre des projets de smart city. Ce fut un fiasco complet. La commission n'a même pas pu obtenir la liste des fournisseurs de logiciels, et encore moins d'information sur le fonctionnement des algorithmes, sous prétexte de secret des affaires. Au Canada, le projet de *smart city* de Google à Toronto a fait long feu en raison de polémiques liées à la masse des données concernant les résidents que Google envisageait de collecter.

C. Comprendre la querelle des biais

Le recours à des outils d'apprentissage automatique, dans la reconnaissance faciale par exemple, a suscité aux États-Unis de nombreuses polémiques quant à l'existence de biais au détriment de certaines minorités. Ceci s'inscrit dans la montée du « politiquement correct » et l'hypertrophie de la question raciale aux États-Unis Les critiques portent notamment sur la représentation des minorités ethniques dans les données.

Selon la National Academy of Sciences, les assistants vocaux type Alexa ont un taux de reconnaissance de 80% pour les Blancs mais de seulement 65% pour les Noirs. Les Blancs sont plus nombreux et aussi plus aisés, donc plus susceptibles d'acquérir ce type de matériel, et donc plus présents dans les données d'entraînement. De même, selon l'agence des normes US National Institute for Standards and Technology (NIST), l'équivalent de l'AFNOR, les Noirs et les femmes sont moins bien identifiés par les principaux algorithmes de reconnaissance faciale. En 2019, *Rekognition*, le logiciel de reconnaissance faciale d'Amazon, a confondu des athlètes connus de Boston avec des délinquants. Le scandale a conduit l'entreprise a en « suspendre » l'usage en 2020.

Les critiques ont porté principalement sur les algorithmes qui seraient selon l'essayiste Cathy O'Neill « des opinions intégrées dans un code ». La question est cependant plus complexe.

Les biais peuvent s'introduire à trois niveaux

- Lors de la conception des algorithmes, la définition des seuils de significativité et celle des objectifs. Les humains gardent souvent une part d'implicite lorsqu'ils fixent des objectifs, quitte à vérifier a posteriori qu'il s'agit bien des objectifs qu'il avait en tête ;
- Lors de la collecte et la mise en forme des données. Les données peuvent n'être pas pertinentes car la sélection est trop restrictive ;
- Lors de l'exploitation et notamment la sélection des variables, avec les risques associés au principe même du Big Data, pour qui corrélation vaut causalité.

Il est alors très difficile d'identifier et de corriger des bais ou des erreurs a posteriori.

1. Au niveau des algorithmes

Dans l'apprentissage automatique, la programmation consiste surtout à définir la structure des réseaux de neurones : le nombre de couches cachées, la taille des filtres et de *feature maps* (matrices de pixels utilisés pour la reconnaissance des images et des visages) pour les réseaux de neurones convolutionnels, puis à déclencher son entraînement avec une boucle de programme qui va scanner un jeu d'entraînement.

Si l'appareil de règle est bien moins important que dans l'IA symbolique, les « *patterns* » sont bien renseignés à la main. Des biais peuvent apparaître dans la construction des *feature maps* ou dans la configuration de l'algorithme : en 2018, l'ACLU a croisé une base de données de 530 membres du Congrès avec une base de données de criminels, avec un algorithme produit par Amazon. Il a détecté 28 criminels parmi les membres du Congrès. Les méthodes traditionnelles de profilage basés sur des jugements humains ne sont pas à l'abri des erreurs, comme le montre le cas de de Richard Jewell, le héros du dernier film de Clint Eastwood.

Mais ces erreurs peuvent être renforcés, dans le cas de des outils algorithmiques par la présomption d'infaillibilité des systèmes (« *mathwashing* »). Le risque est de dire : « les chiffres parlent d'eux-mêmes », ce qui peut aussi se traduire par : « Ce n'est pas moi, c'est l'algorithme».

2. Au niveau des données

Le problème des biais concerne plus les données que les algorithmes eux-mêmes. On parle de biais de disponibilité : on ne tire des conclusions qu'à partir des données disponibles alors qu'on pourrait obtenir des résultats différents avec des données plus complètes. Par exemple, une entreprise (c'est souvent le cas en informatique) dont la majorité du personnel est masculine et qui utilise sa base RH pour entraîner un algorithme de recrutement va probablement privilégier les candidats du même sexe (et à l'inverse dans le domaine de la mode).

Si les données d'entraînement sont en décalage par rapport aux populations réelles, il y a forcément un risque de biais statistique. En effet, si l'échantillon de la base de données d'entraînement n'est pas suffisamment large, il y des cas qui ne seront jamais reconnus ou des populations sous-estimées.

Rappelons que plus de la moitié des données de la base ImageNet utilisée pour la reconnaissance d'images provient des de deux pays, les États-Unis et le Royaume-Uni.

Les populations mal prises en compte par les statistiques (par exemple celles qui se connectent peu ou pas à internet) sont sous-représentées dans les jeux de données. Cela peut générer des faux positifs (sur une faible population, un petit nombre d'individus présentant une caractéristique faiblement présente dans la population générale mais surreprésentées dans l'échantillon) ou de faux négatifs (tendance non détectée par l'algorithme).

Ceci met en lumière l'importance des *data engineers*, qui sont chargés de collecter et de préparer les données en amont de leur traitement, en s'appuyant sur des juniors, les *data analysts*. Le *data engineer* doit avoir, en sus de ses compétences fonctionnelles, une très bonne maîtrise des méthodes statistiques et des enjeux du métier pour évaluer notamment le degré de pertinence d'une population. Ces compétences sont très rares sur le marché, à l'opposé des *data scientists* en charge de l'exploitation des données, qui sont avant tout des développeurs informatiques. Ces derniers conçoivent de moins en moins les algorithmes de bout en bout car on trouve de plus en plus de briques logicielles disponibles en *open source*.

3. La sélection des variables et les problèmes de traitement des données

En 2018, une équipe d'Amazon a créé un algorithme en utilisant 50.000 mots-clés censés représenter les caractéristiques des employés performants afin d'optimiser ses recrutements. Comme les données d'entraînement étaient basées sur des CV très majoritairement masculins, l'algorithme a privilégié les candidatures masculines.

Facebook, de son côté, a développé un algorithme pour les annonceurs avec un filtre d'« affinités ethniques » pour cibler par exemple uniquement les *prospects* hispaniques, ce qui permet à l'inverse d'éliminer certains groupes. Ce n'est pas forcément gênant pour les annonces publicitaires. Ce l'est plus lorsqu'il s'agit d'offres d'emploi. De même, Facebook offrait la possibilité à certaines entreprises de cibler leurs offres d'emploi à partir des CV de leurs salariés.

Les algorithmes permettent d'introduire discrètement des discriminations par l'âge (en sélectionnant certains prénoms représentatifs d'une génération ou, plus subtilement, en sélectionnant un poste en fonction de l'année où il a été occupé). I. Ajunwa, professeur de droit du travail à l'université Cornwell, donne l'exemple d'un abus de corrélation. Dans le cadre d'un recrutement, une entreprise cherche à déterminer le profil du salarié idéal. L'algorithme montre que les dix meilleurs pratiquent l'alpinisme pendant leur temps de loisir. Elle va donc recruter en priorité des alpinistes amateurs, alors que cette qualité n'a aucun rapport avec les compétences exigées pour le poste à pourvoir.

Les analyses souffrent aussi souvent du manque de prise en compte du contexte : un algorithme entraîné sur une population donnée ne sera pas pertinent sur une autre population (piège de la portabilité).

Les procédures de création des algorithmes, de choix et de labellisation des données sont couvertes par le secret des affaires mais il n'est pas exclu que, dans le futur, les entreprises puissent être tenues responsables du choix des données d'entraînement et être accusées de négligence coupable si elles induisent des biais trop évidents.

La « police prédictive » en question

« *Peut-on imaginer un monde dans lequel des programmes réaliseront de meilleures enquêtes policières que les humains ? Non car l'enquêteur logiciel a peu de chances de voir les particularités entourant le contexte d'un meurtre* » Jean-Louis Dessalles.

Depuis quelques années, les services de police des grandes villes américaines ont recours à des logiciels de cartographie de la criminalité, complétés par des algorithmes dits de « police prédictive » visant à optimiser le travail des patrouilles en les concentrant sur les zones les plus sensibles.

Parmi ces acteurs, on peut citer Compass, qui évalue la probabilité de se faire arrêter de nouveau pour récidive dans les deux ans suivant la libération, Palantir, Northpoint et surtout Predpol Software LLC. Née en 2012 d'une collaboration entre les services de police de Los Angeles et l'université publique UCLA, Predpol est aujourd'hui l'outil le plus utilisé pour anticiper les lieux où peuvent se passer les crimes.

Boucles de rétroaction

Le score de criminalité potentielle est fondé en partie sur le nombre d'interactions d'une personne avec la police, sans qu'elle ait forcément commis un délit. Le logiciel Predpol utilisé à LA jusqu'en avril 2020 déterminait des zones sensibles d'une centaine de m2 où un crime a le plus de chances d'être commis, car les gangs ont souvent un ancrage ultra-local. Une cinquantaine d'unités de police aux Etats-Unis utilisent le logiciel Predpol.

Les algorithmes de Predpol déterminent des zones sensibles, où il y a plus de délits que la moyenne. Une méthode manuelle ne donnerait pas des résultats sensiblement différents, mais elle conduit à renforcer la routine, et la répétition des patrouilles dans les mêmes endroits finit par apparaître comme une forme de harcèlement pour les résidents qui n'ont rien à se reprocher.

Le risque est en effet de construire des « boucles de rétroaction » : de suspect vous devenez coupable. Pour les activistes opposés à l'utilisation de ces outils (dont on ne sait pas s'ils ne sont pas liés à la pègre), comme ceux de la «Stop Spying Citizens Coalition », ceux-ci entretiennent les personnes concernées dans la marginalité au lieu de les aider à s'en sortir. Dans le passé, les personnes ayant certaines caractéristiques ont présenté une forte prévalence de comportements criminels ou déviants. Donc toute personne présentant ces caractéristiques va reproduire le même comportement. L'apparition de ces caractéristiques peut générer une « alerte » dans le système. Dès lors, on soupçonne même des innocents qui ressemblent à des criminels du passé.

Risques d'erreur

Les algorithmes de prédiction de la récidive aux USA utilisent des variables liées (quartier pauvre, parent emprisonné, etc.) et qui sont plus présentes chez les Noirs. Du coup, le risque d'erreur de prédiction (faux positif) serait quasiment double chez les Noirs par rapport aux Blancs. Certains algorithmes de « predpol », comme celui développé par Northpoint, font apparaître pour certaines populations des risques de récidives sensiblement différents de ceux constatés par les statistiques. Faute d'intervention humaine, les systèmes d'aide à la décision automatisatés deviennent des systèmes de décision autonomes et ont un effet darwinien. Le même type de reproche a été adressé aux algorithmes de notation des enseignants, ou ceux qui prédisent la récidive de maltraitance des enfants.

En avril 2020, après le décès de George Floyd, la police de Los Angeles (LAPD) a annoncé la fin de tous ses programmes de police prédictive. Un audit de l'inspection générale du LAPD avait conclu en 2019 qu'il n'était pas possible d'évaluer l'impact réel de ces programmes sur la prévention du crime. Ou était redondant avec le travail des patrouilles sur le terrain.

D. Les boites noires : à la recherche de l'explicabilité

Comme on l'a vu, dans l'approche symboliste, on définit des règles, on applique ces règles à des données et on obtient des résultats conformes à ces règles. On peut relier les faits et les règles et on sait donc pourquoi on a obtenu ces résultats.
Dans l'approche connexionniste, on prend des données que l'on qualifie. Puis, par tâtonnement, la machine fait ressortir des corrélations statistiques et des probabilités (phase d'entraînement). C'est pourquoi on parle de *« machine learning* ». Puis on applique ces règles et ces corrélations à de nouvelles données (phase d'inférence). On obtient des résultats, qui parfois peuvent être inattendus. À ce stade, on n'est pas capable de dire pourquoi on a obtenu ces résultats. Plus l'algorithme est complexe (à cause de la « profondeur » du réseau de neurones, d'où le nom de « deep learning »), plus on a du mal à retracer le cheminement de l'information au travers des différentes couches de neurones. C'est toute la problématique de l'explicabilité.

On obtient des corrélations, mais pas des relations de cause à effet. Le problème est que les règles que les réseaux de neurones font émerger sont implicites et générés au fur et à mesure du cheminement de l'algorithme sans que l'on puisse y accéder. Qui plus est, la propriété intellectuelle concernant les algorithmes et le secret des affaires pour les données constituent autant d'obstacles à la transparence des décisions prises à l'aide d'outils d'intelligence artificielle en apprentissage profond.

Lorsqu'une décision prise par un algorithme se traduit par un refus de prêt ou une condamnation en justice, on se doit de comprendre pourquoi. En outre, lorsqu'en *deep learning* les réseaux de neurones décrivent des objets à partir de leur environnement et de leur contexte, il y a un risque d'erreur de contexte. On commence à détecter ces biais dans les outils de reconnaissance de parole, mais c'est beaucoup plus difficile avec la reconnaissance faciale.
Pour Yann Le Cun, directeur de la recherche et du développement-produit en Intelligence Artificielle chez Facebook, l'explicabilité est un faux problème. Son argument est le suivant : on ne sait pas précisément comment fonctionnent les médicaments, mais on a des protocoles qui permettent de tester leur efficacité. Il reste qu'on ne peut mesurer que partiellement les effets secondaires avant l'autorisation de mise sur le marché et que certains ne sont détectés qu'a posteriori, après diffusion auprès du grand public, comme en témoignent par exemple les scandales de l'Oxycontin ou du Mediator.

