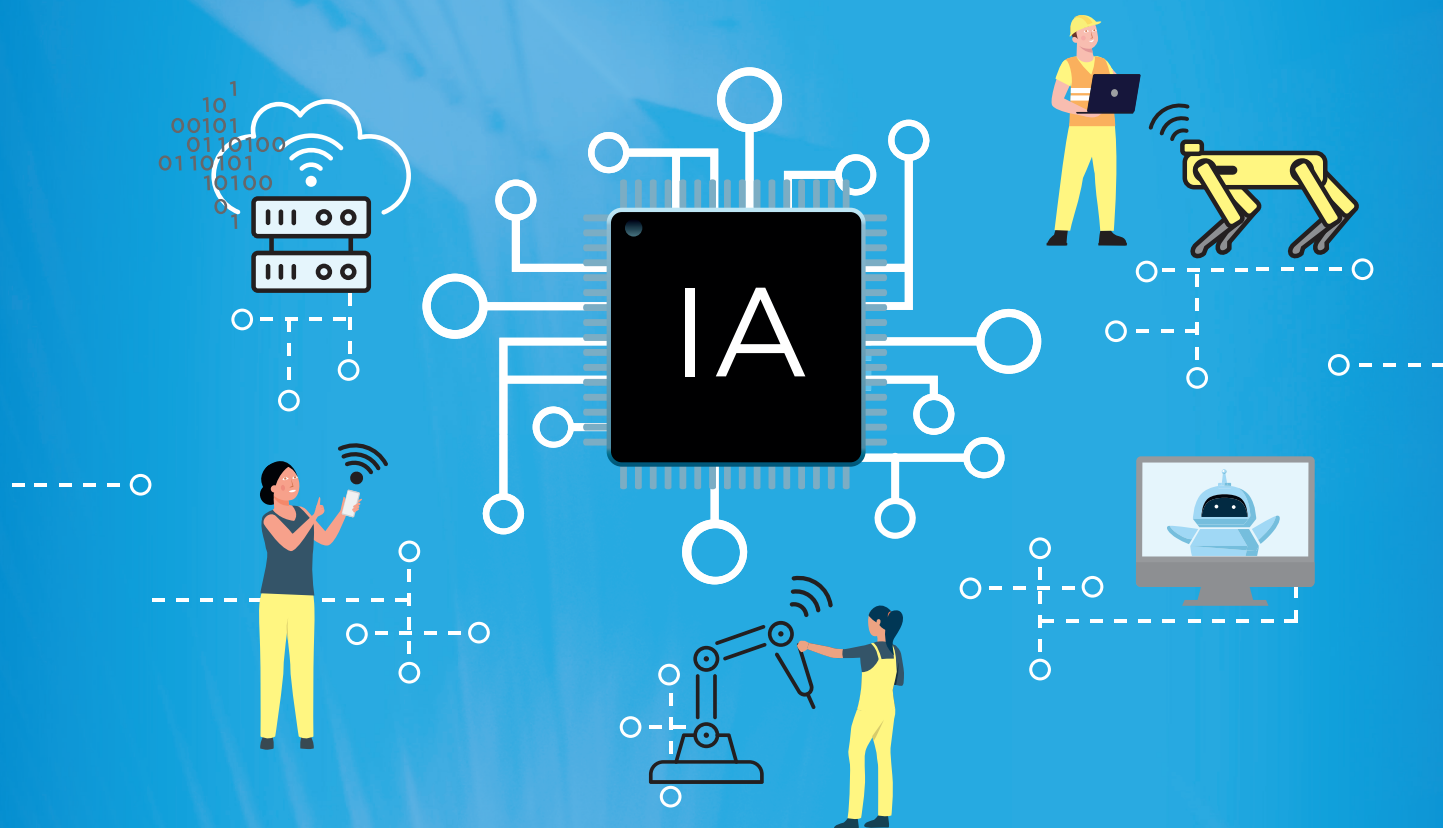




**l'Assurance
Maladie**
RISQUES PROFESSIONNELS



Institut National de Recherche et de Sécurité



L'intelligence artificielle au service de la santé et sécurité au travail

Enjeux et perspectives
à l'horizon 2035

Groupe projet

Anani Olympio (CNP Assurances) ; Bertrand Braunschweig (Bilab, Confiance.ai) ; François de Jouvenel, Jean-François Soupizet (Futuribles) ; Jorge Munoz (Université de Brest) ; Martin Bieri, Régis Chatellier (Cnil/Linc) ; Nazim Fatès (Inria/Loria) ; Nicolas Bouby (Bouygues construction) ; Sylvain Halluin (Cramif) ; Timothée Silvestre (CEA) ; Vincent Mandinaud (Anact) ; Jennifer Clerté, Marc Malenfer, Michaël Sarrey, Michel Héry (INRS).

Participants aux ateliers

Guillaume Boulanger (Santé publique France) ; Sabri Bayouhd (Blaxtair) ; Cegarra Julien (INU Champollion [Laboratoire SCoTE]) ; Joséphine Hurstel (CNNum) ; Sebastien Lissare (Cognitive Engine) ; Claire Tissot, Frédéric Clerc, Guy Hedelin, Lisa Chedik, Adriana Savescu, Lien Wioland, Louis Laurent, Patrice Marchal, Adel Sghaier, David Tihay (INRS).

Autres contributeurs

Loïck Briot (Mines Nancy) ; Yann Ferguson (Icam) ; Ariane Leroyer (Université de Lille) ; Agnès Aublet-Cuvelier, Benjamin Paty, Jean Theurel, Jean-Christophe Blaise, Cordelia Genzel, Nadiège Félicie, Virginie Govaere (INRS).

Les auteurs remercient également Céline Raynal de l'entreprise Safety Data et Florian Ruen de l'entreprise Eyes'R.



Avis au lecteur

La prospective n'est pas une prédiction de l'avenir. Elle n'est pas non plus une prévision qui serait le prolongement des tendances passées.

La prospective prend en compte les tendances et les discontinuités pour décrire des futurs possibles et proposer une aide à la prise de décision.

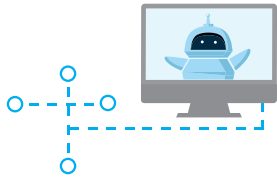
Les travaux présentés dans ce document sont le fruit de productions collectives. Ils ne préjugent en rien des opinions et souhaits des participants y ayant contribué et n'engagent ni les organismes dont ils sont issus, ni l'INRS.



Sommaire

■ Introduction	6
● Intelligence artificielle : de quoi parle-t-on ?	7
● Usages de l'intelligence en S&ST : quel périmètre ?	8
■ Messages clés	9
■ Méthodologie	13
■ Scénarios prospectifs : l'IA au travail en 2035	15
● Scénario 1 : Les géants du numérique imposent leurs solutions et leur vision	15
● Scénario 2 : Les États garantissent un cadre pour l'intégration de l'IA	19
● Scénario 3 : Développement démocratique	22
● Scénario 4 : Déclin de l'IA	26
■ Cas d'usage	29
■ Synthèse des ateliers	30
● Atelier 1 : Usages possibles des outils d'IA en épidémiologie et accidentologie	31
● Atelier 2 : Technologies de sécurisation des environnements de travail recourant à l'IA	35
● Atelier 3 : Robotique avancée recourant à l'IA	39
■ Apport de la chémo-informatique à la prévention des risques associés aux usages de substances chimiques	43
■ Focus thématiques	47
● Encadrement juridique des usages des technologies de l'information et de la communication en milieu professionnel	49
● Un aperçu des initiatives d'encadrement éthique des usages de l'IA	57
● Les machines intégrant de l'IA et la réglementation	61
■ Annexes	63
● Fiche 1 : Avancées technologiques de l'IA	95

- Fiche 2 : Disponibilité des ressources pour la production et le fonctionnement des dispositifs d'IA 83
- Fiche 3 : Disponibilité des ressources humaines nécessaires au déploiement de l'IA (hypothèses seules) 99
- Fiche 4 : Acteurs et dynamique de la diffusion de l'IA, processus de normalisation..... 101
- Fiche 5 : Acceptabilité sociale générale (citoyen) de l'IA 119
- Fiche 6 : Règlementation sur l'IA et les données 137
- Fiche 7 : Acceptabilité de l'IA dans le monde du travail : employeurs/ salariés 149
- Fiche 8 : Les questions éthiques sur l'acceptabilité en matière de prévention de l'usage de l'IA dans le monde du travail..... 159
- Fiche 9 : Croissance économique et géographie de la production..... 175
- Fiche 10 : Démographie du travail (vieillesse, qualification, formation...) 183
- Fiche 11 : Organisation du travail et statut des travailleurs..... 195
- Fiche 12 : Automatisation, cobotisation et équipements du travailleur ..209
- Fiche 13 : Responsabilités et modes de gestion de la prévention 219



Introduction

Parmi les 9 principes généraux de prévention qui guident l'action des acteurs de l'entreprise en matière de préservation de la santé et de la sécurité des travailleurs, figure en cinquième position celui-ci : tenir compte de l'état d'évolution de la technique¹. Le rythme et le foisonnement des innovations rendent cette tâche particulièrement ardue, d'autant plus qu'il s'agit également de respecter les autres principes généraux et notamment le deuxième : évaluer les risques qui ne peuvent pas être évités. Une évaluation des risques rigoureuse nécessite bien souvent du temps, et place le décideur dans une position délicate face aux innovations technologiques. Il s'agit pour lui d'arriver à les utiliser afin d'améliorer la productivité et les conditions de travail, tout en s'assurant qu'elles n'engendrent pas de nouveaux risques.

En matière de technologies de l'information et de la communication (TIC), le domaine de l'intelligence artificielle (IA) mobilise énormément de moyens et d'attention ces dernières années. Que ce soit à des fins d'amélioration de la productivité, dans l'optique de diminuer la vulnérabilité des chaînes d'approvisionnement ou pour des raisons géopolitiques de relocalisation d'activités dans des secteurs stratégiques, le fait de pouvoir disposer de systèmes d'IA est perçu comme un atout qu'il convient de développer, y compris pour la gestion des risques professionnels.

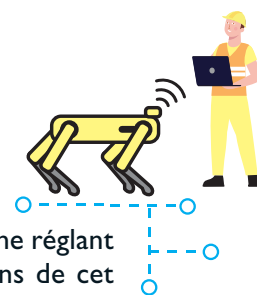
Les acteurs de la santé et sécurité au travail s'intéressent naturellement aux conséquences que peut avoir l'introduction de systèmes d'IA en milieu professionnel. Il s'agit généralement pour eux d'essayer d'évaluer les risques que peuvent faire peser ces dispositifs sur la santé physique et mentale des travailleurs, et d'émettre des préconisations de prévention. L'INRS prend logiquement sa part dans ce travail qui relève pleinement de ses missions.