Le Big Data, qui est la recherche de caractéristiques significatives dans une masse importante de données, a induit un changement épistémologique qui n'est pas sans poser problème, puisqu'on se fie uniquement aux corrélations en faisant l'impasse sur l'explication Un exemple, qui a donné lieu à des polémiques universitaires mémorables, illustre bien le piège. En 2001, deux économistes de l'Université de Yale, aux États-Unis, Stuart Levitt et John Donohue, publient un article intitulé « *The impact of legalized abortion on crime* ». Ils partent du constat qu'à partir de 1992, la criminalité, mesurée par le nombre d'arrestation de délinquants, baisse de manière significative sans qu'on puisse comprendre pourquoi. Pour les deux chercheurs, il faut relier ces évènements à un autre survenu 19 ans plus tôt, à savoir la légalisation de l'avortement (l'arrêt *Roe vs Wade* prononcé par la Cour Suprême) en 1973. À partir de cette date, nombre de futurs délinquants n'auraient tout simplement pas vu le jour, ce qui expliquerait la baisse des crimes et des délits vingt ans plus tard.

Ce type d'analyse a été fortement contesté. Sur la méthode tout d'abord (le nombre d'arrestation ne correspond pas forcément au nombre de crimes puisque nombre de crimes restent impunis, d'une part, et parce qu'avant 1973 il y avait de nombreux avortements clandestins, d'autre part). En 2007, Reyes a montré qu'il existait une corrélation statistique entre la baisse des crimes et la prohibition du plomb comme additif dans d'essence, deux évènements qui n'ont rien à voir, ce qui montrerait l'inanité des analyses fondées sur de simples corrélations. Dans le même genre, deux ans plus tard, le *New York Times* a fait sensation en ressortant un graphique portant sur l'année 1949 montrant une corrélation presque parfaite entre le nombre de crises de poliomyélite et la consommation de

crèmes glacées. Fallait-il arrêté de consommer des glaces pour enrayer la polio ? Il s'avère simplement que les crises de poliomyélites surviennent plus particulièrement au cours des mois les plus chauds de l'année, au cours desquels les enfants consomment plus de glaces qu'en hiver …

Pour tenter d'y voir plus clair, la DARPA – l'agence publique américaine qui a donné naissance à internet - a lancé en 2016 une initiative appelée « *Explanable AI* » pour permettre une meilleure explicabilité des méthodes de *Deep Learning*, l'explicabilité d'un algorithme consistant à comprendre son processus son fonctionnement. Comme on l'a vu, dans l'IA symboliste, on a mis au départ des règles implicites, alors que dans l'IA connexionniste, on demande à la machine de dégager des règles implicites. Il s'agit par exemple de décomposer les étapes qui mènent un réseau de neurones à reconnaître un objet (un poisson, une voiture, etc.) dans une image.

L'approche DARPA repose sur trois piliers : la création de modèles de *Deep learning* explicables avec des couches intermédiaires utilisant des formes labellisables, des modèles interprétables avec des notions de cause à effet pour permettre la compréhension des mécanismes cognitifs des algorithmes et la production d'interfaces utilisateur plus simples. Parallèlement des acteurs de l'informatique ont mis au point des outils pour créer de l'IA explicable (Googlewhat if, Syncnet ou Psibernetics). De son côté, en France, l'INRIA a lancé en 2017 la plateforme Transalgo, destinée à évaluer la transparence des algorithmes, mais il ne s'agit que de tests a posteriori et de « rétro-ingénierie » des algorithmes. Plus prometteur, Thales s'est lancé en 2019 dans la création d'outils certifiables, en combinant *machine learning* et système de règles pour une Intelligence Artificielle explicable et maîtrisée, indispensable pour la gestion des systèmes critiques tels que les centrales nucléaires ou les systèmes d'armes. Une autre piste est de travailler sur l'interprétabilité, c'est-à-dire le lien entre les solutions (données de sorties) et les données d'entraînement, à savoir quelles sont les données qui ont influencé le résultat.

Les choix à opérer ne sont pas simples car il y a forcément un arbitrage entre explicabilité et sophistication des algorithmes. En théorie, une Intelligence Artificielle non explicable utilisée seule n'est pas compatible avec l'article 63 du RGPD. Mais dans certains cas, les modèles sont suffisamment opaques pour empêcher toute explication des résultats. C'est d'ailleurs ce qu'a plaidé Mark Zuckenberg, le président de Facebook, devant le Sénat américain.

L'amélioration des outils d'analyse informatique ne dispense cependant pas d'avoir recours à un jugement humain pour, en dernière analyse, évaluer l'existence ou non de liens de causalité. En effet, ainsi que l'ont bien montré Casude et Longo (2016), plus on a recours à une masse importante de données, plus grand est le risque de voir apparaître des corrélations non pertinentes, voire totalement loufoques, à l'instar de celles répertoriées sur le site www.tylervigen.com, comme la relation entre la consommation de margarine et le taux de divorce ou entre la production des centrales nucléaires américaines et le nombre de noyades dans les piscines ...

E. La cybersécurité : au cœur de tous les dangers

Alors que l'informatisation, jusqu'à la fin du siècle dernier, se déployait pour l'essentiel à l'intérieur d'une entreprise ou d'une institution, l'apparition d'internet a changé la donne. Les entreprises sont désormais connectées à leurs clients et leurs fournisseurs. Nombre d'institutions publiques, qu'il s'agisse des hôpitaux ou des universités fonctionnent en réseau. Les logiciels ne sont plus installés sur les postes de travail des utilisateurs, mais dans des centres de données accessibles à distance (*cloud computing*). Dès lors, il devient tentant pour les pirates informatiques, les fameux « *hackers* » de tenter de pénétrer dans les systèmes à des fins d'espionnage, de vol des données, d'extorsion de fonds (*ransomware*) ou pour altérer les données ou les algorithmes. Ils utilisent pour cela des logiciels appelés « virus » ou « *malware* ».

Les systèmes d'information deviennent de plus en plus intégrés et ne font plus la différence entre l'intérieur et l'extérieur de l'entreprise. Le risque va aller croissant avec le recours aux capteurs

connectés pour centraliser des données d'exploitation dans les entreprises industrielles (Industrie 4.0). Les fournisseurs d'objets connectés ne sont, la plupart du temps, pas des entreprises dont le métier de base est l'informatique. Qui plus est, les objets connectés sont de plus en plus bon marché, ce qui permet de les déployer massivement à un coût raisonnable. Mais de ce fait, ils n'ont pas de capacité de calcul autonome pour supporter des éléments logiciels de sécurité embarqués. Ce sont donc des proies toutes trouvées pour les prédateurs.

En outre, la mise à disposition des informations auprès des salariés implique la multiplication des terminaux d'accès et donc la nécessité de les sécuriser. Cela n'est pas toujours simple en termes de gestion de parc, surtout lorsque les salariés se connectent à partir de leurs propres outils informatiques (BYOD) ou utilisent des clés USB.

Les outils d'intelligence artificielle sont d'autant plus fragiles que les utilisateurs leur font une confiance aveugle. Un récent rapport du gouvernement britannique, *Artificial Intelligence and UK National Security*, notait que « *la dépendance croissante à l'Intelligence Artificielle pour les fonctions et services critiques ne va pas seulement créer des incitations pour l'ennemi de s'en prendre aux algorithmes, mais elle accroît aussi les conséquences de telles attaques* ». Les attaques peuvent se concentrer sur le code informatique des algorithmes ou bien sur la falsification des données d'entraînement (*data poisoning*), de manière à compromettre les processus de décision des entreprises ou des États.

Les pirates informatiques, d'ailleurs, se sont mis eux aussi à l'Intelligence Artificielle. Ils scannent par exemple tous les ports informatiques des systèmes cibles à l'aide de robots d'indexation comme *Nmap* ou *Neotrace* pour identifier les adresses IP qui constituent autant de points d'entrée. Certains logiciels disponibles sur le *darknet* permettent d'effectuer une cartographie des personnes susceptibles de détenir des informations sensibles à partir de leurs profils Linkedin (*footprinting*) pour mieux les attaquer ensuite.

1. Un danger aux multiples facettes

La forme la plus répandue des attaques informatiques, au moins jusqu'à ces dernières années, est l'intrusion dans les systèmes informatiques afin d'y dérober des informations, dans un but d'espionnage. Ainsi le programme *Morecowbell* de la NSA est destiné à cartographier les réseaux internes des grandes entreprises concurrentes des groupes américains. On se souvient que lors de l'affaire Alstom, la justice américaine s'était mystérieusement trouvée en possession d'une énorme quantité de mails appartenant à l'entreprise française. Dans un article publié en janvier 2015 dans la revue *Notices*, le Directeur de la recherche de la NSA, Michael Wertheimer, reconnaissait que la NSA n'avait pas hésité à faire la promotion d'EC_DBRG, un algorithme de cryptographie dont les clés de chiffrement comportaient une faille (*backdoor*) utilisable par la NSA. Les pirates peuvent aussi installer sur l'ordinateur infecté des *key loggers* (qui enregistrent toutes les frappes clavier) pour espionner un poste informatique en temps réel. L'autre objectif des intrusions est le vol de données qui sont ensuite revendues sur le *dark net* : le piratage dont a été victime en 2017 l'assureur crédit Equifax, touchait un américain sur deux. Des pirates chinois sont soupçonnés d'être l'origine de cette opération. Le vol et la revente de données, notamment les identifiants de cartes bancaires, serait également un moyen pour le régime de Pyongyang pour se procurer des devises. Sans parler des groupes criminels, bien évidemment actifs sur ce marché.

Les attaques de type déni de service (DDOS) sont dites « distribuées », c'est à dire effectuées à partir de points multiples, dont les pirates ont pris préalablement le contrôle. Ce peut aussi bien être des serveurs que des caméras de surveillance, voire des ampoules électriques connectées.

Apparus vers 2014, les *ransomwares* sont des virus informatiques qui cryptent toutes les données d'une entreprise et exigent une rançon, généralement payable en crypto-monnaie (Bitcoin, Monero) en échange de la clé de décryptage. Concrètement le *malware* peut transformer tous les suffixes des

fichiers .doc ou .xls par exemple en .locky (lock = fermé). En général, la pièce jointe ne contient pas le virus mais elle renvoie sur un serveur extérieur qui pilote l'attaque. Quand le malware ne crypte pas les données, il se contente de bloquer l'accès aux serveurs. Les rançonneurs s'attaquent de préférence aux systèmes critiques (les hôpitaux sont une cible de choix) car les victimes sont plus enclines à payer la rançon. Les pirates tentent d'infecter un grand nombre d'ordinateurs (*spear fishing*) en demandant des rançons de faible montant (de l'ordre de 250 à 1.000 dollars à chaque fois), inférieur aux coûts de récupération des données, et jouent sur l'effet de masse. Deux affaires ont défrayé la chronique, du fait de leur caractère viral et du montant des dégâts occasionnés.

Le 12 mai 2017, des pirates, probablement nord-coréens s'attaquent à des institutions européennes (dont Renault et le National Health Service britannique) à l'aide d'un *ransomware*, baptisé *Wannacry*, qui aurait été dérobé à la NSA. *Wannacry* a infesté 300.000 ordinateurs dans 150 pays en utilisant une faille dans le système d'exploitation Windows XP. A cette occasion, la cour des comptes britanniques a noté que le succès de *Wannacry* était principalement lié à la négligence du NHS en termes de sécurité informatique.

Un mois plus tard survint une autre attaque avec le virus *NotPetya*, dont ont été victimes d'abord des institutions ukrainiennes, puis par contagion de grandes entreprises occidentales comme Saint-Gobain, Mondelez, Merck ou Maersk. *NotPetya* se présentait comme un *ransomware* (mais en fait les fichiers étaient cryptés définitivement, et le paiement de rançon n'a servi à rien). Des hackers russes sont soupçonnés bien que le groupe pétrolier russe Rosneft ait aussi été infecté. Le montant des dommages infligés par *NotPetya* est estimé à 10Md$ (soit deux fois plus que *Wannacry*).

Plus récemment, le *ransomware Maze* a opéré à différents moments de l'année 2019, d'abord aux États-Unis puis en France (attaque contre Bouygues Construction). *Maze* présentait la particularité de demander des rançons élevées (plusieurs millions de dollars) et d'avoir été « loué » sur le darknet (*ransomware as a service*) aux pirates qui l'avaient conçu.

2. Qui sont les pirates ?

Thales estime que la plupart des attaques de niveau élevé dans le monde sont le fait de groupes étatiques ou parrainés par des États. Leurs principales cibles sont les autres États, les institutions financières, le secteur énergétique et les entreprises (surtout les PME, plus fragiles) de la chaîne logistique.

Il est en fait très difficile de déterminer l'origine des attaques informatiques qu'il s'agisse de la simple intrusion, de la pénétration en profondeur (APT : *Advanced Persistent Threat*) ou du déni de service (DDOS). En effet, on trouve sur le *dark net* des kits prêts à l'emploi (« *exploits* ») pour modifier les codes et empêcher le bon fonctionnement des logiciels.

Parmi les techniques les plus utilisées, on peut citer :

- Le dépassement de tampon (*buffer overflow*), pour introduire des instructions dans la pile d'exécution d'un programme et exécuter un code malveillant ;
- Le dépassement de tas (*heap overflow*), pour introduire des données différentes de celles d'origine ;
- Le dépassement d'entier (*integer overflow*), pour modifier la valeur numérique résultant d'une opération mathématique ;
- Le dépassement de pile (*stack overflow*), pour écraser des informations nécessaires au bon déroulement d'un processus.

Il existe également des virus plus spécialement destinés à tirer profit des failles existant dans les systèmes. La manière dont nombre d'éditeurs de logiciels cherchent à mettre leurs produits sur le marché le plus vite possible sans avoir vérifié la qualité du code induit une prolifération de failles dites « *zeroday* » (présentes dans la version initiale et jamais corrigées depuis). Certains malwares

sont d'ailleurs disponibles en mode *cloud* sur des plateformes en accès temporaire (*Malware as a service*) sur le dark web. C'est la forme la plus commerciale du piratage : les descriptifs de certaines failles *zeroday* sont en vente sur le dark web, à des prix qui peuvent atteindre le million de dollars.

C'est d'autant plus facile que MS Windows, par exemple, comprend 50 millions de lignes de code et Android 12 millions. Selon Steve Connell, un programmeur fait entre 10 et 20 erreurs de code (qui ne sont pas forcément des failles) toutes les 100 lignes. De ce fait une application mobile comporte facilement une dizaine de failles.
Il y a aussi des failles sur processeurs eux-mêmes, comme Meltdown (qui a pu être réparé par des patches de sécurité) ou Spectre (difficilement réparable car les failles sont liées aux méthodes de fabrication des puces).
La disponibilité des outils de piratage sur le *dark web* a donc considérablement réduit les barrières à l'entrée pour les hackers mais rend aussi plus difficile la recherche de l'origine réelle des attaques. De plus, certains hackers vont jusqu'à pénétrer dans des systèmes déjà attaqués par d'autres pour récupérer les malwares qui y ont été déposés.

Citons quelques-uns des principaux suspects :

L'Ukraine et la Chine abriteraient les principaux repères de pirates, dont les noms sont associés aux virus dont ils sont les auteurs. De fortes présomptions existent quant à l'origine russe de certains groupes tels que *NotPetya* ou APT28 (alias *Pawnstorm*, alias *Fancy Bear*) en raison de la sophistication des procédés employés et du type de cible. Mais aucune preuve formelle n'a pu être apportée en raison des techniques utilisées par les pirates (génération dynamique d'adresses IP, utilisation de serveurs zombies, etc.). Un gang de hackers russes (*Lurk*) a réussi à dérober en 2011 plus de 40 millions de dollars à la principale banque du pays, la Sberbank.

Les premières attaques informatiques contre le secteur énergétique ukrainien en décembre 2015 ont été attribuées à des hackers parrainés par les services secrets russes, de même que le virus *Petya* (prédécesseur de *NotPetya*) début 2016. Il est clair que, depuis 2013, le ministère de la Défense russe a lancé une vaste campagne de recrutement de hackers, dans le cadre de la « doctrine Guerasimov » qui théorise la guerre hybride (c'est aussi le cas de la NSA qui tient chaque année un *Summer Camp* à destination de futurs hackers). Ce qui rend les choses compliquées, c'est que *NotPetya*, comme *Wannacry*, aurait utilisé des milliers de fichiers de la NSA, dont le *malware EthernalBlue*, qui ont été dérobés début 2017 par un groupe de hackers indépendants connu sous le nom de *Shadow Brokers*.

Les hackers chinois, comme les groupes APT3 et surtout APT 10 qui vise les fournisseurs de services d'infogérance (et donc les données des entreprises clientes), sont réputés être tous aux ordres du gouvernement de Pékin. Ils sont accusés principalement d'opérations d'espionnage et de vol de données. Le groupe APT *Naikon* est de son côté accusé d'attaques contre les systèmes du gouvernement australien en 2019 à l'aide d'un virus nommé *Ariabody*.
Les spécialistes de la sécurité informatique attribuent au groupe nord-coréen *Lazarus* (alias APT 38) le vol massif de données effectué chez Sony en 2014 et la cyberattaque de la Banque Centrale du Bangladesh en 2016, mais ils seraient surtout derrière le malware *Wannacry*, qui aurait causé 5 milliards de dollars de pertes en 2017. Leur fonction est de ramener des devises par tous les moyens. Plus récemment ils auraient concentré leurs attaques contre des grands sites de e-commerce, toujours pour dérober les coordonnées bancaires des clients, qui sont ensuite échangées contre des crypto-monnaies sur le *dark net*.

Les États-Unis ne sont pas en reste, comme l'a révélé l'affaire Snowden. Selon Wikileaks, la CIA aurait même mis au point un logiciel nommé *Hammerdrill 2.0* destiné à pénétrer un ordinateur coupé de tout réseau informatique. Pour mieux préparer ses intrusions, la CIA a constitué une bibliothèque de failles de sécurité, notamment failles z*erodays* jamais révélées. Selon Symantec, le groupe de hackers *Longhorn* serait probablement lié à la CIA. Le programme *Quantum* de cyber-

espionnage de la National Security Agency s'appuie notamment sur l'agence de renseignement suédoise FRA, équivalent de la NSA. *Quantum* a pénétré le siège de l'OPEC, l'opérateur belge Belgacom, la Commission et le Parlement européens, dans le cadre de l'opération *Winterlight*, associant la NSA, la FRA suédoise et le GCHQ britannique. À cette occasion, ils ont établi la cartographie des systèmes d'information des cibles, introduit des malwares et contrefait certains profils Linkedin.

Début 2019, selon le *New York Times*, les USA auraient infiltré de manière visible les systèmes d'information du réseau électrique russe en guise d'avertissement après l'affaire *NotPetya*. Il ne faut bien entendu pas oublier les hackers indépendants : on voit des individus isolés capables de lancer des attaques massives comme les Britanniques *Bestbuy* ou *Spi(d)ir* alias Daniel Kaye. Ce dernier s'était attaqué à *Lonestar*, l'opérateur historique de télécom au Libéria et à des banques britanniques comme Barclays ou Lloyds. Plus récemment, on a assisté à une multiplication des attaques par ransomware perpétrés par des groupes criminels sur un nombre limité de cibles de grande importance (*Big Game Hunting*), avec un chiffre d'affaires estimé à plus de 2 milliards de dollars à ce jour.

F. La menace pour la civilisation européenne

Les États Unis ont été le premier grand pays à formaliser leur stratégie en matière d'Intelligence Artificielle avec le *National Artificial Intelligence Research and Development Strategic Plan* d'octobre 2016. Ce document reste encore très général quant aux orientations à prendre.

Il faudra attendre l'Executive Order (Décret) N°13859 signé par Donald Trump en février 2019, intitulé « *Maintain American Leadership in Artificial Intelligence* », pour que soient définies des priorités en matière de recherche sous l'autorité du *National Science and Technology Council* et que soit décidée l'ouverture des données publiques gérées par les agences relevant de l'Etat fédéral. En août 2020, l'administration annonce d'ailleurs une enveloppe d'un milliard de dollars supplémentaires sur 5 ans pour l'intelligence artificielle et l'informatique quantique, dont un tiers financé par le secteur privé. Cette enveloppe sera principalement destinée à la création de 12 centres de recherche.

Mais c'est surtout la note de la Harvard Kennedy School (l'équivalent américain – privé - de l'ENA) intitulée *AI for all*, publiée en 2017, qui explicite la philosophie américaine en matière d'Intelligence Artificielle. Elle met en avant la gouvernance par les algorithmes, qui optimiserait les décisions politiques, ce qui revient à verser dans le « solutionnisme numérique » dénoncé par Evgueny Morozov. Plus encore que l'État fédéral, qui s'intéresse en priorité aux applications duales civiles-militaires, ce sont les collectivités locales qui vont porter avec des partenaires privés des projets (*smart cities*, police prédictive, etc.) où le politique se dissout dans la technologie.

1. Le poids des GAFA est devenu insupportable

Aux USA, alors que la recherche fondamentale a été portée par le gouvernement et les universités, la recherche appliquée en Intelligence Artificielle est désormais principalement le fait des acteurs privés. Celle-ci n'est pas toujours traçable, car nombre de recherches privées ne font pas l'objet d'une publication.

La montée en puissance de l'Intelligence Artificielle se conjugue avec celle de l'économie des plateformes comme Facebook, Google ou Amazon, dont la capacité à préempter une technologie génère des rentes d'oligopole voire de monopole (*winner takes all*). Le cas d'Amazon est emblématique : entre 1999, son chiffre d'affaires propre est passé de 1.5 milliard de dollars à 117 milliards, mais le chiffre d'affaires pour compte de tiers (la *market-place*) est passé dans le même temps de zéro à 160 milliards de dollars. La force des plateformes, c'est aussi de pouvoir agréger tout un écosystème autour d'elles, tout en restant le centre (par le biais d'interface type API). Pour

caricaturer, je crois être sur Facebook, mais mes données sont aussi récupérées par Uber et AirBnB, sachant que 80% des visites sur les 25 sites les plus connus dans le monde vont vers des acteurs américains. Et aux États-Unis, 66% des acheteurs en ligne commencent leurs recherches sur Amazon avant d'aller éventuellement visiter d'autres sites.

Si les GAFAM, sont américains, il en va de même des grandes plateformes d'intelligence artificielle : Watson (IBM), Nervana (Intel), Einstein (Salesforce), Alexa (Google). Google a racheté Deepmind et Waze. Intel fournit plus de 95% des serveurs de datacenters dédiés à l'intelligence artificielle, tandis que Nvidia a une position prépondérante dans les cartes graphiques (GPU).

La France ne dispose plus que d'un acteur majeur dans le domaine du logiciel, Dassault Systèmes, les autres éditeurs sont de taille petite ou moyenne. Valeo pourra, on l'espère, compter parmi les acteurs structurants de la voiture autonome, mais l'entreprise française n'est pas seule dans la course.

Les GAFAM constituent un « oligopole global dissymétrique » marqué par l'accumulation de la puissance et de la richesse aux mains de quelques-uns. Cette toute-puissance se dissimule volontiers derrières des préoccupations éthiques. Bill Gates a ainsi exprimé en 2015 son inquiétude sur les progrès de l'Intelligence artificielle, mais Microsoft (dont il est resté administrateur jusqu'en 2020) n'a pas, tant s'en faut, mis fin à ses développements dans ce domaine.

Les acteurs anglo-saxons, en particulier américains, ont multiplié les participations aux forums et *think tanks* se réclamant d'une vision éthique de l'IA, tout en s'efforçant de lutter contre toute réglementation et en agissant au mépris de la vie privée : la meilleure façon d'être du côté du Bien n'est-elle pas de définir soi-même la frontière qui le sépare du côté obscur ?

Des pays comme la France sont-ils en passe de devenir des « colonies numériques » des plateformes américaines ? Le rapport Villani déplore que de grandes entreprises sous-traitent la gestion de leurs données aux GAFAM. Renault, Telefonica ou Carrefour coopèrent ainsi avec les GAFAM. Mais l'État ne donne pas le bon exemple. Ainsi Microsoft « met à disposition » Office 365 dans les écoles et forme les professeurs dans la cadre de la convention signée avec l'Éducation Nationale en 2015. De même, il avait dans un premier temps été décidé que les données du *Health Data Hub* du Ministère de la Santé seraient hébergées sur les serveurs de Microsoft France, tandis que la DGSI est très dépendante des algorithmes de Palantir.

2. Une menace chinoise ?

Sur les principales 4500 sociétés opérant dans le domaine de l'IA, 50% sont américaines et 30% chinoises. Ce qui laisse peu de place aux autres pays, et notamment aux Européens. Les sommes dévolues à l'IA en Chine sont colossales par rapport aux budgets des pays européens, et même supérieures à celui des USA. Alors que France IA est doté d'un milliard et demi d'euros sur 5 ans, la Chine a dépensé une somme huit fois supérieure sur la période 2016-2019. Fin 2018, 45 % des brevets liés à l'IA ont été déposés aux États-Unis et presqu'autant (40%) en Chine selon Wipo, l'organisme mondial de gestion des brevets. Dans le classement *CB insight 2020* qui identifie les 100 start-up les plus prometteuses en matière d'IA, seulement 2 sont localisées dans l'Union européenne et 6 au Royaume-Uni. Et sur les 30 principaux déposeurs de brevets sur l'AI, seules 4 sont européens.

La première formalisation d'ensemble des ambitions chinoises en matière d'IA est contenue dans le Plan « *Made in China* » publié en 2015. L'année suivante, les médias chinois présentent la défaite du joueur de go Lee Sedol contre le logiciel AlphaGo de Google comme leur « moment Spoutnik », par référence au choc qui avait frappé les décideurs américains en 1957. En septembre 2016, le Ministère de l'Industrie publie ses « Modalités de mise en œuvre du projet de fabrication intelligente en Chine 2016-2020 ». L'académie des sciences chinoise a lancé dans la foulée un projet comparable au *Human Brain Project* européen ou à la *Brain Initiative* américaine.