Le parti pris de l'exercice de prospective, dont les résultats sont présentés dans ce document, est de se focaliser sur les usages possibles des dispositifs d'IA à des fins de protection de la santé et de la sécurité au travail à un horizon d'une douzaine d'années. En pratique, on verra que certaines conclusions vont au-delà de ce cadre initial, mais il a été jugé pertinent de les conserver. Ces technologies sont d'ores et déjà utilisées sur certains lieux de travail, à un niveau encore marginal, mais qui soulève déjà des questions.

La démarche suivie reprend les fondamentaux de la pratique de la prospective à l'INRS, à savoir un travail collaboratif et pluridisciplinaire, impliquant de nombreux experts internes et externes à l'institut ; l'application d'un cadre méthodologique permettant de donner à voir différents futurs possibles, tous cohérents, pas toujours souhaitables ; un approfondissement de certains enjeux et l'extraction de messages clés destinés aux acteurs de la prévention, au premier rang desquels figurent les partenaires sociaux siégeant au conseil d'administration de l'institut.

1. Article L4121-2 du Code du travail : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033019913/

Intelligence artificielle : de quoi parle-t-on ?



Il existe de très nombreuses définitions de l'intelligence artificielle (IA), aucune ne réglant totalement la question, car la discipline a des contours flous. Pour les besoins de cet exercice, le groupe de travail a retenu trois définitions (proposées dans la fiche variable I), qui sont compatibles entre elles. Deux sont des définitions par intension.

- Celle proposée par la Commission d'enrichissement de la langue française et publiée au Journal officiel du 9 décembre 2018 :

« Champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines. »

- La seconde est celle publiée par la Commission européenne, la même année :

« Même s'il n'y a pas de définition consensuelle, le terme « IA » fait référence à des systèmes qui exhibent des comportements intelligents en analysant leur environnement et en effectuant des actions avec un certain degré d'autonomie, pour atteindre des objectifs spécifiques. Les systèmes d'IA peuvent être purement logiciels, agissant sur le monde virtuel. Des exemples sont les agents conversationnels, les logiciels d'analyse d'images, les moteurs de recherche et les systèmes de reconnaissance de visages. D'autres systèmes d'IA sont ceux où l'IA est embarquée dans des appareils, comme les robots avancés, les voitures autonomes, les drones ou les applications de l'Internet des Objets. En contraste avec les systèmes d'IA logiciels, de tels systèmes perçoivent leur environnement grâce à des capteurs, et agissent dessus ou s'y déplacent, exigeant donc des mesures de sécurité plus robustes. »

- La troisième relève d'une autre approche consistant à définir l'IA en extension, par la somme des diverses sous-disciplines qui la composent : apprentissage automatique, raisonnement automatique, traitement de la langue naturelle, vision artificielle, représentation des connaissances, etc. C'est le choix fait par l'Académie des Technologies dans son rapport « Renouveau de l'intelligence artificielle et de l'apprentissage automatique », également publié en 2018² :

« – Perception, un domaine très vaste contenant la reconnaissance des images, des formes et des sons (et de toutes formes de signaux produits par des « senseurs »). La reconnaissance des images est le domaine le plus important (on parle de machine vision) à cause de ses applications et de par les progrès spectaculaires de ces dernières années, cités en introduction, et liés au deep learning ;

– Traitement du langage naturel, sous sa forme écrite ou orale (ce qui implique de le coupler avec la perception de la parole) ;

– Planning et navigation, qui pourraient être étendus aux domaines de la résolution de problèmes formalisés dans l'univers de la recherche opérationnelle ;

– Représentation de connaissances, qu'il s'agisse de les manipuler, de les rechercher ou de les combiner. Ce domaine a fortement évolué avec le développement des mégadonnées (Big Data) ;

– Raisonnement logique lié à la forme symbolique de l'intelligence artificielle, rendue célèbre dans les années 80 par les systèmes experts. »

Dans ces trois définitions complémentaires, on retrouve notamment les capacités données à des machines (matériel et logiciel) de réaliser des tâches demandant de

2. https://academie-technologies-prod.s3.amazonaws.com/2018/04/06/13/49/30/183/Rapport_IA_DEF.pdf

l'intelligence quand elles sont exercées par des humains, des exemples d'applications de ces capacités (véhicule autonome, agent conversationnel, reconnaissance d'images, etc.), et des technologies au service de ces fonctions (représentation des connaissances, raisonnement, apprentissage, planification, etc.).

Usages de l'intelligence artificielle en santé et sécurité : quel périmètre ?

Comme indiqué en introduction, l'ambition de cet exercice n'est pas de traiter de manière exhaustive l'ensemble des liens entre IA et S&ST. Il s'agit d'explorer les usages des systèmes d'IA spécifiquement dédiés à la prévention des risques professionnels. Ce périmètre n'est pas simple à définir.

La première difficulté est de définir quelles innovations technologiques mobilisent ou non de l'IA. Sur ce point, comme évoqué ci-dessus, le groupe de travail a retenu une définition assez englobante.

La seconde est de s'accorder sur les usages à considérer, car certains développements peuvent être motivés par des objectifs autres que la préservation de la santé au travail (amélioration de la qualité, réduction des consommations de matières premières ou d'énergie), mais avoir des conséquences favorables pour celle-ci. Le parti pris du groupe a été d'essayer de se concentrer sur des développements de systèmes d'IA motivés en premier lieu par des objectifs d'amélioration de la santé et de la sécurité au travail. Il peut s'agir d'usages amont, à des fins d'étude et de recherche sur les risques professionnels (par exemple, en accidentologie) ou bien de systèmes visant à supprimer des expositions de travailleurs à des dangers (par exemple, en recourant à des robots autonomes) ou de dispositifs de détection de situations à risques avant la survenue d'un dommage (par exemple, par le traitement en temps réel de données collectées par des objets connectés).

Certains usages « indirects » ne sont donc pas traités ici. C'est par exemple le cas des dispositifs visant principalement la substitution de l'humain par la machine à des fins avant tout productivistes, comme l'automatisation de tâches répétitives dans les services : les conséquences possibles (positives ou négatives) en S&ST sont importantes, mais relèvent du sujet plus global de l'impact des technologies sur le travail, qui a déjà été traité dans d'autres travaux de prospective conduits par l'INRS³. Les développements de dispositifs de réalité virtuelle à des fins de formation à la prévention ne sont pas abordés non plus, dans la mesure où il s'agit essentiellement d'appliquer au domaine de la S&ST des technologies qui ont été initialement développées à d'autres fins ; ce sujet a par ailleurs également été évoqué dans le cadre d'un précédent exercice de prospective⁴.

Tout au long des travaux, les membres du groupe ont identifié des circonstances dans lesquelles l'introduction de l'IA est susceptible de modifier de façon significative les situations de travail et l'approche globale de la prévention des risques professionnels. Sans remettre en cause *a priori* l'utilisation de ces technologies, ces conséquences possibles imposent une vigilance et une réflexion particulières de la part des partenaires sociaux et plus généralement des organismes chargés de la prévention des risques professionnels. Certains des messages clés issus de l'exercice relèvent ces points de vigilance.

3. « Modes et méthodes de production en France en 2040 : quelles conséquences pour la santé et la sécurité au travail ? », <https://www.inrs.fr/media.html?reflNRS=VEP%203>

4. « Quelle formation à la santé et sécurité au travail en 2030 ? », <https://www.inrs.fr/media.html?reflNRS=PV%2015>

Messages clés



Messages clés

Les principaux enseignements issus de cet exercice de prospective sont rassemblés ici sous forme de messages clés afin de fournir une synthèse facilitant la lecture du document.

Un marché en plein développement

1. Les différentes avancées en cours en matière d'IA, permises par des investissements massifs de la part d'acteurs privés et publics, préfigurent le développement d'un marché conséquent dans les prochaines années. La sécurisation des environnements de travail est l'un des domaines d'usages professionnels de ces innovations.
2. Plus l'automatisation sera intelligente, plus la machine (ou l'algorithme) remplira les tâches effectuées jusqu'alors par des travailleurs. Cette automatisation peut permettre de soustraire certains travailleurs à des risques. Elle entraînera aussi une modification des tâches d'autres travailleurs vers des fonctions d'entraînement, d'accompagnement et de contrôle.
3. Il y a un enjeu pour les acteurs à favoriser le développement de systèmes d'IA qui soient compatibles avec les valeurs essentielles de l'approche européenne et française de la S&ST (approche collective, protection des données, dialogue social). La position hégémonique actuelle des géants du numérique américains et chinois pose donc question.
4. Face aux problématiques éthiques que posent ces nouvelles technologies, de nombreux organismes ont formulé des principes à respecter afin que le développement de l'IA puisse se faire au bénéfice de tous les acteurs de la société. Il est nécessaire d'identifier et de promouvoir auprès des entreprises les recommandations pertinentes pour un usage éthique des technologies d'IA en S&ST.
5. Le scénario d'un nouvel hiver de l'IA (du fait de blocages technologiques ou « murs », de crises énergétiques, de rejets sociétaux, de cybercriminalité, etc.) n'est pas exclu. Il est donc important de ne pas faire reposer toutes les avancées en S&ST sur ces solutions technologiques. La recherche de solutions de prévention efficaces ne recourant pas à ces systèmes ne doit donc pas être abandonnée.
6. L'usage des systèmes d'IA en S&ST peut se heurter au mur de l'explicabilité (notamment pour les dispositifs faisant appel à l'apprentissage profond) et donc de la compréhension des messages et décisions générés par l'IA. Le développement des usages de l'IA à des fins de prévention des risques professionnels supposera à la fois de favoriser la bonne compréhension de ces outils (vertus et limites) par les salariés et employeurs, et de favoriser l'émergence de solutions transparentes (par exemple, des systèmes d'IA hybrides : puissance de l'IA et transparence des systèmes de raisonnement logique), facilitant les débats entre parties prenantes.

Des promesses en santé et sécurité au travail

7. Les avancées de l'IA sont potentiellement porteuses d'usages divers en S&ST. Des avancées sont à attendre dans le domaine des traitements de grandes masses de données au service de l'accidentologie et de l'épidémiologie, dans la toxicologie *in silico*, dans celui de la sécurisation des environnements de travail, et dans le développement de technologies de robotique avancée.

8. Les progrès de l'IA mettant en œuvre des techniques d'apprentissage ouvrent des perspectives intéressantes, par exemple en épidémiologie et en accidentologie, sous réserve de disposer de masses de données fiables et de ne pas occulter certaines dimensions de la S&ST pour lesquelles on ne dispose pas forcément de données exploitables (dimension organisationnelle notamment).

9. L'IA ouvre aussi des possibilités de supervision d'un environnement de travail, un chantier ou un site industriel, par exemple. Au-delà de la détection et de l'alerte, on doit attendre de ces dispositifs qu'ils fournissent des informations utiles à l'élaboration de mesures de prévention durables (organisationnelles) : cela implique une exploitation par des personnes en capacité de les analyser.