En 2017, la Chine publie un « Plan de développement de la prochaine génération d'intelligence artificielle 2016-2030 : l'objectif des Chinois est d'atteindre une (voire *la*) position de leader mondial de l'IA à l'horizon 2030, c'est-à-dire dans moins de dix ans. En 3 étapes :

2020 : réalisation d'une découverte majeure dans la théorie de l'IA ;

2025 : l'IA doit devenir le principal moteur de la croissance chinoise, avec un objectif centré sur la santé, l'agriculture, les villes intelligentes et, bien entendu, la défense ;

2030 : la Chine devient leader mondial de l'Intelligence Artificielle.

Parallèlement, la Chine déploie des efforts sans précédent dans le domaine des infrastructures numérique et en particulier les semi-conducteurs. La Chine cherche à combler son retard dans les semi-conducteurs, surtout après les menaces d'embargo américaines, avec des acteurs significatifs comme SMAC ou Cambricon.
En Chine ce sont les grands groupes (notamment les BATX : Baidu, AliBaba, Tencent, Xiaomi) qui portent une grande part de la recherche appliquée en Intelligence Artificielle, en liaison étroite avec les autorités. Baidu, le Google chinois, a dépassé son concurrent américain en matière de reconnaissance d'images et se lance dans les véhicules autonomes. Avec IFytek, la Chine dispose d'un leader de la reconnaissance vocale. L'empire du milieu a également pris des positions de leader dans les calculateurs à haute puissance. Sunway Taiku, fabriqué uniquement avec des composants chinois est le troisième plus gros ordinateur du monde.

Le développement des *smart cities* chinoises repose sur la gestion du trafic, mais aussi la surveillance de la population (plus de 200 millions de caméras de télésurveillance couplées aux systèmes de reconnaissance faciale développés par le groupe Cloudwalk). Une personne qui ne traverse pas dans les clous voit son visage s'afficher sur écran géant jusqu'à ce qu'elle ait payé son amende. Une surveillance qui va jusqu'à la quantité de papier consommé dans les toilettes publiques …
Avec le système appelé « Crédit Social », les « mauvais citoyens » peuvent se voir privés d'accès aux moyens de transport à longue distance ou aux meilleures écoles par exemple.

3. Les particularités du système chinois

La Chine dispose d'un triple avantage :

- Elle dispose d'une masse considérable de données grâce à sa population nombreuse, mais aussi du fait d'une culture de recueil de données dans les entreprises de commerce électronique ou de messagerie. À titre d'exemple, les deux principaux réseaux sociaux chinois, *Weibo* et *Wechat*, rassemblent 800 millions d'utilisateurs. À cela s'ajoutent toutes les données liées aux paiements mobiles, car en Chine la moitié des smartphones sont dotés d'une application de paiement mobile. La crise du COVID a encore renforcé la part des paiements faits sur mobile, et donc le volume de données concernant les utilisateurs.
- Une coopération intense entre les géants privés (les BATX mais pas seulement) et l'appareil politico-militaire. D'où la polémique autour de Huawei, qui détient 25% du marché mondial des équipements télécom et est également devenu début 2020 le premier vendeur de téléphones mobiles. Les licornes d'IA chinoises comme Mojvii, Sensetime, YiTu ou Cloudwalk s'internationalisent rapidement dans le cadre du volet numérique des Routes de la Soie. Elles bénéficient pour se développer d'importantes commandes publiques.
- Elle n'a aucun état d'âme concernant la vie privée et dans la politique de collecte et d'exploitation des données (y compris la reconnaissance faciale). Tout est possible, pourvu qu'on ne mette pas en cause le pouvoir de l'État-Parti. On notera dans un autre domaine que les trois savants chinois auteurs de manipulations génétiques sur des fœtus n'ont écopé – et encore, après la médiatisation de l'affaire – que de trois ans de prison, peine dont on ne sait si elle sera vraiment effectuée. La Chine privilégie l'efficacité sur l'éthique (au sens occidental). C'est une société holistique, d'autant qu'après 40 ans de relatif libéralisme, la

Chine semble revenir à un « néo-maoïsme » qui répondait à certaines attentes de la société chinoise, comme en témoigne le roman *Dragon Bleu, Tigre Blanc* de Qiu Xiaolong. Le système de Crédit Social des individus appliqué en Chine, c'est Big Brother pour les Occidentaux, mais les Chinois semblent s'y résigner. Et ceux qui ne seraient pas d'accord n'ont pas – il est vrai - le droit à la parole.

Tout cela conduit à un très grand optimisme : En 2017 des chercheurs de Yale et d'Oxford ont interrogé 352 « sachants » américains et asiatiques sur les perspectives de l'Intelligence Artificielle. Il en résulte que la probabilité que l'IA dépasse l'humain dans toutes les tâches (singularité) était de 50% à horizon 2090 pour les Américains mais seulement 2047 pour les Asiatiques. Rappelons que la plupart des cadres dirigeants du PC chinois ont un bagage d'ingénieur. Ils sont donc pro-technologie.

L'étude Life de l'agence de publicité BETC portant sur la question « pensez-vous que l'IA fera progresser l'humanité ? » a montré le fossé qui sépare la Chine (très optimiste) des pays européens (beaucoup plus réservés). Les Chinois seraient prêts sans état d'âmes à sacrifier les libertés (qu'ils n'ont d'ailleurs pas) en échange de plus de performance technologique, et donc économique.

4. Les thèses de Kai Fu Lee

Dans ce contexte, il est particulièrement intéressant de suivre les thèses développées par Kai Fu Lee, un chinois de Taiwan qui, après avoir dirigé la filiale de Google en Chine, y a créé un fonds d'investissement dédié à l'Intelligence Artificielle. Pour Lee, rejoignant l'analyse d'un Olivier Ezratty en France, le temps n'est plus à la course aux armements en termes de grandes découvertes, il s'agit surtout d'être le premier à appliquer les derniers développements de l'Intelligence Artificielle dans un maximum de situations. Rien ne sert de courir après une illusoire « IA forte ».

« *Alors que s'ouvre l'ère de l'IA appliquée*, dit-il, *l'avantage de la Silicon Valley en termes d'expertise et d'excellence a fait son temps »*. Pour que les choses soient bien claires, il ajoute que, dès lors que « *le plus gros du travail théorique a été accompli, il est temps (pour les entrepreneurs) de retrousser leurs manches et de s'atteler aux basses œuvres - autrement dit transformer les algorithmes en entreprises durables* ».

Selon cette thèse, le champ de bataille de l'IA n'est plus exclusivement à la frontière technologique, dans les laboratoires de recherche. Nous sommes passés de l'ère des compétences à l'ère des données. L'IA n'est pas tant une technologie nouvelle qu'un assemblage de process incluant l'analyse des données, des algorithmes de calcul et des briques logicielles. La Chine dispose en quantité de ce qui va lui permettre de devenir un leader mondial : des données en très grande quantité, de grandes capacités de calcul et la disponibilité d'ingénieurs qui n'ont pas besoin d'être des génies mais qui maîtrisent la capacité à assembler intelligemment des briques logicielles disponibles en open source sur des places de marché de modèles d'IA pour les développeurs comme ONNX, MLIR ou Algorithmia.

On retrouve ici le schéma évoqué par Clayton Christensen dans son désormais classique *The Innovator's Dilemma* : dans le processus d'innovation, la première manche appartient aux scientifiques, mais la seconde est le domaine des « fermiers », ceux qui savent exploiter cette innovation pour en faire un produit de masse. Kai Fu Lee parle d' « innovation de masse », une expression qui combinerait les Cent Fleurs maoïstes et la « théorie du bordel ambiant » chère à Roland Moreno …

Et c'est justement, selon l'auteur, ce qui fait la force de la Chine : Les algorithmes et la puissance de calcul sont disponibles pour tous, et la Chine regorge d'entrepreneurs, d'ingénieurs de bon niveau et surtout de données qui vont lui permettre de surpasser ses concurrents : « *Dans le domaine du deep learning (…) dès lors qu'on lui procure de plus grosses masses de données, un*

algorithme conçu par des ingénieurs de niveau moyen se révèle souvent plus performant que son concurrent, pourtant mis au point par un scientifique de renommée mondiale » affirme-t-il.

Cette stratégie trouve son application dans le 13e plan quinquennal chinois, structuré autour de 15 grands projets d'innovation structurants, dont quatre dédiés à l'IA (recherche sur le cerveau, bigdata, réseaux intelligents, smart factory). La stratégie chinoise est très centrée sur les applications concrètes. Il s'agit d'utiliser au maximum les acquis théorique des occidentaux et développer en Chine des applications de reconnaissance d'images, la voiture autonome, etc. Le R (Recherche) est largement importé pour permettre une hégémonie dans le D (Développement).

5. Les limites du modèle

La Chine, de son propre aveu, est moins avancée que les USA sur la recherche fondamentale en Intelligence Artificielle. Certes, les Chinois sont leader dans les publications sur le *deep learning* mais l'Europe et les USA disposent de beaucoup plus de chercheurs de haut niveau en IA que la Chine (5000 pour chacun contre 1000 pour la Chine). Et, comme on l'a vu précédemment, l'histoire de la recherche n'est pas terminée.

Le fin mot de l'histoire est que la Chine développe en partie, sans trop s'en cacher, une « stratégie du coucou » en copiant les innovations occidentales. Un des points soulevés par Acemoglu et Restrepo est que l'Intelligence Artificielle peut servir au *reverse engineering* des innovations de la concurrence. On se souvient que Samsung avait ainsi copié les smartphones d'Apple. Pour un pays comme la Chine qui fait beaucoup de copie et ne respecte pas la propriété intellectuelle, ce peut être une arme redoutable. Les Chinois ont la réputation –justifiée - de copier tout ce qu'ils peuvent, mais si l'on regarde le succès des frères Sammler en Allemagne, dont le groupe Rocket Internet a copié nombre de sites de e-commerce américains, les Européens sont capables d'en faire autant.

Cependant, rien ne prouve que les algorithmes (développés en Chine ou adaptés à partir d'outils disponibles en open source), qui sont entraînés sur des millions de Chinois, puissent être utilisés à l'identique aux États-Unis ou en Europe. On retrouve le biais d'extrapolation bien connu des chercheurs.

Par ailleurs, la Chine a bâti son modèle sur l'abondance de données. L'IA chinoise serait dans une logique d'imitation intelligente plutôt que d'innovation. Du coup, selon Osborne et Frey (2020), elle risque de ne pas être très performante sur les domaines où la donnée est spécifique ou peu nombreuse. Elle court le risque de ne pouvoir produire qu'une IA rudimentaire. Plutôt que les Cent Fleurs, elle pourrait connaître à nouveau l'échec du Grand Bond en Avant.

À moins qu'au contraire, déjouant les prophéties pessimistes, les acteurs chinois, comme cela s'est produit dans les énergies renouvelables, ne parviennent à remonter ensuite la chaîne de valeur en utilisant les profits générés par cette Intelligence Artificielle moins sophistiquée, mais efficace, pour développer ensuite des outils de plus en plus performants.

V. Que faire ?

A. Les stratégies nationales comme France IA et le rapport Villani

Alors que les gouvernements américain et chinois ont largement communiqué sur leur stratégie en matière d'intelligence artificielle et d'apprentissage automatique, le secrétariat d'État à l'industrie, au numérique et à l'Innovation lance en janvier 2017 l'Initiative France IA qui se concrétise deux mois plus tard, en mars 2017, par un Plan pour l'Intelligence Artificielle. Celui-ci prend acte de l'extrême fragmentation du paysage français, entre les différents centres de recherches, entre les nombreuses initiatives privées – émanant tant des start-up que des grands groupes – et entre le monde de la recherche et celui de l'entreprise, la recherche publique portant, contrairement aux États-Unis, l'essentiel de l'effort.

Un an plus tard, la nouvelle majorité, à vrai dire en partie issue de l'ancienne, décide de s'attaquer au problème des priorités afin d'éviter que le foisonnement d'initiatives empêche de dégager un minimum de visibilité stratégique, comme cela avait été le cas avec le saupoudrage des crédits entre les pôles de compétitivité et les 37 priorités industrielles d'Arnaud Montebourg. La mission coordonnée par le député LREM Cédric Villani, qui rend son rapport en mars 2018, définit quatre secteurs prioritaires : santé, environnement, transport-mobilité et défense-sécurité. La mesure-phare est la création, en réponse aux préoccupations exprimées dans le rapport France IA, de plusieurs Instituts Interdisciplinaires d'Intelligence Artificielle (3IA), avec trois missions : créer des conditions d'environnement de recherche attractives, diffuser une formation scientifique de haut niveau et fluidifier les interfaces recherche-industrie.

Les universitaires seront répartis en chercheurs statutaires (affiliés), compagnons (*fellows*) à plein temps, associés à temps partiel et chercheurs invités (3 à 12 mois). Sur le modèle israélien, les entreprises sont invitées à contribuer financièrement en échange de l'accès aux travaux de recherche. D'ores et déjà, des structures ont été créées à Toulouse, Grenoble, Nice et Paris. Les 3IA ont vocation à être des « bacs à sable » c'est-à-dire des lieux d'expérimentation supervisés dégagés des contraintes ordinaires (réglementation, grilles salariales, etc.). Il s'agit de constituer des « zones franches de l'IA » avec des allègements de formalités administratives, salaire hors grille fonction publique.