10. Certaines technologies de robotique avancée (embarquant de l'IA) offrent des solutions potentiellement bénéfiques à la S&ST. C'est le cas notamment des dispositifs de téléopération et de robotique collaborative qui peuvent atténuer, voire supprimer, des expositions à des facteurs de risques. L'implémentation de ces dispositifs doit cependant faire l'objet d'une évaluation systématique permettant de garantir qu'ils ne génèrent pas de nouveaux risques (intensification du travail, perte de sens...).

Limites et points de vigilance sur les usages de l'IA en S&ST

11. La logique de rentabilisation des investissements dans ces technologies, parfois coûteuses, peut amener à positionner ces systèmes au centre de l'organisation du travail, au risque de placer le travail humain au second plan.

12. De façon générale, un usage inadapté, détourné ou l'absence d'une réflexion préalable sur l'organisation de l'intégration de ces nouvelles technologies pourrait conduire à des effets délétères en matière de santé et de sécurité au travail. La facilité apparente d'usage et d'implémentation de ces solutions « intelligentes » risque d'entraîner un effet de contentement des acteurs, les incitant à ne prendre en considération que les risques identifiés par le système d'IA sans évaluer régulièrement les risques plus organisationnels ne faisant pas l'objet d'un *monitoring* technologique.

13. L'usage de l'IA en S&ST peut conduire à développer des outils de surveillance des travailleurs et d'alerte lorsque les conditions d'un travail en sécurité ne sont pas remplies (consignes non respectées, état de santé du travailleur hors norme, etc.). Cette surveillance permanente peut générer des RPS et également conduire à une individualisation de la S&ST, et à une responsabilisation exclusive du travailleur au détriment de la mise en place par l'employeur de mesures de prévention collectives.

14. Une attention devra être portée aux risques éventuels associés à des usages dans des dispositifs de S&ST d'algorithmes d'IA qui n'ont pas été développés spécifiquement à cette fin (librairies *open source*, produits standards vendus sur étagères).

15. Les technologies d'apprentissage profond (*deep learning*) reposent sur l'entraînement d'un modèle sur un jeu de données. L'utilisation de l'IA dans les entreprises va donc impliquer la collecte et le stockage de nombreuses données. Dès lors qu'un usage en prévention est envisagé, se pose la question de la constitution, de la qualification et de l'étiquetage du jeu de données utilisé lors des phases d'apprentissage dans les systèmes supervisés. Une attention particulière doit être apportée aux jeux de données afin qu'ils correspondent aux domaines d'applicabilité pouvant varier selon les activités et situations de travail, et qu'ils ne soient pas biaisés.

16. Les accidents du travail surviennent fréquemment lors de situations atypiques par rapport au déroulement classique d'un processus de production : situations dégradées,

pannes, opérations de maintenance... Ces situations sont souvent imprévues, donc non anticipées dans les procédures, ce qui les rend particulièrement dangereuses. Elles constituent ainsi une limite possible à l'entraînement des systèmes d'IA, les jeux de données nécessaires n'étant pas en mesure d'intégrer de manière exhaustive la palette des aléas qui peuvent survenir dans de nombreux contextes de travail (chantiers, grands sites industriels, travaux sur la voie publique...).

Pistes d'actions

17. Du fait des opportunités offertes par ces nouvelles technologies, ainsi que des risques potentiels qu'elles sous-tendent, la formation des acteurs de la prévention (employeurs, représentants du personnel, préventeurs) constitue un enjeu primordial de l'intégration à venir de l'IA aux équipements de travail et aux solutions de prévention. Ces formations devront permettre une bonne compréhension du mode de fonctionnement de ces outils, des enjeux éthiques, du cadre réglementaire qui les régit, des possibilités de les piloter, des risques qu'elles peuvent représenter, mais aussi l'acquisition de méthodes permettant la définition des besoins, la rédaction de cahiers des charges, et l'intégration des dispositifs dans l'entreprise. L'acculturation des acteurs du dialogue social, aussi bien au niveau des branches professionnelles que des entreprises (notamment dans les TPE-PME) est indispensable afin qu'ils soient en mesure d'appréhender et de discuter en amont les transformations des modes et méthodes de travail que ces nouveaux systèmes induisent.

18. Le développement et la mise sur le marché de dispositifs utilisant des techniques d'IA présentés comme des outils de prévention doivent être assurés par des personnes disposant de solides compétences en matière de santé et sécurité au travail. Au-delà des formations à prévoir dans le cadre de l'entreprise, des modules de formation doivent également être mis en œuvre dans les cursus d'écoles de management et d'ingénieurs afin de sensibiliser les futurs commanditaires et développeurs de systèmes d'IA aux opportunités et risques qu'amènent avec elles ces nouvelles technologies en matière de S&ST.

19. Il convient de promouvoir auprès des entreprises les démarches reposant sur l'expérimentation et l'évaluation qui permettent de mesurer en conditions réelles les conséquences des nouveaux systèmes sur l'organisation de l'entreprise et le travail des opérateurs, et de conserver une possibilité de revenir en arrière.

20. Les normes et réglementations encadrant l'IA se développent (AI Act). Il est donc essentiel que les principes de la S&ST soient portés au sein des instances d'élaboration. C'est notamment vrai au niveau des réglementations européennes, mais également dans les comités de normalisation.

21. Une réflexion collective (de type conférence de consensus) devra être menée sur la question des données utilisées dans des dispositifs d'IA touchant à la S&ST. Il s'agira notamment de définir des règles pour la constitution des jeux de données, l'encadrement de leur utilisation en fonction des domaines d'application. Au-delà des utilisateurs, il s'agira d'impliquer les partenaires sociaux, des personnalités qualifiées (experts, philosophes spécialistes de l'éthique, juristes...).

22. De façon générale, les avancées de l'IA offrent des perspectives de progrès pour la prévention des risques professionnels. Elles entraînent aussi, comme toute modification, certains risques. Le développement et la diffusion d'outils méthodologiques permettant de guider les acteurs face à ces innovations constituent un enjeu fort pour les organismes de prévention.

Méthodologie

Un exercice en cinq phases

Cet exercice a été mené par un groupe de travail pluridisciplinaire composé d'experts et accompagné sur le plan méthodologique par le cabinet Futuribles. Il a été conduit en cinq phases.

Première phase : Définition collective du sujet et de son périmètre

Du fait de la technicité du sujet traité, la définition des concepts a constitué une étape importante au moment d'engager cet exercice. Une note de cadrage a servi de base à la première réunion du groupe de travail visant à clarifier les concepts et délimiter le sujet. L'horizon temporel de l'exercice de prospective a également été défini durant cette phase.

Deuxième phase : Analyse du sujet

L'analyse des facteurs clés a permis d'établir une liste de 13 variables réparties au sein de trois composantes : évolution de l'offre en IA, acceptabilité des usages possibles de l'IA, travail et prévention. La rédaction de chacune des fiches variables a été confiée à un membre du groupe, puis discutée en réunions.

Troisième phase : Élaboration des scénarios

Sur la base des hypothèses d'évolution des différentes variables, quatre scénarios ont été établis. Ils ont été construits de manière contrastée afin d'explorer un champ des possibles suffisamment large.

Quatrième phase : Étude de cas d'usage

Trois ateliers ont ensuite été organisés afin de poursuivre la réflexion sur les usages possibles de l'IA et de ses conséquences potentielles en prévention aussi bien pour les acteurs de la santé et la sécurité au travail que pour les conditions de travail des salariés. Ces ateliers ont été conduits sur la base des différents contextes envisagés dans les scénarios, en essayant de considérer toutes les conséquences tant favorables que défavorables dans trois champs d'usages possibles : épidémiologie et accidentologie, surveillance des lieux de travail et des travailleurs, robotique avancée.

Cinquième phase : Élaboration des messages clés

Une dernière série d'échanges avec le groupe de travail et des experts extérieurs au groupe a permis de stabiliser une liste de vingt messages clés, principaux enseignements issus de cette démarche.

Scénarios prospectifs : l'IA au travail en 2035

Les quatre scénarios présentés ici ont été construits par le groupe sur la base de l'instruction des treize variables clés. Ils donnent à voir des évolutions possibles et contrastées du contexte de développement des usages de l'intelligence artificielle en milieu de travail à l'horizon 2035. Ils constituent avant tout un outil de projection et d'aide à la réflexion.

Scénario 1 : Les géants du numérique imposent leurs solutions et leur vision

Synthèse

Dans ce scénario, l'exubérance technologique se poursuit dans un cadre de compétition pour la maîtrise de l'intelligence artificielle, mais les acteurs principaux en sont les géants du numérique, essentiellement les AMAMA¹ à l'Ouest et les BATX² en Chine. Leur puissance dépasse largement le champ du numérique et, de fait, ils contrôlent l'essentiel de l'innovation et dominent des pans entiers de l'économie mondiale. Dans un contexte de rivalité internationale systémique, les États doivent composer avec ces acteurs et s'appuyer sur eux pour maintenir leur puissance comme pour assurer les fonctions clés de leur souveraineté. La réglementation est morcelée entre les États et largement influencée par ces géants. Les utilisateurs acceptent ces normes par commodité et parce qu'elles sont devenues indispensables. L'automatisation progresse et la surveillance s'installe comme outil privilégié de la sécurité au travail. Celle-ci justifie un contrôle des activités de chacun dans un contexte de collaboration étroite entre les hommes et les machines.

1. Alphabet, Meta, Amazon, Microsoft, Apple

2. Baidu, Alibaba, Tencent, Xiaomi

Scénario

Dans un contexte de compétition internationale, l'intelligence artificielle poursuit son développement technologique et sa diffusion commerciale. Le modèle actuel qui repose sur les systèmes apprenants à partir de masses de données montre son efficacité dans de nombreux domaines. De la logistique à la tarification des services, de l'économie de plateformes à l'industrie 5.0, leur utilisation se généralise. Elle est facilitée par l'offre qui l'intègre dans l'informatique en nuage (*cloud computing*), champ désormais dominé par les géants du numérique, accentuant une tendance déjà notable en 2021, année où quatre entreprises détenaient 70 % de ce marché. Ces acteurs consentent des efforts de recherche considérables sur le sujet et leur puissance financière leur permet de contrôler les innovations extérieures, par des rachats stratégiques de licornes ou la neutralisation de concurrents émergents. Dans le même temps, ils étendent leur contrôle aux infrastructures de télécommunications essentielles à leurs activités, que ce soit les câbles sous-marins ou les satellites en orbite basse.

Les États ont conscience de la montée en puissance de ces acteurs et des tentatives de sanctions financières ont été appliquées sans effet. Les promesses de modération des réseaux sociaux se succèdent sans résultats notables non plus. La compétition géopolitique prime, les États-Unis accordent un soutien sans faille à leurs champions nationaux et évitent de durcir la réglementation ; le Parti communiste chinois s'appuie sur le numérique pour le contrôle de la population par le crédit social, et l'Europe peine à faire vraiment appliquer des réglementations que les jeux de la mondialisation affaiblissent. Bref, les géants du numérique jouent sur les rivalités et les divergences des intérêts nationaux, y compris par le chantage³ pour contenir les tentatives de régulation. De surcroît, les États composent avec ces géants dans de nombreux domaines comme la sécurité, le spatial, la logistique militaire, ou même l'exploitation de masses de données publiques, comme l'illustre le cas des données de santé françaises dont le stockage a été confié à Microsoft. Aussi, les réglementations sont influencées par les géants du numérique et n'ont qu'une portée territoriale limitée, morcelant et complexifiant les règles établies par les pouvoirs souverains.

Les ressources ne sont pas un frein. Les financements se portent aisément sur un secteur dont les promesses restent très attractives, y compris parce qu'une partie de son dynamisme dépend de la captation d'une rente au détriment des activités traditionnelles. Cette attractivité joue également en faveur de la disponibilité des ressources humaines dont le marché s'est mondialisé : des spécialistes provenant de toutes les régions du globe se voient offrir des facilités de visa et d'établissement ou d'études dans les pays leaders. Et même si des tensions apparaissent ici ou là, des dispositifs de formation se multiplient, facilités par des plateformes de e-learning mises en place par les géants du numérique.

En matière d'environnement, les pénuries de certaines ressources naturelles (terres ou métaux rares) accroissent les coûts, mais des solutions alternatives sont développées. Le domaine le plus sensible concerne la consommation d'énergie électrique. La généralisation de l'usage de l'IA s'accompagne d'une croissance importante des consommations, et le numérique devient l'un des grands secteurs consommateurs d'électricité. Là encore, les grandes entreprises promettent des technologies moins gourmandes en métaux rares et des *data centers* aux faibles émissions de gaz à effet de serre. Parallèlement, un marché très actif de régulation de la pollution s'est mis en place à travers la négociation de quotas d'émissions. À terme, les perspectives de la fusion nucléaire entretiennent le mirage d'une énergie inépuisable et peu coûteuse. La dynamique est donc celle d'un développement effréné sous l'impulsion d'un secteur privé dont la

3. Par exemple, la menace de Meta (Facebook) de quitter l'Europe.

croissance annule les réductions d'émissions et de consommations obtenues grâce aux innovations.

La question de l'acceptabilité de ces technologies se pose, compte tenu de leur caractère intrusif. Elle reste toutefois le fait de groupes restreints, et si épisodiquement des mobilisations plus importantes ont lieu, cela reste minoritaire, car les utilisateurs sont partagés entre les risques qui pèsent sur leur vie privée et les services apportés qui leur sont indispensables dans la vie courante. De plus, ces entreprises revendiquent une « neutralité politique » sur les informations qu'elles détiennent et se targuent d'agir pour le bien⁴ ; elles inspirent moins de méfiance que les États.

Le monde du travail et le grand public se satisfont de cette situation dans laquelle les géants du numérique se conforment à des règles ou principes qu'ils ont largement définis. Dans les domaines critiques (sécurité, médecine, justice, recrutement, etc.), les processus maintiennent une « supervision humaine » des décisions, même si cela relève souvent d'une mesure cosmétique.

L'automatisation se développe fortement dans tous les secteurs, poussée par les Big Tech et encouragée par les États. Elle est mise au service de l'amélioration des techniques de culture sans désherbants et un tissu de services se constitue autour des technologies numériques agroalimentaires. La France se réindustrialise en recourant aux avancées technologiques, l'industrie 5.0 bénéficie de gains de productivité importants. Les activités logistiques sont optimisées. Par exemple, la chaîne de tâches allant de la commande en ligne à la livraison au consommateur final est totalement assurée par des équipements autonomes. Dans les cas où la main-d'œuvre humaine reste indispensable, des orthèses mécatroniques lui apportent leur aide, tandis que les signaux biométriques analysés garantissent sa sécurité. Cela s'étend aux services financiers et assuranciers ou administratifs. L'industrie française est de plus en plus intégrée et dépendante de quelques grands opérateurs internationaux⁵, les seuls capables financièrement et technologiquement d'offrir des services compétitifs en matière d'intelligence artificielle.

Mais des secteurs entiers échappent partiellement ou totalement à ces évolutions. On enregistre une forte polarisation du marché du travail : d'un côté, des emplois conceptuels, bien rémunérés, occupés après des formations longues et coûteuses et où des tensions se manifestent ; de l'autre, des emplois d'exécution aux salaires faibles et aux conditions précaires qui se traduisent par une rotation rapide des travailleurs. Ceux-ci, en effet, alternent travail déclaré et travail au noir, y compris dans les emplois en tension au début de la période (hôtellerie-restauration, transports, etc.). L'économie des plateformes, dopée par les géants du numérique, poursuit son essor et s'étend aux cols blancs. Elle affecte principalement les emplois moyennement qualifiés qui sont les plus exposés aux délocalisations ou à l'automatisation, avec les risques de déclassement que cela comporte pour les intéressés. Les plateformes soutenant les espaces virtuels (*metavers*) sont suffisamment avancées pour instrumenter leur développement : après leur succès populaire pour les activités sociales et commerciales, elles deviennent des outils essentiels pour les télétravailleurs, qui peuvent y effectuer une part croissante de leurs activités. Les systèmes d'IA y trouvent une application naturelle et à moindre risque que dans le monde industriel réel, où les limites inhérentes à la technologie sont toujours présentes.

La fragmentation du travail et les cohabitations de plus en plus fréquentes des travailleurs avec les machines transforment la nature des risques d'accidents et de maladies. Dans le contexte d'euphorie technologique, les législateurs cherchent à favoriser les innovations. Et la réglementation évolue vers un allègement de la responsabilité de l'employeur en

4. Voir par exemple l'initiative de Data for Good de Facebook : <https://dataforgood.facebook.com/>

5. Bleu, nouvel opérateur pour un Cloud de confiance et qui résulte d'une alliance entre Orange et Cap Gemini, opère sur la base de technologies développées par Microsoft.

matière de réparation de préjudices. Les contraintes sont réduites *a minima* (respect des normes, auto-contrôles et information des personnels). Les acteurs privés affichent leurs normes en matière d'intelligence artificielle et les comités d'éthique se multiplient pour garantir des protocoles adéquats autour de caractéristiques rassurantes (robustesse, équité, loyauté, sécurité, explicabilité...). Elles peuvent cependant s'avérer parfois éloignées de la réalité du travail. Des labels en la matière concourent à l'acceptabilité sociale des usages de l'IA pour la prévention et la &SST. Dans la réalité, de nombreux dispositifs assurent un suivi de l'activité des travailleurs, pour des objectifs de rentabilité, de prévention des risques, et de protection de l'entreprise en cas de contentieux ultérieurs.

Scénario 2 : Les États garantissent un cadre pour l'intégration de l'IA

Synthèse

Le développement de l'IA sans maîtrise a poussé les États européens à travailler sur un règlement commun afin de pouvoir en encadrer l'écosystème et les principes éthiques. Au cours de la période, la multiplication des problèmes – entraînant des préjudices pour les citoyens, les entreprises et les travailleurs – les a poussés à durcir cet encadrement dans un contexte de montée des préoccupations environnementales. Les États, afin d'orienter l'allocation des ressources, ont décidé de ne développer que des systèmes d'IA sobres, répondant à des critères élevés (notamment autour du concept d'intérêt général), sur des secteurs non critiques et sous supervision humaine. Cela passe par un meilleur contrôle des fabricants, par le développement de pôles d'expertises européens et nationaux, mais également par des expérimentations qui permettent de démontrer l'innocuité de l'IA et des mesures de contrôle et d'audit. L'IA au travail se développe donc dans un environnement assez encadré, permettant une implémentation, dès lors que l'intérêt et la non dangerosité des dispositifs ont été approuvés.

Scénario

Dans les années 2020, un encadrement réglementaire des usages de l'IA se met progressivement en place

Les États européens – et en particulier la France – sont acteurs d'une régulation assez forte, qu'il s'agisse de cibler les usages de l'IA dans l'espace européen ou d'obliger les grands fabricants internationaux à plus de transparence quant à l'application de leurs dispositifs d'IA, et à l'impact que ces derniers peuvent avoir sur les utilisateurs. Le règlement européen sur l'IA s'est vu assez vite adopté en 2023 : son cadre limite les usages pouvant porter atteinte à la liberté des individus, tout en permettant aux États de garder leur souveraineté dans le développement d'outils mobilisant l'IA pour des usages régaliens. Les critiques de ce texte, trop généraliste et pas forcément adapté à certains secteurs, n'ont pas spécialement été entendues dans un premier temps. Pour certains acteurs (syndicats, chercheurs), il s'avère insuffisant pour prévenir les effets négatifs de certaines applications sur la santé des travailleurs. Ils pointent notamment les dispositifs de pilotage de l'activité qui tendent à intensifier les rythmes de travail tout en réduisant les marges de manœuvre des travailleurs.

En introduisant des obligations d'audits ou des évaluations des risques concernant les systèmes d'IA, d'autres grands textes européens, notamment développés autour de la question des données, ont un impact sur les grands acteurs de l'économie numérique, qui sont également de grands pourvoyeurs de services fortement dépendants d'algorithmes d'IA.

À la recherche d'une souveraineté européenne en matière d'IA

Cette volonté d'encadrement du développement et des usages de l'IA va de pair avec un choix de s'affranchir du bras de fer sino-états-unien, et notamment d'assurer l'indépendance stratégique et la sécurisation de l'approvisionnement de l'industrie européenne. La puissance normative de l'Europe ne suffit plus, il devient nécessaire de développer une expertise européenne en matière d'IA. Les États ont décidé à la fois de reprendre en main le secteur par une régulation accrue et de dynamiser la recherche européenne, à travers :

- Le développement de filières de formation afin de créer les compétences et le savoir-faire nécessaires
- Un renforcement de la recherche européenne par la création de pôles de recherche qui se veulent particulièrement attractifs : ils visent notamment le développement de dispositifs d'IA plus sobres
- La recherche d'une moindre dépendance aux autres pôles internationaux, à travers le développement de filières de production de composants, la constitution de stocks en Europe et l'élaboration d'accords commerciaux plus équilibrés

C'est donc une sorte de relocalisation qui s'installe sous l'impulsion de l'UE. La volonté est de se détacher des géants du numérique américains : ceux-ci sont déjà sous le feu de fortes critiques aux États-Unis et les propositions de démantèlement se succèdent dans une approche anti-trust. Celle-ci a déjà convaincu les Européens, qui cherchent à valoriser les initiatives européennes et à protéger les solutions locales innovantes. Pour des raisons de souveraineté numérique, elles sont donc privilégiées face aux outils des acteurs étrangers et, dans certains domaines, les rachats de start-ups sont encadrés (défense, Cloud...).