Parallèlement, Emmanuel Macron annonce une enveloppe de 1,5 milliard d'euros pour l'Intelligence Artificielle, dont un peu plus de 40% est dédié à la recherche. Le solde est notamment consacré à 6 projets pilotes centrés sur des applications concrètes de l'Intelligence Artificielle (preuves de concept ou « POC » en franglais): détection d'occupation irrégulière des sols à partir d'images satellitaires, amélioration des contrôles environnementaux, lutte contre les risques sanitaires dans la restauration, optimisation de la sécurité nucléaire, diminution des risques post-opératoires à l'hôpital, robot vocal pour les utilisateurs de chèques emplois associatifs.

Certains observateurs ont dénoncé le manque d'ambition de ce plan. Pour eux, à lire le rapport Villani on pense que pour un marteau, tous les problèmes ont la forme de clou. On peut critiquer son élitisme et sa focalisation sur la dimension universitaire. Il faut penser l'ensemble de la chaîne de valeur de l'économie de la donnée, et pas seulement le haut comme le fait le rapport.

Comme le reconnaissaient les rédacteurs du rapport, il y a un fossé en France entre la recherche et les applications. La situation de la filière robotique a de quoi donner à réfléchir : alors que la France est très bien placée en matière de recherche, ce secteur n'emploie que 7.000 personnes dans notre pays (2.3% du total mondial). Le marché est dominé par le Japon, les États-Unis, la Corée et l'Allemagne. Une bonne partie des entreprises robotique françaises sont des filiales de groupes étrangers (ABB, Adept, Comau, Yaskawa, Europlacer, Staübli) et il y plus d'intégrateurs que de

fabricants. Et pourtant, les entreprises européennes de robotiques (notamment allemandes) ont un tiers du marché mondial et la robotique est bien plus répandue en Allemagne, en Suède, en Belgique ou au Danemark que dans notre pays.

La critique est en partie fondée, mais il faut à l'inverse lui reconnaître la volonté de définir des priorités. Le rapport de 2018 est une brique stratégique, pas une stratégie. Il faudra à l'avenir le compléter, et notamment dans deux domaines qui en constituent les angles morts : l'interface entre la recherche et l'entreprise (même si le rapport évoque la mise en place de « plateformes sectorielles de mutualisation » accessibles aux chercheurs et aux entreprises), et l'ajout d'une véritable stratégie industrielle sur le modèle allemand.

B. Les coopérations européennes, subsidiarité

1. La situation de l'Europe

La diffusion de l'Intelligence Artificielle dans les entreprises est moins importante en Europe qu'aux États-Unis. Beaucoup de projets sont encore à l'état de pilote. De manière générale, selon le cabinet McKinsey, les grands groupes européens investissent moins dans le numérique que leurs homologues américains. De fait, fin 2017, le secteur numérique ne représentait 1,7% du PIB en Europe contre 3.3% aux USA et 2.2% en Chine. L'Europe est en retard dans l'internet des objets, l'internet mobile et le Big Data, mais dispose semble-t-il d'une légère avance sur d'automatisation des processus de gestion (RPA).

Pourtant, notre continent ne manque pas d'atouts :

Il y a plus de développeurs en Europe (près de 5 millions) qu'aux USA et il s'est publié dans les pays de L'Union européenne depuis 2008 autant d'articles sur l'Intelligence Artificielle dans des revues à comité de lecture que pour les USA et la Chine réunis. L'Europe a une très bonne expertise en IA symbolique de par sa culture rationnelle, ce qui la place en bonne position pour développer la voiture autonome, qui repose largement sur des moteurs de règles.

Les Européens conservent environ 10% du marché des semi-conducteurs (contre 50% pour les USA et seulement 5% pour la Chine), avec Infineon (très présent dans l'automobile) et STMicro (centré sur les télécoms), mais ils sont quasi absents du marché des composants dédiés à l'intelligence artificielle, aujourd'hui dominé par les Américains (Advanced Micro Devices, Nvidia). En revanche, les Français ont développé d'excellentes compétences en architecture, pour optimiser la capacité de traitement des composants.

2. La stratégie de l'UE

Faut-il attendre notre salut de l'Union Européenne ? Dans un premier temps, le parlement européen a adopté une attitude très restrictive, ce qui a poussé la Commission, appuyée par les gouvernements des États-membres, à reprendre la main. Le programme Horizon 2020, piloté par la DG Recherche et Innovation, a mobilisé 80 milliards d'euros sur 7 ans (dont 17 milliards pour les technologies de pointe), mais pas de ligne spécifique à l'Intelligence Artificielle. Le budget 2021-2028 a vocation à atteindre 120 milliards d'euros. 50% des 13 milliards du Fonds européen de Défense (2021-2027) aurait dû être alloués aux innovations de rupture et notamment à l'IA, mais celui-ci a été fortement réduit dans le cadre des mesures de relances post-COVID. Par comparaison, la Chine a alloué 10 milliards de dollars rien que pour l'informatique quantique et 60 milliards rien que pour l'IA entre 2017 et 2025.

C'est en prélude du nouveau budget que l'UE a publié en février 2020 trois documents stratégiques : le Livre blanc sur l'Intelligence artificielle, la Stratégie européenne en matière de

données et le manifeste « *Shaping Europe's digital future* ». Il s'agit d'investir 20 milliards d'euros dans l'IA chaque année dans les secteurs définis comme prioritaires : la santé, les transports, la sécurité et l'énergie. Par comparaison, Google a investi 24 milliards d'euros en R&D en 2019 et Alibaba 6 milliards d'euros. Les fonds seront investis dans 70 projets cofinancés avec les budgets nationaux et des investisseurs privés.

La démarche de l'UE sur le marché unique numérique (*Digital Single Market*) doit s'appuyer sur un marché unique des données. L'Europe a perdu la bataille des données personnelles, elle veut gagner celle des données publiques (open data) et l'exploitation des données des entreprises, qui bénéficieront de la 5G et de l'internet des objets. Reste à savoir si cela ne va pas constituer une aubaine pour les GAFAM, qui sont déjà au cœur de la place ? Il existe pourtant certains grands acteurs européens présents dans l'IA : Capgemini, SAP, ATOS, mais aussi les équipementiers automobiles qui visent à devenir des plateformes au sein de leur écosystème.

L'Europe n'a pas vraiment l'équivalent de la DARPA, et le projet d'Agence Européenne des Innovations de rupture préconisé par le rapport Hennion-Auconie d'octobre ne verra pas le jour au mieux avant 2021. L'Europe saura-t-elle dépasser l'approche purement déclamatoire de la stratégie de Lisbonne il y a vingt ans et fixer des objectifs plus normatifs et concrets ? Le problème de la stratégie européenne est qu'elle repose sur le partage des données entre les entreprises, mais comme le reconnaît Thierry breton, celles-ci sont très réticentes. Les constructeurs automobiles allemands semblent vouloir réserver les données à leur stratégie de plateforme Industrie 4.0 et ne sont pas prêts à en faire profiter d'autres acteurs.

3. Coopération bilatérale: le Royaume-Uni plutôt que l'Allemagne ?

De son côté, le rapport Villani propose de consolider d'abord les écosystèmes nationaux de données avant d'envisager des collaborations européennes. On en peut que se réjouir de cette approche pragmatique.

En Europe, le Royaume-Uni et les pays scandinaves sont probablement les plus avancés en matière d'IA, mais la France n'est pas très loin. Le cabinet allemand Roland Berger a mis au point un index de financement des start up spécialisées en IA, basé sur le nombre de start-up pour 10.000 habitants. L'indice était en 2019 de 0.9% pour le Royaume-Uni (590 start up), 0,4% pour la France (236), 0,3 pour l'Allemagne (218) mais 2,1% pour Israël (189). Les levées de fonds en IA sont bien plus importantes au Royaume-Uni qu'en France mais l'écart diminue. Sur la période 2009-2019 la moyenne des levées de fonds était de 4,5 millions de dollars pour la France contre 2,5 millions pour le Royaume-Uni.

Plus que l'Allemagne, qui a centré ses efforts sur son propre projet *Industrie 4.0,* le partenaire naturel de la France est donc probablement le Royaume-Uni, qui a mené une réflexion stratégique sur le déploiement de l'IA sous l'égide de la Royal Society et l'Institut Alan Turing. Reste à savoir si nos voisins d'outre-Manche sont vraiment demandeurs …

C. L'Intelligence Artificielle est-elle vraiment intelligente ? Poser les limites pour trouver les bons usages de la machine

Comme le rappelait Margaret Vestager, vice-présidente de la Commission Européenne, en introduction du Livre Blanc, « *l'IA n'est ni bonne ni mauvaise intrinsèquement. Tout dépend de pourquoi et comment elle est utilisé*e ». L'apprentissage automatique, dont le potentiel est, on l'a vu, avéré – mais pas infini – a généré bien des phantasmes, dans un sens comme dans l'autre. On est passé insidieusement de systèmes dits d' « aide à la décision » à l'illusion que la machine pourrait prendre les décisions seules, puisqu'elle serait plus « intelligente que les humains.

N'oublions pas, cependant, que ce que nous appelons de manière un peu rapide « intelligence » des machines repose sur la conjugaison d'une très grande capacité de mémoire, permettant de traiter de grandes quantités de données, et de la vitesse de traitement. Comme le rappelle opportunément Jean-Louis Dessalles, la machine ne « raisonne » pas, elle calcule, ce qui n'est pas la même chose. La manière dont nous organisons les connaissances dans notre cerveau nous permet ou non de distinguer l'information importante, et de mettre des informations en relation les unes après les autres. La machine ne dispose pas de cette capacité : dans les cabinets d'avocats par exemple, toute la jurisprudence est disponible en ligne, indexée avec des liens hypertextes, mais le bon avocat sera celui qui saura construire une stratégie juridique avec cette information.

Quant à l'apprentissage automatique, il permet, pour reprendre d'Aurélie Jean, de formuler des hypothèses et non des certitudes. Dès lors, plus encore que la question des « biais » éventuels, c'est l'usage irraisonné de des données et des hypothèses qui pose problème. Plus encore que d'une « éthique » de l'IA, c'est de normes dont nous avons besoin. Le RGPD, qui pose que toute décision importante doit être complétée par une intervention humaine, va dans le bon sens.

Loin de faire disparaître le besoin d'une bonne expertise-métier, l'usage des capacités offertes par la machine en renforce le besoin. Et, comme le souligne Pierre Veltz, « la performance repose moins sur la qualité et le coût des diverses ressources que sur l'intelligence de leur combinaison, autrement dit l'efficacité de l'organisation du tissu relationnel ». L'automatisation des processus transactionnels va rehausser le rôle du relationnel, qui permet de faire la différence face à la concurrence : modes de management, échanges d'idées, facilitation, bon sens, créativité, goût du travail bien fait.

Sylvain Duranton, président de BCG Gamma, la structure du BCG dédiée à l'IA, constate que l'IA est comme la bureaucratie. Elle édicte des règles et des procédures au détriment du jugement humain, réputé imparfait ou déficient. L'hubris de l' « algocratie », du gouvernement par les nombres (Alain Supiot, Olivier Rey) et les algorithmes conduit à la centralisation. C'est un autoritarisme qui se cache derrière la science. L'on retrouve toute la problématique de la gouvernance, qui vise à éliminer le politique. Comme le note le philosophe Denis Collin à propos de l'éthique d'Aristote, « *Il faut que l'homme juste soit capable de corriger la règle pour l'appliquer au cas particulier qu'il a sous les yeux* ».

Refuser l'algocratie n'implique pas, tant s'en faut, le refus d'un usage raisonné de l'Intelligence artificielle. Bien définir la place du robot et de l'Intelligence Artificielle n'est rien de moins qu'un projet de société. Il appartient donc à la France et à l'Europe de faire peser toute leur puissance normative face à celle des GAFAM qui ne s'embarrassent pas de scrupules superflus.

L'accent mis sur l'éthique de l'Intelligence Artificielle dans le discours de nombreux Européens dans la foulée de la « Déclaration de Montréal pour le développement responsable de l'Intelligence Artificielle » de Décembre 2018, ne sert-il pas à cacher l'absence de vrai projet industriel européen dans ce domaine et l'absence de vrais acteurs européens permettant d'établir un rapport de force ? Après tout, le gouvernement chinois a pu faire plier Apple parce qu'il y avait une concurrence chinoise (Huawei, Xiaomi). En Europe, nous n'avons pas le choix et il ne nous reste que le poids de la norme.

« *L'avenir est dans les volontés, non dans les prophéties* » déclare une des personnages de Montherlant. Encore faut-il que cette volonté soit éclairée, par la connaissance des circonstances, ses moyens mais aussi et surtout sur ses buts, à savoir « la vie bonne » (eudaimonia) dont nous parle Aristote dans l'Éthique, le « souverain bien » de Montaigne. Faute de quoi, elle n'est que ce qu'Heidegger, dans son séminaire sur Nietzsche, appelait la « volonté de volonté ». Il appartient aux peuples européens de prendre leur destin en main.