Une approche « distribuée » est particulièrement développée en France : elle repose sur des écosystèmes impliquant des start-ups technologiques, des laboratoires universitaires et des réseaux d'accompagnement des entreprises (chambres de commerce, centres techniques, organisations sectorielles). Elle permet une adaptabilité des systèmes au plus proche des besoins des entreprises, mais également de profiter de la main-d'œuvre experte arrivant sur le marché du travail, qu'il s'agisse d'une immigration qualifiée ou de jeunes issus des filières spécialement lancées autour de l'IA.

Dans les années 2030, les impératifs climatiques conduisent à un renforcement de l'interventionnisme de l'UE, y compris en matière d'IA

Au cours des années qui suivent, dans un contexte d'instabilité géopolitique, la recherche d'une souveraineté européenne connaît une accélération liée aux difficultés d'approvisionnement en certaines ressources (énergies fossiles, minéraux) et aux dérèglements engendrés par le réchauffement climatique (sécheresses, catastrophes naturelles). Les gouvernements sont obligés d'intervenir de manière encore plus importante pour imposer des choix éclairés en termes d'allocation de moyens aux technologies. Ce virage écologique a aussi concouru à recentrer et restreindre les domaines de développement de solutions recourant à l'IA, et a également conforté la légitimité des textes l'encadrant, les justifiant. Il a fallu prioriser les développements selon

des finalités supposées positives pour l'humanité, délaissant tout ce qui a été considéré comme non viable sur le plan environnemental, non prioritaire ou accessoire. Le principe du « Règlement IA » est alors renforcé : sont interdits les systèmes pour lesquels des précautions insuffisantes ont été prises sur leur applicabilité. Il fait l'objet d'une actualisation en 2033 qui le complète sur deux volets. D'une part, en introduisant un critère de coût environnemental des dispositifs afin de dissuader le développement des solutions les moins sobres ; d'autre part, en imposant des autorisations préalables des dispositifs pour certains usages, notamment dans les domaines de la santé et du travail. Une agence d'expertise européenne est créée à cet effet.

En 2033, création de l'EA-DIR (European agency for the digital innovation regulation); agence européenne de régulation des innovations numériques.

Cette nouvelle structure a pour vocation d'apporter son expertise aux administrations chargées du contrôle de la mise sur le marché d'un certain nombre d'innovations, telles que les véhicules autonomes, les dispositifs médicaux de type cabines de téléconsultation, etc.

Cette agence dispose d'un département spécialisé sur les usages professionnels, qui traite notamment des aspects robotiques et des algorithmes de pilotage des activités humaines. Selon les usages, les dispositifs d'IA sont soumis à différents niveaux d'audits (inspirés de ceux préconisés par l'OCDE) avant d'être autorisés au sein de l'UE.

Les critères de sélection ont été fixés par les États lors de sommets tenus en réaction à des rapports mettant en évidence les dangers de certains dispositifs. Ils deviennent désormais obligatoires pour toutes les entreprises se trouvant sur le territoire de l'Union européenne ou ciblant des citoyens européens (critère de ciblage). En plus de ces critères de développement, s'ajoute la nécessité de prouver l'impact positif de l'IA : des systèmes de test des dispositifs d'IA pouvant avoir un impact sur les individus sont mis en place, avant qu'il ne soit possible de les déployer à large échelle. Cela est particulièrement vrai dans le monde du travail, dans lequel sont créés des comités de suivi et d'éthique de l'implémentation des innovations technologiques en entreprise, qui regroupent à la fois des spécialistes issus des pôles d'expertise, mais aussi tous les corps intermédiaires habituels, notamment les partenaires sociaux. La santé et la sécurité au travail est un élément déterminant pour ces comités qui veillent à bloquer les développements des dispositifs réputés néfastes, à documenter les dangers potentiels des innovations, et à encourager l'implantation des dispositifs dont l'intérêt en prévention a été démontré.

L'IA au travail, qu'il s'agisse de robotisation, de *monitoring* ou de logiciels de pilotage de l'activité, engendre une nécessaire montée en compétences au sein des entreprises, créant de nouveaux métiers et expertises pour encadrer cette utilisation (supervision obligatoire, nécessité d'explicabilité d'une décision automatique, etc.).

Scénario 3 : Développement démocratique

Synthèse

Dans ce scénario, les années 2020 voient se mettre en place des processus de contrôle démocratique par les travailleurs et citoyens, nécessaires au bon développement des IA et à leur expansion encadrée dans la vie civile et professionnelle. Dans un contexte de croissance économique mondiale générant de l'emploi dans l'industrie et les services, et permettant d'investir dans la formation, les conditions sont réunies pour que les systèmes d'IA se déploient largement dans le monde du travail, et concourent au façonnage progressif d'une maîtrise collective de ces projets technologiques. Le recours aux dispositifs d'IA est facilité via l'essor des outils *open source* et le développement de solutions très accessibles (*low code, no code*). De plus, les recherches en IA lancées depuis les années 2010 finissent par aboutir, dans les années 2030, à la conception de systèmes d'IA hybrides combinant la puissance de l'apprentissage automatique à la transparence des systèmes de raisonnement logique. En restaurant le principe éthique d'explicabilité comme clef d'appropriation, ces résultats contribuent à construire la confiance collective dans l'IA et à la mettre au service de la performance, de la santé et de la sécurité dans les organisations du travail.

Scénario

Années 2020 : développement démocratique, chartes internationales, protection des travailleurs

Des expérimentations de nouvelles applications d'IA sont lancées dans différents secteurs. Un certain nombre d'usages (dans les domaines de la santé, des loisirs, de la politique) mettent en évidence la nécessité d'un encadrement. Il s'agit notamment d'erreurs dans les domaines de la santé (prescription par des IA de traitements inadaptés) ou de la justice (décisions biaisées), d'accidents impliquant des véhicules autonomes et des cobots, et de la propagation de fausses informations (*deepfake*).

Un cyberactivisme qui met en défaut les systèmes d'IA

En s'inspirant du principe que certains développements de l'IA particulièrement sensibles n'étaient accessibles qu'à quelques entreprises et aux États, la branche banque et assurances s'était octroyé les mêmes prérogatives au nom de la protection du secret bancaire. Rapidement, les consommateurs des pays occidentaux sont passés à l'action. Guidés par des groupes de hackers éthiques, des milliers de citoyens ont multiplié les actions fortement consommatrices de techniques d'IA (identifications visuelles et digitales biaisées, opérations complexes sur de très petites sommes, utilisation d'assistants personnels pour leurs démarches auprès des banques, etc.) afin d'embouteiller les systèmes, de les mettre en défaut et d'augmenter leurs coûts d'exploitation. Rapidement, banques et assurances sont venues à résipiscence.

Cette fronde a inspiré les citoyens russes et chinois, qui l'ont appliquée dans tous les domaines mettant en œuvre l'IA. Face aux dysfonctionnements qui en ont résulté, les États russes et chinois ont adhéré au GPAI (partenariat mondial sur l'IA).

Les pouvoirs publics (notamment européens), sous pression de la société civile, imposent que la conduite de tout nouveau projet s'accompagne non seulement de comités d'éthique, mais aussi associe les parties prenantes à la gouvernance des projets (de leur conception à leur mise en œuvre et leur suivi). Les représentants des travailleurs sont embarqués dans ces processus. Les critiques ne sont pas repoussées, stigmatisées ni rejetées : elles sont incorporées au processus pour soutenir, consolider et affiner des développements. L'évaluation sociale de ces technologies est explicitement recherchée. L'examen critique et démocratique est institutionnalisé et sa valeur reconnue. L'acceptabilité sociale des IA est ainsi renforcée et la qualité des projets s'en trouve sensiblement améliorée. Les expérimentations contrôlées d'IA sont un levier de renouvellement et de renforcement de l'exercice démocratique dès lors qu'elles créent les conditions pour augmenter conjointement les capacités des IA, mais aussi des intelligences individuelles et collectives. Les technologies d'IA peuvent ainsi être mobilisées pour des enjeux socio-économiques : écologie (maîtrise des consommations d'énergie), lutte contre la fraude fiscale, inclusion... Ces expérimentations participent d'un développement de connaissances et de compétences, et s'accompagnent d'une nécessaire mise en culture partagée des données exploitées pour développer et faire fonctionner ces IA de qualité.

Pour accompagner et encadrer ces développements, les acteurs privés et publics nationaux (CNIL, ministère de l'Économie) et internationaux (UNESCO, OCDE) s'accordent sur des règles en termes d'éthique, et un cadre normatif global est adopté pour favoriser la protection des individus au travail et des collectifs, sans freiner le développement de la recherche, de l'innovation et d'applications d'IA.

Le GPAI (Global Partnership on AI – Partenariat mondial sur l'IA), qui regroupe peu à peu toutes les nations majeures de l'IA et l'Union européenne, vise le développement et l'utilisation responsable d'IA fondées sur les droits de l'homme, l'inclusion, la diversité, l'innovation et la croissance économique. Il soutient des projets les mettant en œuvre, produit des chartes pour une IA responsable, sur la gouvernance des données nécessaires à l'IA, sur la commercialisation des innovations en IA et sur sa bonne utilisation dans le cadre du travail. Les quelques pays qui n'avaient pas souhaité s'associer au GPAI lors de sa création (en particulier la Chine et la Russie) adoptent également ces chartes sous la pression de leurs entreprises et de leurs citoyens pour accéder aux marchés occidentaux.

L'IA accompagne les transformations du travail dans le respect des acteurs

Une attention particulière est portée au maintien d'une protection sociale pour tous les travailleurs, y compris les indépendants : l'objectif est de limiter les situations de précarité. Des efforts importants sont conduits en matière de formation pour faire évoluer les compétences numériques des travailleurs, les entreprises du secteur y contribuent. Parallèlement, les nouvelles formes d'organisation telles que les coopératives, les entreprises libérées, etc. prennent de l'ampleur et permettent aux travailleurs de gagner en autonomie dans un contexte d'affaiblissement des positions de management. Cette ambition a permis de passer les crises successives (crises sanitaires et perturbations des chaînes de valeurs en raison de conflits et de catastrophes naturelles) sans effondrement économique. L'ensemble des acteurs a déployé des moyens d'achats, de travail et de loisirs efficaces dans ces environnements grâce à l'IA.