D. Inventer la vie qui va avec : formation et organisation, économie de la connaissance et rôle des entreprises

Mettre l'Intelligence Artificielle au service de la société et en tirer le meilleur parti implique de mettre l'accent sur les priorités, et notamment l'organisation, la collaboration homme-machine et entre les acteurs et la mise en place d'un dialogue social capable d'anticiper l'avenir du travail.

1. Organisation

En France, la baisse ces dernières années du rythme de progression de la productivité (et l'écart avec l'Allemagne) s'explique semble-t-il par ce que les économistes appellent la productivité globale des facteurs, c'est-à-dire par la dimension organisationnelle, l'apprentissage de la technologie, l'évolution des qualifications.

On constate de vraies difficultés à mobiliser les filières autour des enjeux de l'Intelligence Artificielle. En effet, beaucoup d'entreprises ont compartimenté leurs données par service, par business unit, par pays, et ne sont pas en mesure d'utiliser à plein les possibilités ouvertes par la transformation numérique. Les entreprises ont dû mal à appréhender et valoriser les compétences transversales. Le *lean management* a conduit à supprimer les métiers de « middle office » où les connaissances et l'expérience pratiques des salariés leur permettent de résoudre des problèmes.

Les PME ne disposent pas toujours des compétences internes pour accompagner la mise en place de robots ou des outils numériques. Mais elles ne sont pas les seules. Selon une étude Qlik et Accenture France portant sur 1000 personnes en 2020, 70% des salariés présentent des signes de stress et d'épuisement lorsqu'ils doivent travailler sur des données. Les entreprises ont mis en place des outils en négligeant l'accompagnement et la formation

2. Collaboration

Toutes les complémentarités homme-machine ne sont pas souhaitables si elles induisent une soumission de l'homme à la machine. On cite souvent l'exemple de la logistique d'Amazon dans certains pays, où les pickers (ceux qui assemblent les commandes) sont en permanence surveillés par l'algorithme, y compris lorsqu'ils veulent se rendre aux toilettes (surveillance algorithmique). On ne peut que constater l'insuffisance des réflexions sur le déploiement d'une automatisation réellement collaborative, avec, à la clé, le risque de tomber dans les pièges dénoncés par Acemoglu et Restrepo, à savoir un écart entre des investissements et des résultats décevants.

3. Mutualisation

Si une entreprise accepte de mutualiser ses données, elle bénéficie de l'ensemble des données des entreprises participantes. C'est ce que semblent avoir compris Total et Thales, qui ont décidé de mutualiser leurs modèles d'apprentissage. L'Institut Mines Telecom a créé en 2014 Teralab, une plateforme de mutualisation et de partage des données entre industriels et académiques pour les projets de R&D. Au niveau européen, le projet AI4EU rassemble 90 partenaires, dont Thales en France.

4. Dialogue social

Pour l'OCDE, la moitié des effectifs des grands groupes européens vont avoir besoin de reconversion. Villani note justement qu'il faut des expérimentations en ciblant prioritairement l'accompagnement des transitions professionnelles des personnes qui sont aujourd'hui employées. On constate un décalage criant entre la mobilisation qui existe en Allemagne autour du travail du futur (Zukunft der Arbeit) qui associe entreprises, instituts de recherche, syndicats et collectivités

locales et ce qui se passe de notre côté du Rhin. Le dialogue social lancé en 2015 (initiative *Arbeitviernull*) a d'ailleurs donné lieu à un intéressant relevé de conclusions (Weissbuch 2017).

Stiglitz et Greenwald : La nouvelle société de la connaissance

Reprenant les thèses de Kenneth Arrow, qui avait montré que l'augmentation du taux d'épargne, et donc l'accumulation du capital, n'entraînait une augmentation du taux de croissance qu'en présence de gains de productivité, Bernard Stiglitz et Glenn Greenwald rejoignent l'argumentation développée par Acemoglu et Restrepo (*cf. supra*) : l'investissement et la technologie ne contribuent à la croissance que dans le cadre d'une société apprenante, qu'ils appellent la *Nouvelle Société de la Connaissance*.

Entreprises apprenantes

Il ne suffit pas d'un bon système éducatif, encore faut-il que les entreprises prennent le relais, car c'est souvent au sein des entreprises que l'on apprend réellement un métier. L'apprentissage passe souvent par la pratique et toute innovation se fonde sur des innovations antérieures.

En effet, lorsqu'on maîtrise déjà une technique ou un processus, on a moins à apprendre pour s'améliorer, puisqu'on a l'acquis technique et l'acquis d'apprentissage. À l'inverse, si l'on est loin de la frontière technologique, il y a plus à gagner mais c'est plus difficile et plus aléatoire. On le voit avec la filière nucléaire française, qui s'est laissé distancer et a perdu du savoir-faire, comme le montrent les difficultés de l'EPR en Finlande, bien analysées par le rapport Folz, et le fiasco de l'ATMEA en Turquie. Comme nous l'avons vu dans les analyses de la Dares, les entreprises qui n'innovent pas laissent s'accumuler un déficit de savoir-faire qui les conduit à se battre uniquement sur les coûts, ce qui conduit des stratégies d'automatisation destructrices d'emploi ou à des délocalisations.

Rente

Pour que les entreprises soient incitées à innover – car apprendre est un investissement -, il faut qu'elles aient une perspective de rente plus ou moins solide pour rembourser les coûts encourus. Si l'apprentissage génère des économies de coût dans le futur, l'entreprise peut actualiser cette économie dans ses coûts actuels, baisser son prix de vente et conforter dès aujourd'hui sa rente d'innovation. Il en va de même des individus : On peut penser que les seniors ne sont pas incités à investir dans la connaissance car la rente qu'ils peuvent en tirer est moindre que celle des plus jeunes. Les grandes entreprises sont capables d'internaliser une plus grande partie des externalités produites par l'apprentissage et les transformer en rente. Le modèle à suivre pour les Européens ne serait pas nécessairement la « start up nation », même si les start up ont un rôle fondamental dans la « destruction créatrice » chère aux Schumpetériens. Les États-Unis sont plus doués que les Européens pour transformer l'innovation en rente financière. Les États-Unis protègent leurs inventions, ils sont donc peu donneurs de savoir. Mais ils profitent des inventions des autres qui sont dans le domaine public (preneurs de savoir). La propriété intellectuelle est aux entreprises ce que le protectionnisme est aux nations. Pour le reste de la société, il ne faut pas que cette rente soit trop exclusive, ce qui explique les débats sur un éventuel démantèlement des GAFAM.

Les limites du marché

Pour Schumpeter : les secteurs où l'innovation est importante connaîtront probablement une concurrence imparfaite. Les marchés à fort apprentissage risquent d'être dominés par une seule entreprise et il existe une prime au premier entrant, car il peut amortir plus vite ses coûts fixes. Dès lors, « *rien n'indique que les marchés soient efficaces pour produire et diffuser les connaissances et les pratiques acquises* ». Face à cette situation, concluent les auteurs, « *l'objectif de l'État est de développer les secteurs qui ont de fortes élasticités d'apprentissage (qui génèrent des rentes d'apprentissage) et de gros débordements d'apprentissage (savoirs qui profitent à l'ensemble des acteurs)* », y compris par le protectionnisme.

E. Investir dans la recherche et lutter contre le Brain Drain

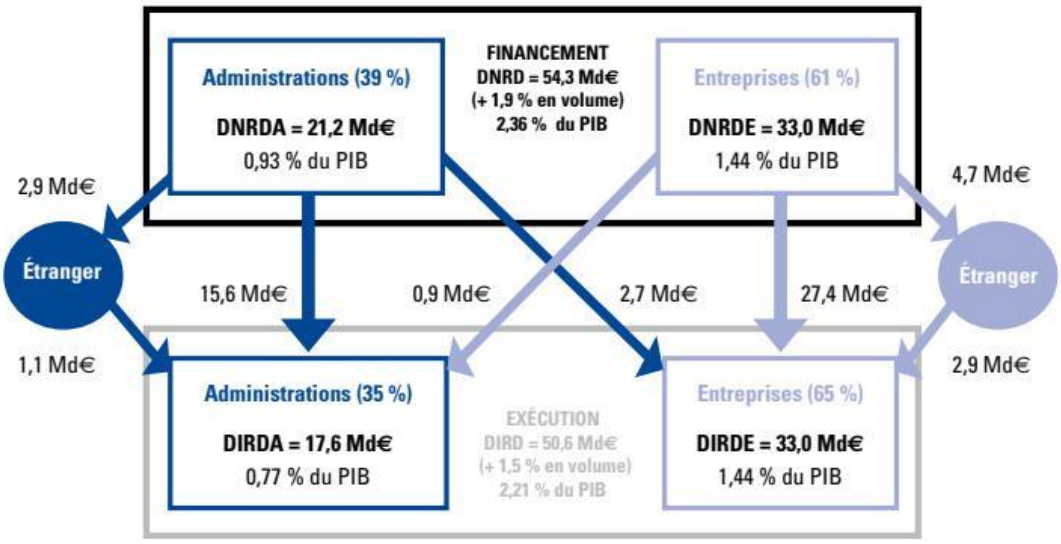
La recherche est confrontée, comme l'a bien identifié le rapport Villani, à un triple défi :

1. Des moyens insuffisants pour la recherche publique en Intelligence Artificielle et une faible interdisciplinarité

L'effort de recherche annuel de la France se situe aux environs de 50 milliards d'euros par an, soit 20 milliards issus de l'État et 30 milliards venant des entreprises. Ces dernières bénéficient en retour du Crédit d'Impôt Recherche (CIR) qui s'élève à environ 6 milliards d'euros par an. Les dépenses totales de recherche représentent 2.2% du PIB, proportion à peu près constante ces dernières années. C'est un peu moins que la moyenne de l'OCDE (2.4%) et surtout moins que certains pays comme l'Allemagne ou les pays Scandinaves qui se situent au-dessus de 3%. Notons que les États-Unis affichent en ratio d'environ 2,7% et que des pays comme Israël ou la Corée du Sud dépassent les 4%.

La meilleure performance de l'Allemagne par rapport à la France s'explique un peu par la force du *Mittlestand*, mais surtout par le poids des industries automobile et chimie-pharmacie, qui sont comme en France les premiers contributeurs à la recherche privée.

GRAPHIQUE 1 - Le financement et l'exécution de la recherche en France en 2017 (résultats semi-définitifs)



Source : INSEE

Sur ces montants, les crédits publics accordés à l'Intelligence artificielle sont principalement concentrés sur les principaux pôles de recherche comme l'INRIA, le CEA, le CNRS, Mines-Télécom et Paris VI. Concomitamment au rapport Villani, 660 M€ ont été débloquées au profit de la recherche en Intelligence Artificielle, complétés par 120 M€ venant du Fonds de Dotation pour l'Innovation (financé par la cession de titres par l'État : Engie, Renault, FDJ).

La Loi de Programmation pour la Recherche 2021-2030 vise à augmenter les crédits consacrés à la recherche publique de 25 milliards d'euros sur 10 ans. À horizon 2030, la dotation publique consacré aux administrations de recherche serait portée à 20 milliards d'euros contre une peu plus de 15 milliards aujourd'hui. Au total, les montants consacrés à la recherche atteindraient 3% du PIB d'ici 2030, un objectif ambitieux qui avait déjà été affiché il y a dix ans à horizon 2020. Entretemps, il n'est pas exclu que, passé l'effet de stupeur dû au Coronavirus, l'Allemagne et les Pays-Bas n'exigent le retour au respect des objectifs de frugalité budgétaires énoncés dans le Pacte de Stabilité et de Croissance.

Cette augmentation serait, qui plus est, graduelle : 400 millions d'euros en 2021, 800 millions d'euros en 2022 et de 1,2 milliard d'euros en 2023, auxquels s'ajouteront 4,6 milliards d'euros du Plan de Relance et du volet 4 du Plan d'Investissements d'Avenir.

Faut-il par ailleurs envisager de repositionner certains établissements universitaires qui mobilisent aujourd'hui des budgets conséquents mais offrent peu de débouchés vers les métiers en tension ? C'est complexe politiquement mais pas infaisable sur la durée. Encore faut-il trouver le personnel enseignant, ce qui n'est pas une mince affaire comme le montrent les difficultés de mise en œuvre liés à la suppression du numerus clausus en médecine. Cela ne signifie pas, tant s'en faut, qu'il faille négliger la recherche en Sciences Humaines, car seule une interdisciplinarité bien menée pourra permettre d'éviter les dérives observées aux États-Unis dans la conception et l'utilisation des outils fondés sur l'apprentissage automatique.

2. Une fuite des cerveaux qui peut prendre des formes insidieuses

Mais la France fait face aujourd'hui à un véritable exode des cerveaux en matière d'Intelligence Artificielle. Au-delà de quelques noms largement médiatisés (comme Yann Le Cun), ce sont pratiquement des promotions entières de chercheurs qui quittent le territoire national à l'issue de leurs études. L'une des raisons est la faiblesse des rémunérations offertes aux jeunes chercheurs. C'est pourquoi, à partir de 2021, le Gouvernement s'engage à ne plus recruter aucun jeune chercheur en-dessous de deux SMIC (contre environ 1,5 SMIC en moyenne aujourd'hui).