Nouveau dynamisme, croissance portée par la réindustrialisation et l'essor des services

Sous l'influence des contraintes environnementales, économiques et sociales, on assiste à une revitalisation de l'industrie française, notamment par le rapatriement de certaines activités, la poursuite de l'automatisation, et la transition environnementale de l'économie (transports, énergies vertes, économie de la fonctionnalité). En 2024, l'implantation d'industriels comme TSMC ou Intel redynamise la filière électronique en Europe et renforce son autonomie vis-à-vis de l'Asie. Les innovations liées à l'IA contribuent à abaisser les coûts de production. Cela ne va pas sans poser de problèmes pour certains métiers peu qualifiés fortement automatisés : préparateur de commande dans la logistique, grande distribution... Dans ces secteurs, les transformations du travail induites par l'introduction de l'IA – y compris lorsqu'elles visent à réduire les pénibilités – sont parfois mal acceptées par les travailleurs, qui voient leur métier changer.

En matière de prévention et de S&ST, employeurs et travailleurs européens bénéficient d'un cadre clair et transparent. En France, l'IA est intégrée dans le cadre d'une obligation de résultat, à l'instar d'autres technologies numériques utilisées en milieu de travail. Dans cette perspective, la responsabilité de l'employeur est dite « objective », à savoir engagée *de facto* à partir du moment où un accident du travail survient. De plus, l'intégration de systèmes d'IA fait l'objet de négociations collectives obligatoires, à l'échelle des branches professionnelles ou des entreprises.

Les fonds alloués aux formations initiales et continues permettent de répondre aux besoins de main-d'œuvre qualifiée et les carrières peuvent se prolonger, dans la mesure où les innovations réduisent les pénibilités physiques les plus lourdes. Des investissements publics et privés importants permettent le développement de filières de formation initiale et continue en IA. Un vivier de main-d'œuvre se développe en France, ce qui permet une meilleure distribution des solutions IA, y compris dans de petites entreprises. Par conséquent, ces solutions sont moins standardisées et plus proches des besoins spécifiques des entreprises.

Années 2030 et suivantes, une IA sobre se déploie massivement

Sur le plan technique, d'importants progrès sont réalisés pour permettre la croissance des usages de l'IA. Cela passe par le recours à des composants moins gourmands en énergie et nécessitant moins de métaux rares, d'importants progrès en matière de récupération et de recyclage des métaux rares, une optimisation du partage des capacités de calcul, et le développement d'un Cloud souverain faiblement émissif en gaz à effets de serre. Ces développements sont en partie possibles grâce à des progrès en électronique, comme les circuits neuromorphiques, le *in-memory computing* et le développement de l'IA embarquée, qui permet la réduction de consommation des systèmes et le désengorgement du Cloud. Néanmoins, tous ces progrès n'ont pas permis de réduire la consommation énergétique globale : en raison d'un effet rebond, les systèmes – plus efficaces énergétiquement – permettent une augmentation des usages.

Les plateformes pour le codage et l'utilisation des outils d'IA se déploient largement et stimulent le tissu économique dans le domaine du numérique. Des librairies *open source* et des services IAaaS (Intelligence Artificielle as a Service) se développent et offrent des applications variées dans le domaine de l'énergie, l'agriculture, la micro-mobilité. Ce déploiement effréné est accéléré par les progrès de la programmation *low code* et *no code*, qui rend la programmation accessible au plus grand nombre. Cette prolifération de l'IA pouvant entraîner son utilisation par des organisations terroristes et des réseaux de

cybercriminels, certains développements particulièrement sensibles ne sont accessibles qu'à quelques grandes entreprises habilitées par les États et à l'armée.

Les conditions sont réunies pour une intensification de la robotisation dans tous les secteurs. En agriculture, un écosystème dense de concepteurs, constructeurs, concessionnaires de robots émerge, et un tissu de services se constitue autour des technologies numériques. Dans l'industrie, la simplification des systèmes de pilotage et la baisse des coûts permettent une automatisation proche de 90 %, les 10 % restants étant souvent assurés à distance. Les personnes restant dans l'atelier sont aidées dans leurs efforts par des exosquelettes mécatroniques, leurs signaux biométriques analysés par leurs vêtements intelligents. Dans les secteurs de la logistique et de la grande distribution, toute la chaîne allant de la commande en ligne à la livraison au consommateur final est progressivement assurée par des équipements autonomes. Dans le secteur tertiaire, le perfectionnement des outils de traitement automatique du langage permet d'automatiser de nombreuses tâches, de s'affranchir des problèmes de traduction et de déployer des environnements plus inclusifs, tant pour les usagers que pour les travailleurs. Les dossiers demandant une décision sont instruits par des systèmes d'IA, mais l'humain garde la maîtrise de la décision et en assume la responsabilité dans un certain nombre de domaines encadrés : santé, assurance, banque...

Les recherches sur l'hybridation neuro-symbolique de l'IA aboutissent à des systèmes capables de dépasser le niveau perception, pour arriver au niveau système 2 de Kahneman : capacité à conduire des raisonnements à partir des données et des connaissances disponibles, avec une certaine dose de sens commun, et à expliquer leur fonctionnement à leurs concepteurs, utilisateurs et certificateurs. Ces systèmes demandent beaucoup de développement et de ressources humaines (conduite autonome, robotique industrielle...). Ils sont donc mis en œuvre avec parcimonie, mais donnent totalement satisfaction dans ce cadre.

Au final, l'IA acceptée

Plus explicables grâce à ces évolutions technologiques, basées sur des données plus traçables, plus discutables grâce à des évolutions institutionnelles et culturelles, certaines formes et applications d'IA deviennent acceptables et praticables pour un nombre grandissant d'employeurs et de travailleurs. Un grand travail de pédagogie et de montée en compétence a été nécessaire pour élaborer des processus de conception et des instances de contrôle démocratiques efficaces, dont la saisine et la gouvernance permettent d'installer la confiance nécessaire à l'innovation socio-technique.

Scénario 4 : Déclin de l'IA

Synthèse

En début de période, les usages des systèmes d'intelligence artificielle se développent dans tous les domaines professionnels. Poussée par les progrès technologiques, la numérisation généralisée de la société et les nouvelles organisations de travail, l'IA est majoritairement bien acceptée dans le monde du travail. Depuis 2022, elle est considérée comme un atout pour les employeurs (automatisation, productivité, qualité, etc.) et pour les travailleurs (pénibilité, sécurité, etc.). Cette considération est principalement basée sur les promesses que les systèmes d'IA apportent. Progressivement, la déception face aux applications de terrain, les failles de ces systèmes qui provoquent des incidents, accidents ou crises, génèrent un rejet des systèmes d'IA dans le monde du travail. À partir de 2030, ce rejet amène à un déclin de cette technologie et de ses usages professionnels.

Scénario

Années 2020 : la croissance de l'IA se poursuit

Depuis 2010 et les progrès du *deep learning*, les usages de l'IA sont de plus en plus répandus et s'accroissent dans de nombreux domaines. La diffusion des services et des objets intelligents est assurée par des géants mondiaux du digital aux moyens financiers colossaux. Des initiatives d'encadrement et de réglementation émergent, mais peinent à aboutir et se confrontent à la globalisation de l'offre et la compétition géopolitique, qui fait des systèmes d'IA des enjeux stratégiques. Ces projets de régulation de l'IA mettent en avant les notions d'éthique, de protection de la vie privée, de transparence des algorithmes et de responsabilité des acteurs. Cependant, l'acceptation des systèmes d'IA dans la société reste bonne et le ratio bénéfice/risque perçu est toujours très favorable. Cette croissance des systèmes d'IA est poussée dans le monde du travail notamment par l'évolution de la démographie des travailleurs.

Cependant, cette croissance s'érode progressivement à la fin de la décennie :

- Les instabilités économiques et démographiques créent des tensions sur le marché de l'emploi. En recherche permanente de compétitivité, les employeurs favorisent les systèmes d'IA à différents niveaux de la chaîne de création de valeur. Avec le développement du télétravail, les outils d'IA suivent les employés chez eux et deviennent progressivement intrusifs.
- Pour améliorer les couvertures d'assurances des salariés et négocier des primes à la baisse, les données récoltées sont partagées avec les assureurs (notamment la Sécurité sociale) afin de tenir compte de la prévention des risques dans la tarification, voire la personnalisation de certaines garanties.
- En même temps, certains acteurs peu scrupuleux n'hésitent plus à utiliser les données récoltées pour tracer l'activité des employés de manière systématique et sans consentement. Les organisations syndicales et les politiques prennent la menace au sérieux. Les tensions au sein des entreprises et les propositions de lois limitatives se multiplient.

- Les organisations du travail ont des positionnements différents selon les pays et selon les branches d'activités. L'absence de réglementation harmonisée entraîne des inégalités statutaires entre secteurs, et les divergences en matière d'éthique entre les grandes zones économiques conduisent à une augmentation des contentieux en matière du droit des salariés.

Ainsi, la méfiance/défiance grandit au sein des entreprises quant à l'usage de l'IA, et notamment pour la S&ST.

La crise de l'IA dans le monde du travail

Une grande crispation des acteurs privés et publics autour du sujet de l'usage des systèmes d'IA dans la S&ST émerge. La crise de l'usage des systèmes d'IA est déclenchée par une désillusion de plus en plus évidente de la société, le tout corroboré par des incidents malheureux qui se multiplient.

Même si l'argument des employeurs est de faciliter la fluidité et la flexibilité du travail, la perception des salariés est celle d'une « intrusion » et d'un abus de pouvoir. Ces ressentis sont amplifiés par le contexte d'un marché de l'emploi très tendu.

Cybersécurité : diffusion de données biométriques de travailleurs suite à une cyberattaque

« Biometric-Leak : ce jeudi 28 avril 2028, une attaque des serveurs AméliA (site de l'assurance maladie), à l'aide de technologies d'intrusion basées sur l'IA, a permis à des cybercriminels d'accéder à la base nationale des données biométriques des salariés. L'usage de l'IA est de plus en plus récurrent pour venir à bout de systèmes de sécurité biométriques utilisant des réseaux de neurones. Les données piratées, initialement recueillies à des fins de prévention des maladies professionnelles, ont été publiées sur Internet. Elles contiennent l'ensemble des informations monitorées durant l'activité des salariés, ainsi que leurs numéros de Sécurité sociale. Un employeur mal intentionné pourra désormais accéder aux informations médicales de ses salariés ou candidats pour opérer une sélection. Ce que l'on qualifie déjà de choc dans le monde de la prévention des risques professionnels risque de mettre un point d'arrêt au déploiement des applications de *biomonitoring* des activités professionnelles. »

Face aux mouvements de contestation et aux risques de cybersécurité, certains acteurs font le choix d'externaliser les compétences en IA (notamment via des plateformes) pour continuer de bénéficier des avantages concurrentiels qu'elle offre, tout en limitant les risques pour leur image.

Pour jouer l'apaisement et faire face aux imprévus majeurs, de plus en plus d'acteurs se limitent aux usages uniquement sécuritaires.

Un premier pas vers une tentative de régulation européenne est réalisé avec la mise en place d'une obligation de moyens renforcés de l'employeur (celui-ci doit prouver qu'il a pris les mesures nécessaires pour sécuriser les données et éviter tout effet sur la santé), malgré des divergences persistantes entre États.

De plus, les facteurs techniques et surtout environnementaux viennent contraindre les développements et les usages de l'IA. Les progrès attendus ne sont pas arrivés.

L'IA souffre toujours de problèmes structurels et théoriques (explicabilité, sécurité, certification, équité, robustesse, etc.), et elle n'est pas perçue comme digne de confiance par l'opinion publique. Durant cette fin de décennie, les effets du bouleversement climatique sont devenus prégnants dans l'opinion mondiale.

L'acceptabilité sociale, les promesses non tenues de l'IA et les enjeux environnementaux ont eu raison du déploiement des applications de l'IA dans le monde du travail au tournant de cette décennie. Les usages de l'IA au sein des entreprises et chez les travailleurs indépendants vont décroître significativement. L'IA amorce son déclin.

2030-2035 : le déclin de l'IA

La bulle de l'IA des années 20 a explosé !

Dans les ateliers, les tâches automatisables le sont totalement (sans présence humaine) ; les étroites collaborations homme/machines envisagées et testées lors de la décennie ne sont déployées que dans certains cas de figure (usines automobiles) et abandonnées sur les chantiers ou dans le bâtiment, car jugées improductives.

Les grands acteurs privés du marché de l'IA, après une forte concentration dans la décennie 2020, ne revendiquent plus les systèmes d'IA partout et dans tous les secteurs. Les investissements financiers se réorientent vers un numérique frugal, réparti et distribué. Le buzz du Metaverse, où tout le monde se rencontre pour échanger, acheter ou travailler, n'a pas eu lieu et se limite à une population de *gamers*. Les systèmes d'IA sont utilisés en embarqué la plupart du temps, consommant peu ou alors dans quelques domaines niches (militaire, informatique quantique...). Le buzz de l'IA laisse place à des systèmes digitaux qui consomment très peu et fonctionnent en essaim (SWARM, Smart Dust...). La recherche en IA se concentre sur l'IA générale et/ou de système 2 : seule cette IA est perçue comme capable de répondre aux grands enjeux sociétaux, mais le chemin à parcourir reste long.

Les enjeux sociaux et environnementaux sont jugés prioritaires. La crise énergétique, climatique et environnementale flèche les efforts de développement et de recherche vers des enjeux vitaux (réduction des émissions de gaz à effet de serre, captation de CO₂, réduction des pollutions atmosphériques et océaniques).

En France, le mouvement de réorientation de l'économie vers le tourisme et les services à la personne est opérant. Il a pour effet de mobiliser une grande partie de l'emploi vers des métiers non automatisables, non délocalisables.

Cas d'usage

Cas d'usage

La deuxième phase de cet exercice a consisté à sélectionner trois types d'applications possibles de l'IA en matière de S&ST et à les confronter aux futurs possibles décrits dans les scénarios. Ce travail a été conduit sous la forme d'ateliers rassemblant des membres du groupe de travail et des experts en santé et sécurité au travail n'ayant pas participé à la phase initiale.

Les trois domaines d'usages explorés lors de ces ateliers sont :

- L'aide à l'analyse de données et au diagnostic, notamment par des traitements intelligents de grandes quantités de données à des fins d'évaluation des risques, d'épidémiologie ou de suivi médical.
- La sécurisation d'environnements de travail en recourant à des dispositifs en mesure d'analyser en temps réel des informations provenant d'objets connectés (équipements de protection individuelle, capteurs divers, engins...) et de fournir aux travailleurs des informations sur la situation ou de provoquer des alertes, voire d'engager des actions d'urgence de manière autonome.
- La robotique autonome, permettant notamment de déléguer à des machines des opérations complexes intrinsèquement dangereuses ou à mener dans des environnements dangereux (environnements radioactifs, pollués, difficilement accessibles, confinés...).

Les synthèses de ces trois ateliers sont présentées dans cette partie, ainsi qu'un développement consacré spécifiquement à l'intérêt pour la prévention des usages de l'IA en chémo-informatique.

Synthèse des ateliers

Atelier 1 : Usages possibles des outils d'IA en épidémiologie et accidentologie

Thème

De par son potentiel de traitement intelligent de données massives, l'IA ouvre des possibilités prometteuses pour les acteurs en santé et sécurité au travail. On pense notamment ici à certains domaines d'usages :

- **En épidémiologie**, l'IA peut offrir de nouvelles possibilités de traitements sophistiqués et rapides des données collectées pour des populations d'intérêt en termes de vulnérabilité, d'exposition, etc. Mais aussi sans doute une meilleure exploitation croisée de bases de données différentes (indicateurs de santé, trajectoires professionnelles, données de mesures individuelles ou d'ambiance, etc.).
- **En accidentologie**, les systèmes de traitement automatique du langage ouvrent également des possibilités de meilleure exploitation de données mal structurées, de données textuelles (certificats de décès, mais demain pourquoi pas déclarations d'AT, données provenant des services de santé au travail relatives aux aptitudes médicales, bases de données de retour d'expérience type EPICEA, données intra-entreprise, telles qu'un registre de dysfonctionnements de machines, données AT-MP stockées via des logiciels spécifiques, enquêtes QSE, etc.). L'usage de ces outils pour optimiser l'exploitation des données contenues dans la littérature scientifique est également à considérer, puisqu'il dépasse le champ de l'accidentologie.

Exemples de cas d'usage

Pour illustrer ces domaines d'usage, des exemples ont été discutés.

- **PLUS « Processus language upgrade safety »** est une solution de traitement de données textuelles par des techniques de traitement automatique de langage, développée par la société Safety Data¹. L'outil offre d'une part des fonctionnalités classiques de moteur de recherche (opérateurs booléens et autres), et d'autre part des fonctionnalités d'analyse intelligente des similarités textuelles via un système algorithmique. Les similarités textuelles sont présentées sous forme graphique sur une échelle de temps. Une exploitation des récits d'accidents du travail similaires pourrait s'avérer intéressante. Elle pourrait notamment pallier la difficulté à reconstituer un contexte d'accidents du travail (cas des coupures dans le secteur de la restauration, par exemple). On retrouve assez facilement les informations concernant le moment de l'accident (service soir ou matin), mais on manque d'information sur le contexte (comment la coupure s'est produite : couteau, verre, éplucheur). Un outil de ce type pourrait permettre d'avoir une analyse plus fine des récits d'accidents.
- Une étude débutant à l'INRS vise à **évaluer des risques liés aux polyexpositions**. Elle a pour objet le croisement de plusieurs bases nationales afin de mettre en évidence l'ensemble des expositions des travailleurs concernés, les niveaux de celles-ci, et faire le lien avec les pathologies qu'elles génèrent. Les données intégrées couvrent une période rétrospective de 20 ans. La disponibilité des données et l'augmentation des capacités informatiques rendent aujourd'hui ce type d'étude possible. À ce jour, il s'agit de caractériser les liens entre des expositions passées et des pathologies. À terme, l'IA pourrait permettre d'exploiter ces informations pour développer des outils prédictifs utiles en prévention, car ils permettraient d'anticiper de manière précoce d'éventuels effets sur la santé et d'identifier les secteurs, postes, etc. les plus à risques.
- La Cramif (Caisse régionale d'Assurance maladie d'Île-de-France) a développé un **outil d'aide à la codification des déclarations d'accidents du travail (DAT)**. Celui-ci récupère les textes rédigés par les employeurs dans les DAT pour proposer aux agents des caisses chargés de la tarification les codes les plus adaptés. Le développement de l'outil s'est appuyé sur une base d'apprentissage de 300 000 DAT et une base témoin de 65 000 DAT. Cet outil vise à la fois à améliorer la qualité de la codification des accidents du travail, et à aider les personnes chargées de cette tâche en leur présélectionnant les codes les plus pertinents.

Potentiel de développement

Les principaux facteurs susceptibles d'accélérer le développement de ces types d'usages sont :

- La disponibilité de grandes masses de données de qualité
- La volonté des acteurs de partager et d'exploiter leurs données et la mise en place

1. <https://www.safety-data.com/>

d'un cadre permettant de le faire dans de bonnes conditions (protection des données, compatibilité des bases...)

- L'augmentation des capacités de stockage et de traitement de ces données
- L'amélioration des performances des outils d'IA pour atteindre un niveau de fiabilité important

Les bénéfices attendus sont des progrès dans l'évaluation et l'analyse des risques, qui devraient permettre des avancées en prévention.

Atouts et points de vigilance

Le développement de ces outils ouvre des perspectives intéressantes en matière de prévention des risques et de suivi médical des travailleurs. Elles pourraient notamment permettre de détecter de nouveaux facteurs (ou combinaisons de facteurs) de risques, que ce soit pour les accidents ou pour les maladies. Il est également possible d'imaginer des progrès dans le suivi de santé des travailleurs tout au long de leur carrière, et notamment de populations dont le suivi pose des difficultés aux acteurs institutionnels (comme les travailleurs saisonniers ou les intérimaires).

Les principales interrogations portent sur les données nécessaires au fonctionnement de ces systèmes. La protection de ces données est un enjeu important : il s'agit de garantir la sécurisation des données personnelles, notamment des données de santé, d'être transparent sur les objectifs de ces traitements, de prévenir les détournements d'usage. De plus, la qualité des jeux de données, notamment ceux qui vont servir à entraîner les systèmes d'IA, est déterminante. La nature des données d'entraînement devra être représentative des situations traitées, et inversement : les usages devront être limités aux domaines d'applicabilité prévus, correspondant aux données d'entraînement utilisées.

Ces systèmes peuvent à terme faire évoluer les approches de la santé au travail. Ils peuvent notamment inciter à une individualisation accrue du suivi des travailleurs en intégrant des facteurs de risque individuels (génétique, hygiène de vie) aux côtés des facteurs de risques professionnels, et par conséquent conduire à des mesures de prévention de plus en plus individualisées, ce qui pose des questions d'éthique. Il y a donc un enjeu important autour des compétences, des objectifs et du contrôle des acteurs qui développent et utilisent ces outils.

Implications pour les acteurs de la S&ST

Les acteurs de la veille sanitaire et de la prévention sont confrontés à la question de leur montée en compétences sur la compréhension de ces solutions. Il s'agira à la fois de développer leurs compétences internes pour pouvoir conduire des études sur ces systèmes, mais également de faire monter en compétences les acteurs concernés, notamment les administrations, les préventeurs et les partenaires sociaux, et de pouvoir développer une capacité de conseil et d'assistance à leur intention.

Une implication des organismes de la santé et sécurité au travail dans les instances de normalisation et le dialogue avec les régulateurs seront également des enjeux importants pour s'assurer que les solutions développées constituent des avancées en prévention et ne conduisent pas à une dérive de l'approche de la santé au travail.

Il leur faudra aussi arriver à gérer quelques écueils possibles. Il y a d'une part le risque de se focaliser sur les outils eux-mêmes, au détriment d'autres missions et approches.