Mais, plus que les salaires, c'est l'environnement de travail qui les attire aux États-Unis ou au Royaume-Uni. Par ailleurs, Facebook, Microsoft, Google, IBM ou Intel ont ouvert des unités de recherche en IA en France. Les GAFAM viennent-il siphonner les compétences en France ? La réponse doit être nuancée.

D'un côté, on peut considérer que, quand Google et Facebook embauchent une part importante des chercheurs en Intelligence Artificielle en Europe, ils préemptent les ressources disponibles au détriment de la stratégie de la France dans ce domaine. Il y aurait dans le moindre environ 25.000 chercheurs en IA de haut niveau et les grands groupes technologiques mondiaux entrent en concurrence avec la recherche publique des États et avec les acteurs privés nationaux. Il serait, dès lors, légitime d'envisager un mécanisme analogue à la « pantoufle » en vigueur à l'ENA ou a Polytechnique.

À l'inverse, lorsqu'un ingénieur est recruté par Facebook ou Google en France, est-ce que cela génère des externalités d'apprentissage au profit de l'économie française ? On peut envisager que le passage par un GAFA soit une « école » dont bénéficiera l'économie française (sur le modèle des *returnees* » en Inde). Il faudrait donc appliquer les mécanismes de rétention des compétences avec souplesse.

3. Une faible valorisation de la recherche

La France se situe au troisième rang mondial pour la recherche en robotique derrière les USA et le Japon et a une forte tradition mathématique de haut niveau, avec 13 médaillés Fields, juste derrière les États-Unis. De même, on estime que le tiers des modèles d'apprentissage ont été inventés en Europe, mais que la diffusion dans les entreprises ne suit pas. Il existe pourtant en France des communautés actives de chercheurs et de développeurs, comme l'Association Française pour l'Intelligence Artificielle (AFIA), Deep Learning Paris Meetup, Paris Artificial Intelligence, Kaggle Paris Meetup, etc. Mais elles manquent souvent d'ouverture vers le monde de l'entreprise.
Si l'on se situe à l'échelle macro-économique, le ratio recherche privée / recherche publique est de 2 aux Etats-Unis contre 1,5 en France. C'est la recherche appliquée, plutôt menée dans les institutions privées, qui pêche dans notre pays.

La France, comme le Royaume-Uni, consacre une part bien moins importante de ses dépenses de R&D au développement de nouveaux produits que ses grands concurrents internationaux

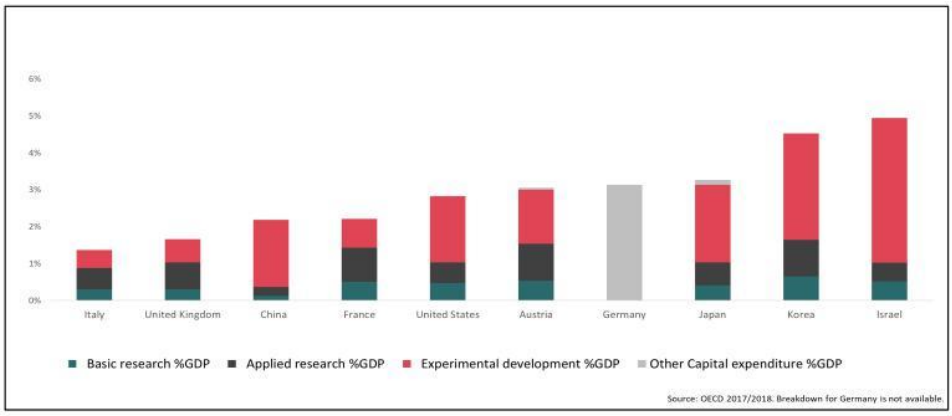


Figure 1 - Basic research, Applied research and Experimental Development as % of GDP

Source : *UK Research and Development Roadmap, juillet 2020*

Pour y remédier, la loi PACTE a facilité la possibilité pour les chercheurs universitaires de travailler jusqu'à 50% de leur temps en entreprise sans passer par la commission de déontologie, mais ce type de mesure est en vigueur aux États-Unis depuis 1980. En dépit de la mise en place des SATT (sociétés d'accélération du transfert de technologies), les transferts de l'université vers les acteurs économiques sont limités en France. On note beaucoup de publications en robotique et en IA, mais pas d'acteurs économiques significatifs. Comme le disait Nicolas Dufourq, DG de la BPI, les start up hightech ne forment qu'une « *forêt de bonsaïs* ».

La LPR 2021-2030, précédemment citée, vise à horizon de 10 ans la création de 500 entreprises issues de la recherche publique chaque année. Depuis leur naissance en 2010, les SATT ont donné lieu à la création de 500 entreprises, soit une moyenne de 50 par an. C'est dire si le nouvel objectif affiché par la ministre de l'enseignement supérieur place la barre très haut …

4. La recherche doit marcher sur deux pieds

Comme l'a bien montré l'économiste italienne Maria Mazzucato, la recherche fondamentale, menée par les institutions publiques, est à l'origine de tous les grands progrès technologiques depuis la Seconde Guerre Mondiale. On notera que Geoffrey Hinton, précédemment cité, a fait toutes ses découvertes au cours des trente-cinq ans qu'il a passés au sein d'universités publiques, avant de rejoindre Google en 2013.

Aux États-Unis, un document publié par l'Office of Science and Technology Policy (rattaché à la Maison Blanche) en 2006 démontre que tous les composants de smartphone d'Apple sont issus de la recherche publique américaine, notamment dans le domaine militaire (DARPA, Département de la Défense, Nationale Science Foundation, etc.). À quelques exceptions près, seul l'État a les moyens de se lancer dans des projets à long terme et sur lesquels pèse un grand degré d'incertitude. Le capital-risque privé n'intervient en général qu'au stade suivant, lorsque la technologie a été validée par une « preuve de concept ». Il finance donc plutôt la recherche appliquée au côté des grands groupes.

Dans un pays comme la France, le « passage de relais » entre recherche privée et recherche publique constitue un défi qui reste à relever. L'autre défi est la capacité des entreprises à développer, pour leur usage interne (innovation de procédé) ou à usage commercial des applications permettant de générer des rentes technologiques, comme savent si bien le faire les GAFAM. D'autant que du point de vue économique, nous ne sommes pas seulement dans une « course aux armements » (à l'innovation) perpétuelle. On peut aussi raisonner comme Kai Fu Lee en appliquant la théorie de Clayton Christensen sur le *good enough* : les rentes d'innovation ne sont pas seulement

issues de produits se situant à la frontière technologique mais aussi de produits et services « raisonnablement innovants », comme le prouve le succès de la *fintech* chinoise Ant Financial, filiale d'AliBaba. Il s'agit aussi – et c'est le cas pour l'Intelligence Artificielle - de mettre rapidement sur le marché, des produits et des services afin de capter une base de clients, pour se constituer de grandes bases de données que l'on puisse entraîner, sachant que les briques logicielles qui servent à bâtir les algorithmes sont accessibles en *open source* et qu'il n'est donc pas nécessaire de réinventer la roue à chaque fois. Le succès du Japon, puis de la Corée du Sud repose sur un important effort de R&D, mais aussi et surtout sur la présence de grands groupes comme Sony ou Samsung qui ont su transformer les découvertes scientifiques en rentes d'innovation.

F. Lutter contre la polarisation, recréer une classe moyenne

Comme on l'a noté précédemment, les principaux pays occidentaux subissent une érosion de la classe moyenne et une tendance à la polarisation du spectre social. Ceci est particulièrement vrai aux États-Unis, mais l'Europe n'est pas en reste. Une récente étude de McKinsey Global Institute, *The future of work in Europe*, distingue 48 métropoles, regroupant 20% de la population européenne (EU 28+ Suisse), qui représentent 43% du PIB. Elles ont bénéficié de 35% de la croissance de l'emploi depuis 2007. Et ce chiffre pourrait atteindre 50% d'ici 2030. De l'autre côté du spectre, 438 régions représentant 30% de la population à faible niveau d'éducation, notamment en Europe du Sud et de l'Est devraient voir leur population baisser, sous l'effet du déclin démographique et exode vers les régions plus prospères.

Comme le note la DARES, les régions d'emploi industriel traditionnel (Nord-Est) devraient plutôt continuer à souffrir de l'automatisation alors que les nouveaux emplois qualifiés seraient créés dans les métropoles du fait de l'effet d'agglomération (la « classe créative » chère à Richard Florida). Ceci reflète aussi la politique de métropolisation comme « porte d'entrée de la mondialisation » - mais irrigant assez peu leur environnement - prônée par la DATAR, qui n'a pas vraiment été remise en cause par ses successeurs, le GGET et l'Agence nationale de la cohésion des territoires en dépit de leur intitulé.

En France, cette évolution a été compensée par la montée des transferts sociaux, mais la désindustrialisation et la concentration des emplois bien rémunérés dans les métropoles représentent des problèmes sociaux et politiques majeurs car, à terme, cette situation est difficilement tenable.

1. Formation

Une des réponses à cette situation est le développement de la formation. L'OCDE note que les pays qui ont pu développer une classe moyenne liée aux professions du numérique sont ceux, comme les pays scandinaves, qui ont investi dans les formations qui préparent à ces métiers. Le paradoxe est que la France dispose d'une recherche de haut niveau en mathématiques, tandis que les classements de type PISA (Programme International de Suivi des Acquis des élèves, coordonné par l'OCDE) portant sur les adolescents de plus de 15 ans, la situent à un niveau relativement médiocre dans cette discipline (juste dans la moyenne de l'OCDE au classement 2019 portant sur les données 2018). De plus, chaque année, 100.000 jeunes sortent du système scolaire sans diplôme.

La formation professionnelle, à laquelle la France consacre 30 Mds€ par an, a bien du mal à compenser ces handicaps et à doter les salariés des compétences dont ils auront besoin dans le futur. Le rapport sur la formation professionnelle de 2008 a mis en lumière l'accès très inégal des salariés à la formation. Différentes réformes et notamment l'institution du Compte Personnel de Formation, qui est un droit individuel opposable, n'ont pas donné les résultats escomptés, surtout après que la réforme de 2018 a transformé les droits en heures en droits en euros.

Mais l'essentiel n'est pas là. Il existe des difficultés structurelles, comme le fait qu'on ne sait pas bien comment former des adultes, en particuliers les personnes qui n'ont pas un bagage scolaire de

base. De son côté, la validation des acquis de l'expérience (VAE) hésite toujours entre s'adapter aux aptitudes des publics concernés et sacrifier aux exigences académiques. Quant à l'enseignement en ligne, on constate que les MOOC s'adressent en fait à un public de diplômés. Si l'on suit le raisonnement de Greenwald et Stiglitz, précédemment évoqué, c'est aux entreprises de prendre en charge cette démarche de mise à niveau des compétences, en allant plus loin dans la gestion prévisionnelle des emplois et des compétences.

2. Améliorer la GPEC

La loi de programmation pour la cohésion sociale, dite loi Borloo-Larcher, de 2005, a instauré une négociation sur la Gestion Prévisionnelle des Emplois et des Compétences (GPEC) obligatoire tous les trois ans dans les sociétés de plus de 300 salariés. Elle a permis une meilleure visibilité sur les parcours de cours de carrière dans les grands groupes, mais son application a été très inégale. Ainsi, la loi visait notamment, à l'origine, à instaurer une seconde partie de carrière pour les seniors, mais cela a plus débouché en pratique sur une préparation à la retraite que sur un véritable rebond. Ce n'est pas neutre quand on sait que l'âge moyen des opérateurs de l'industrie automobile est supérieur à 50 ans. La loi n'a donc pas toujours permis d'enrayer le déclin du capital-savoir dans les entreprises.

Les entreprises de taille moyenne, qui n'étaient soumises qu'à la présentation annuelle d'un plan de formation, ont été rebutées par la difficulté de formaliser des profils de postes, souvent gérés de manière informelle.

Enfin, certaines entreprises, notamment dans les secteurs les moins innovants, ne voient pas l'intérêt d'une démarche prospective. Les abattoirs GAD en Bretagne, rendus célèbres par une phrase maladroite, mais non dénuée de vérité, du président de la République, constituent un exemple emblématique de l'impasse liée à un manque de formation initiale conjugué à l'absence d'évolution des compétences.

Pour en revenir aux problématiques d'automatisation, l'analyse du potentiel de destruction d'emplois évoquée précédemment repose trop souvent sur un état des tâches inchangé. Elle ne prend pas en compte l'évolution des tâches et l'apparition de nouveaux modèles économiques qui rendront peut-être certains métiers inutiles ou non compétitifs.

Dès lors, est posée la question de l'élargissement des dispositifs envisagés par la loi de 2005. Cela suppose de mobiliser la représentation nationale et les partenaires sociaux, dans le cadre d'un Accord National Interprofessionnel par exemple. Une entreprise comme Orange a lancé en 2019 une vaste campagne de mise à niveau des compétences (« *reskilling* ») qui pourrait faire exemple de modèle.

Dans son essai évoqué précédemment, Kai Fu li considère que « *ceux qui prônent la requalification des travailleurs ont tendance à croire que l'IA va opérer un changement graduel des compétences requises sur le marché de l'emploi* » (…) mais « *à mesure que l'IA va s'étendre à des nouvelles professions, les salariés devront se reconvertir à intervalles de plus en plus rapprochés* ». Loin de plaider pour une approche pessimiste, ceci dénote un besoin d'adaptation permanente, sachant que la mutation ne sera pas la même selon que l'on est en croissance (réallocation des emplois facilitée) ou en stagnation économique.

3. Pour une IA inclusive

La question que se pose dans les périodes de mutation est double : Comment développer la résilience des emplois ? Comment donner les meilleures chances aux nouveaux entrants sur le marché du travail et fournir aux entreprises les compétences dont elles ont besoin ? Cela ne passe pas par la transformation de chacun en super-développeur informatique mais surtout par la formation d'utilisateurs avisés des données ou en combinant tâches automatisées et tâches

humaines. En d'autres termes, il s'agit au moins autant d'adapter les outils aux utilisateurs que les utilisateurs aux outils.

Pour combler le « gap de compétences », il faut donc combiner l'amélioration du système éducatif et de la formation continue pour adapter les salariés aux technologies (c'est le discours dominant) mais il est nécessaire de penser également l'adaptation des technologies aux salariés.

Comme le montre l'histoire de l'informatisation des entreprises, les besoins ne sont cantonnent pas à des compétences pointues. Ils se situent au niveau des développeurs, voire des utilisateurs finaux. On constate qu'il y a aussi un fort besoin de professionnels de l'IA de niveau Bac+3 : industrialisation, production d'analyses, visualisation de données, intégration de composants logiciels. Comme l'a montré Geoffrey Moore (*Crossing the Chasm*), au fur à mesure que les technologies deviennent matures et se répandent dans la société, il y a de plus en plus besoin de compétences « moyennes ». C'est aujourd'hui un territoire peu couvert par l'enseignement supérieur et qui conditionnera notre capacité à maîtriser ces technologies. Ce n'est pourtant pas tant la sophistication des outils qui permet d'améliorer la productivité que leur adoption en masse dans les entreprises.

Les salariés présents dans les entreprises représentent un formidable gisement de compétences. Leurs capacités peuvent être augmentées par l'informatisation et le recours à des outils de *business* intelligence ou d'Intelligence Artificielle pourvu qu'ils soient suffisamment conviviaux. Comme le note Dani Rodrik, les entreprises feraient mieux d'exploiter ce gisement avant de s'escrimer à trouver à l'extérieur des compétences.

Lorsqu'on observe le succès d'une entreprise comme Toyota, on constate qu'elle a toujours favorisé la promotion interne et la montée en qualification où chacun monte d'une marche au fil des années. Le succès de Toyota à Onnaing repose sur la forte mobilisation des collectivités locales, mais aussi sur des méthodes de gestion (Toyota Way), la formation continue et la flexibilité. Cet exemple devrait nous inciter à réfléchir à ce que sera l'avenir du travail, non pas sans l'homme, non pas contre mais avec la machine. Car ne l'oublions pas, demain commence aujourd'hui.

VI. BIBLIOGRAPHIE

Ouvrages

Antonio Casilli : *En attendant les robots Enquête sur le travail du clic*, Seuil, 2019

Jean-Louis Dessalles : *Des intelligences TRES artificielles*, Odile Jacob, 2019

Jean-Gabriel Ganascia: *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?* Seuil, 2017

Jean-Gabriel Ganascia: *Intelligence artificielle : vers une domination programmée ?* Le Cavalier Bleu, 2017

Aurélie Jean : *De l'autre côté de la machine*, Éditions de l'Observatoire, 2019

Brittany Kaiser : *L'affaire Cambridge Analytica*, Harper Colins, 2020

Kai-Fu Lee : *IA, la plus grande mutation de l'histoire*, Les arènes, 2019

Mariana Mazzucato : *L'État entrepreneur*, Fayard 2020

Olivier Rey : *Quand le monde s'est fait nombre*, Stock, 2016

Gary Smith : *AI Delusion*, Oxford University Press, 2018

Barbara Stiegler : *Il faut s'adapter*, Gallimard, 2019

Joseph Stiglitz et Bruce Greenwald : *Une nouvelle société de la connaissance*, Les liens qui libèrent, 2017

Pierre Veltz : *La société hyperindustrielle*, Le Seuil, 2017

Rapports et articles :

Cédric Villani : *Donner un sens à l'intelligence artificielle*, Rapport de mission parlementaire (mars 2018)

Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans : *When Will AI Exceed Human Performance ? Evidence from AI Experts*, Oxford, mai 2017

Daron Acemoglu et Pascual Restrepo : *The race between Man and Machine*, NBER, mai 2016

Daron Acemoglu et Pascual Restrepo: *Robots and jobs : Evidence from US Labor Markets*, NBER Working Paper, mars 2017

Daron Acemoglu et Pascual Restrepo : *The wrong kind of AI : Artificial Intelligence and the Future of Labor Demand*, JEP, mars 2019

Daron Acemoglu et Pascual Restrepo: *Automation and new tasks, How technology displaces and reinstates labor*, Journal of Economy Politics, printemps 2019

Ifeoma Ajunwa : *The paradox of automation as anti-bias intervention*, Cornell University, mars 2016/ Cardozo Law Review, juin 2020

Automation and a changing economy, Aspen Institute 2019

Massimo Anelli, Italo Colantore, Piero Stanig : *We were Robots*, IZA, 2019

Philippe Aghion, Benjamin Jones, Charles Jones : *Artificial Intelligence and Economic Growth*, NBER Working Paper 23928, octobre 2017

David Autor, Frank Levy, Richard Murnane : *The skill content of technological change*, The Quarterly Journal of Economics, novembre 2003

David Autor, : *Why are there still so many jobs ?* Journal of Economic Perspectives, été 2015

David Autor, David Dorn, Gordon Hanson : *Untangling Trade and Technology: Evidence from Local Labour Markets*, The Economic Journal, mai 2015

David Autor, David Dorn, Gordon Hanson : *The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade*, Annual Review of Economics, octobre 2016

David Autor, Salomons : *Is Automation Labor Share–Displacing ?* (Brookings Papers on Economic Activity, 2018)

David Autor, David Mindell, Elisabeth Reynolds : *The world of the future : Shaping Technology and Institutions*, MIT, 2019

Charline Bardet : *Comment ont évolué les métiers en France depuis 30 ans*, DARES, mars 2017,

Andrew Berg, Edward F. Buffie, and Luis-Felipe Zanna : *Should We Fear the Robot Revolution ?* (The Correct Answer is Yes), IMF Working Paper, mai 2018

Erik Brynjolfsson, Tom Mitchell, and Daniel Rock : *What Can Machines Learn and What Does It Mean for Occupations and the Economy ?* AEA Papers and Proceedings, 2018

Cristian S. Calude and Giuseppe Longo : *The Deluge of Spurious Correlations in Big Data*, Foundations of Science, mars 2016

Francesco Chiacchio, Georgios Petropoulos, David Pichler : *The impact of industrial robots on EU employment and wages* : A local labour market approach, Bruegel Working Paper, 2018

Conseil d'Orientation pour l'Emploi : *Automatisation, numérisation*, Emploi, 2017

Adrian Cooper et al. : *How robots change the world. What automation really means for jobs and productivity*, Oxford Economics, juin 2019

Charles-Marie Chevalier et Antoine Luciani : *Informatisation, productivité du travail et emploi*, INSEE, 2018 Antoine

Wolfgang Dauth, Sebastian Findeisen, Jens Suedekum, Nicole Woessner : *Adjusting to Robots : Worker-Level Evidence*, unpublished, 2018

Duco de Vos, Evert Meijers & Maarten van Ham : *Working from home and the willingness to accept a longer commute*, Annals of Regional Science, juillet 2018

Olivier Ezratty : *Les usages de l'intelligence artificielle*, 2017

Olivier Ezratty : *Que devient l'IA symbolique ?* 2018

France Stratégie : *Intelligence artificielle et travail, risques et opportunité*, mars 2018

Carl Benedikt Frey et Michael Osborne : *The future of employment: How susceptible are jobs to computerisation ?* Technological Forecasting and Social Change, 2017

Carl Benedikt Frey et Michael Osborne : *China Won't Win the Race for AI Dominan*ce, Foreign Affairs, Juin 2020

Claude de Ganay et Dominique Gillot : *pour une intelligence artificielle maîtrisée, utile et démystifiée*, Rapport au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, 15 mars 2017

Lawrence Katz, Robert Margo : *Technical Change and the Relative Demand for Skilled Labor*: The United States in Historical Perspective, NBER, 2014

McKinsey Global Institute , James Manyika et al : *Jobs losts, jobs gained : Workforce transitions in a time of automation*, 2017

McKinsey Global Institute : *Tackling Europe's Gap in digital and AI*, 2019

McKinsey Global Institute : *The Future of Work in Europe*, juin 2020

Mark Muro, Robert Maxim, Jacob Whiton : *Automation and Artificial Intelligence. How Machines are affecting People and Place*. Brookings Metropolitan Policy Program, janvier 2019

Mark Muro, Robert Maxim, Jacob Whiton : *What jobs are affected by AI*, Brookings Metropolitan Policy Program, novembre 2019

Mathilde Pak et Aurélien Poissonnier : *Accounting for technology, trade and final consumption in Employment : an Input-Output decomposition*, INSEE, décembre 2016

Eliza Strickland : *How IBM Watson Overpromised and Underdelivered on AI Health Care*, IEEE spectrum 02 Apr 2019

VII. Table des matières

I. En guise d'introduction ... 3
II. De quoi parle-t-on ? / Où en est-on ? ... 5
A. Quelques concepts ... 5
B. Les intelligences artificielles ... 5
C. L'IA symboliste ... 6
D. L'apprentissage automatique (machine learning) ... 6
E. Le deep learning ... 8
F. L'histoire n'est pas terminée *:* ... 10
G. L'apprentissage non supervisé, nouvelle frontière de l'IA ... 11
III. Entre réalisations impressionnantes et déconvenues cinglantes ... 12
A. Vers l'industrie 4.0 ? ... 12
1. La robotique industrielle, un engouement inégalement réparti ... 12
2. L'internet des objets et le grand défi de la connectivité ... 13
3. Réalités et mirages de la data driven corporation ... 13
4. Et la France ? ... 14
B. Les applications médicales : Watson, robots médicaux et diagnostic automatique ... 14
1. Du chirurgien augmenté ... 14
2. …au diagnostic automatisé ... 15
C. Finance : la prudence s'impose ... 16
D. Vers la domination de la machine ? Le mythe de l'Intelligence Artificielle Générale ... 18
1. La thèse de la Singularité ... 18
2. Des performances mitigées ... 18
3. L'Intelligence Artificielle ne « réfléchit » pas ... 19
4. L'Intelligence Artificielle ne réplique pas vraiment le cerveau humain ... 19
IV. Un parcours encore semé d'obstacles ... 20
A. L'emploi et l'inégalité ... 20
1. Un monde sans emploi ? ... 20
2. Un peu d'histoire ... 22
3. Polarisation sociale : vers la fin de la classe moyenne ? ... 23
4. Une réallocation régressive ? ... 25
5. La fin du compromis fordien ... 25
6. Tout ça pour ça ? ... 26
7. Et en France ? ... 27
B. Les libertés publiques ... 29
1. La rué vers la collecte des données personnelle ... 29
2. Vers un « capitalisme de surveillance » ? ... 29
3. Les dangers de la surveillance de masse ... 30
C. Comprendre la querelle des biais ... 31
1. Au niveau des algorithmes ... 32

2. Au niveau des données ... 32
3. La sélection des variables et les problèmes de traitement des données ... 33
D. Les boites noires : à la recherche de l'explicabilité ... 35
E. La cybersécurité : au cœur de tous les dangers ... 36
1. Un danger aux multiples facettes ... 37
2. Qui sont les pirates ? ... 38
F. La menace pour la civilisation européenne ... 40
1. Le poids des GAFA est devenu insupportable ... 40
2. Une menace chinoise ? ... 41
3. Les particularités du système chinois ... 42
4. Les thèses de Kai Fu Lee ... 43
5. Les limites du modèle ... 44
V. Que faire ? ... 45
A. Les stratégies nationales comme France IA et le rapport Villani ... 45
B. Les coopérations européennes, subsidiarité ... 46
1. La situation de l'Europe ... 46
2. La stratégie de l'UE ... 46
3. Coopération bilatérale: le Royaume-Uni plutôt que l'Allemagne ? ... 47
C. L'Intelligence Artificielle est-elle vraiment intelligente ? Poser les limites pour trouver les bons usages de la machine ... 47
D. Inventer la vie qui va avec : formation et organisation, économie de la connaissance et rôle des entreprises ... 49
1. Organisation ... 49
2. Collaboration ... 49
3. Mutualisation ... 49
4. Dialogue social ... 49
E. Investir dans la recherche et lutter contre le Brain Drain ... 52
1. Des moyens insuffisants pour la recherche publique en Intelligence Artificielle et une faible interdisciplinarité ... 52
2. Une fuite des cerveaux qui peut prendre des formes insidieuses ... 53
3. Une faible valorisation de la recherche ... 53
4. La recherche doit marcher sur deux pieds ... 54
F. Lutter contre la polarisation, recréer une classe moyenne ... 55
1. Formation ... 55
2. Améliorer la GPEC ... 56
3. Pour une IA inclusive ... 56
VI. BIBLIOGRAPHIE ... 58