Ces dispositifs étant dépendants des données disponibles, ils peuvent par conséquent orienter l'attention des acteurs vers des domaines ou des populations pour lesquels ils disposent de données, au détriment d'autres, rendus « invisibles » faute de données exploitables disponibles (cela pourrait être le cas pour des facteurs organisationnels, par exemple). D'autre part, si les développements de ces dispositifs nécessitent des compétences spécifiques et des capacités de stockage, de protection et de traitement de données importantes, ils ne pourront être gérés que par des acteurs disposant de moyens importants (publics ou privés). Cela pourrait contribuer à une centralisation du pilotage de la santé et sécurité au travail, et possiblement au creusement d'un écart entre les informations de pilotage centralisées issues de ces outils et les contextes de travail réels dans les entreprises.

Atelier 2 : Technologies de sécurisation des environnements de travail recourant à l'IA

Thème

L'intelligence artificielle permet d'extraire du sens à partir de volumes considérables de données. C'est le cas notamment pour l'analyse de flux vidéo ou audio en temps réel, de paroles ou encore de données biométriques. Combinés à des dispositifs de captation connectés, les algorithmes d'IA peuvent réaliser une surveillance instantanée ou consolidée des situations observées. Pour la S&ST, de nombreuses initiatives promettent d'alerter les dangers imminents ou de prévenir des troubles chroniques en informant les opérateurs de mauvaises pratiques. Ces applications se basent sur deux typologies : les objets de surveillance de scène (par exemple, caméra intelligente de détection de situation dangereuse dans un entrepôt) et les équipements à porter (« *wearable technology* ») qui mesurent et transmettent des données biométriques du travailleur (comme la mesure de la fréquence cardiaque pour la prévention des maladies cardiovasculaires) ou des positions articulaires (pour la prévention des lombalgies).

Ces solutions visent à sécuriser les environnements de travail en recourant à des dispositifs permettant d'analyser en temps réel des informations provenant d'objets connectés. Ces dispositifs fonctionnent grâce à des capteurs capables de mesurer différents types de valeurs, et à des actionneurs. On distinguera ici deux types d'usage de ces systèmes de surveillance :

- Les systèmes visant à surveiller l'environnement de travail et capables d'actionner une alerte avant la survenue d'un phénomène dangereux : émission d'un produit toxique, proximité d'un équipement en mouvement... Ces systèmes sont définis ici comme des solutions de détection.
- Les systèmes visant à surveiller le travailleur lui-même. Il pourra s'agir d'un EPI connecté effectuant des mesures régulières de données biométriques ou encore d'un équipement de travail muni de capteurs de mesures biométriques.

Exemples de cas d'usage

Solutions de détection de situations dangereuses

L'entreprise Arcure, *spin-off* du CEA, a développé le système de caméras embarquées sur véhicule de chantier Blaxtair², permettant de détecter les piétons grâce à un système d'intelligence artificielle de reconnaissance des personnes. En cas de trajectoire convergente entre l'engin et un piéton, une alarme est activée. Blaxtair fonctionne avec un avertisseur sonore qui s'active uniquement en cas de situation dangereuse.

La technologie permet également d'enregistrer la typologie, la position et l'heure des détections, pour constituer une base de données des presque accidents. Cette information signalée aux responsables prévention (sous réserve d'autorisation de la collecte de données par les salariés) permet de mettre en place des mesures correctives afin d'éviter que ces situations ne se reproduisent. L'outil propose donc deux niveaux de service : une action curative, complétée par une action préventive. Pour être détectés, les salariés n'ont pas d'équipement à porter, ce qui permet d'éviter les risques liés à l'oubli ou à la présence d'une personne non équipée.

Quelques limitations sont à noter : la solution est adaptée aux véhicules qui se déplacent lentement (en situation de manœuvre), car la détection est effective à 7 m, ce qui est trop court si la vitesse est supérieure à 25 km/h. La sauvegarde des données nécessite un réseau 4G, et n'est donc pas possible dans les souterrains, par exemple. En revanche, la détection de piétons et l'alerte restent opérationnelles sans réseau. Un système d'autodiagnostic existe pour alerter le conducteur en cas de dysfonctionnement du dispositif, mais certaines perturbations ne sont pas diagnostiquées, et l'outil ne peut pas se substituer à la vigilance du conducteur. En l'état, le système ne peut donc pas être considéré comme un système de sécurité d'un point de vue réglementaire, selon les normes en vigueur. Cette exigence réglementaire de fiabilité pose la question de la viabilité économique d'une solution de sécurité qui serait trop chère pour disposer d'un marché.

Systèmes visant à surveiller le travailleur lui-même

La société Eyes'R³ propose une solution d'assistance numérique de prévention et d'anticipation des risques basée sur un système d'IA de reconnaissance visuelle relié aux caméras de vidéosurveillance de l'entreprise, qui permet de détecter le port des EPI ou la présence d'un opérateur dans un espace donné à un moment précis. La solution propose trois modules :

- Un module de détection de situations à risque : les algorithmes d'IA sont en mesure de reconnaître différents types de situations dangereuses (défaut de port d'EPI, matériels mal rangés, flaques d'eau ou d'huile, présence d'individus dans des lieux ou lors de phases d'activité dangereuses...). La technologie peut être reliée à des systèmes d'alarme (sonore ou visuelle), à un système d'ouverture de portes, et peut générer des alertes sur smartphones ou écrans destinées aux salariés ou à leurs responsables.
- Un module de pilotage des indicateurs, visant à exploiter les données collectées pour générer des rapports d'analyse des risques et aider à la définition de plans d'actions.
- Un module « d'intelligence collective » ayant pour objectif d'impliquer les collaborateurs dans l'identification de situations à risques. Via une application mobile, les salariés ont ainsi la possibilité de photographier des situations dangereuses constatées dans l'entreprise pour les signaler.

2. <https://blaxtair.com/>

3. <https://eyesr.fr/>

Potentiel de développement

Actuellement, les avancées sont liées à l'évolution de puissance de calcul, qui permet une augmentation de la performance et de la précision. La consommation d'énergie des puces embarquées est de plus en plus faible. Les développements dans le domaine quantique pourraient encore augmenter cette capacité de calcul et ainsi augmenter la pertinence, réduire la compacité, de même que la consommation d'énergie (sur le véhicule). Tout cela pourra en définitive permettre de réduire les prix de ces solutions, qui limitent actuellement fortement leur déploiement.

La 5G permettra aussi d'assurer un niveau de sécurité élevé en réduisant les temps de latence.

Atouts et points de vigilance

La facilité d'usage de ces technologies permet le développement de nouveaux outils performants dans le domaine de la prévention des risques professionnels, qui peuvent aider à réduire de façon significative la sinistralité. Elles apportent la possibilité de sécuriser les environnements, y compris pour des salariés peu formés ou sensibilisés aux règles de sécurité du site (comme les intérimaires, les nouveaux embauchés ou les intervenants extérieurs). La disponibilité de ces outils constituant une promesse d'amélioration de la sécurité dans les environnements de travail, elle pourrait conduire le législateur et le juge à renforcer leurs exigences en matière d'obligation de moyens et de résultats de l'employeur. Enfin, du fait des informations qu'ils peuvent apporter sur les situations et comportements à risques, ils peuvent aider au ciblage des actions de prévention en fonction de l'occurrence des presque accidents, et à prioriser les efforts de formation sur certains publics : intérimaires, nouveaux embauchés, chargés de la maintenance (cf. scénario 3).

Cependant, un usage inadapté, détourné, ou l'absence d'une réflexion préalable sur l'organisation de l'intégration de ces nouvelles technologies pourrait conduire à des effets délétères en matière de santé et sécurité au travail. La facilité d'usage et d'implémentation de ces solutions risque d'encourager les acteurs à ne prendre en considération que les risques identifiés par la machine, sans évaluer régulièrement les risques plus organisationnels. La méconnaissance des possibles défaillances de ces outils pourrait entraîner une baisse de la vigilance des salariés, comme des employeurs. Cette déresponsabilisation à la fois collective et individuelle en matière de prévention pourrait être accompagnée de l'apparition de comportements déviants (comme le fait de chercher à éviter la surveillance opérée par ces technologies), possiblement générateurs d'accidents. Ces outils pourraient aussi favoriser le développement d'approches plus coercitives que préventives, avec pour conséquence de reporter la responsabilité sur le salarié en cas d'accident.

Détournées de leur finalité première de prévention, ces technologies sont aussi susceptibles de faciliter une surveillance intrusive des salariés, à la fois sur leur état de santé par la collecte de données (rythme cardiaque, suivi médical sans accord formalisé), mais également sur leurs comportements et rythmes de travail. Une telle surveillance peut induire une intensification du travail chez le salarié qui se sait surveillé, et conduire à des risques psychosociaux ou d'accidents accrus (cf. scénario 1).

Implications pour les acteurs de la S&ST

La collecte et l'exploitation de masses importantes de données sur les lieux de travail offrent une opportunité d'analyse inédite pour objectiver les expositions et révéler certains facteurs de risques.

Le regroupement et la mutualisation de données de différentes entreprises d'un même secteur pour constituer une base sous l'égide d'organismes de recherche pourraient permettre de faire avancer à grands pas la recherche dans le domaine de la prévention en santé et sécurité au travail.

Parallèlement, face aux conséquences d'usages potentiellement délétères de ces technologies émergentes, un effort d'information et de formation en direction des préventeurs et des entreprises est souhaitable pour faire connaître les points de vigilance à garder à l'esprit et les bonnes pratiques à mettre en œuvre lors de leur implémentation.

Enfin, les préventeurs auront pour responsabilité d'alerter et de rappeler aux entreprises que ces outils ne s'apparentent pas à des systèmes de sécurité et qu'ils ne les exonèrent pas de la démarche d'évaluation des risques et de prévention à mettre en œuvre.

Atelier 3 : Robotique avancée recourant à l'IA

Thème

Certaines technologies de robotique avancée (embarquant de l'IA) offrent des solutions potentiellement bénéfiques à la S&ST. Pour cet atelier, deux types de cas d'usage ont été identifiés : la téléopération, qui permet d'éloigner l'opérateur de situations délétères ou dangereuses ; et la collaboration homme/robot, qui permet de faire exécuter les tâches physiquement pénibles de l'opérateur par un robot apte à la collaboration.

- **La téléopération** : les compacteurs ou les brise-roches téléguidés du BTP permettent d'éloigner l'opérateur de ces machines dangereuses. Mais pour les tâches plus complexes ou à réaliser dans un espace dangereux, le téléguidage n'est pas toujours possible. Dans ces situations, la machine doit embarquer une partie de l'expertise et des capacités de l'opérateur humain (voire des capacités supérieures à celles de ce dernier). Il s'agit notamment de la vue stéréoscopique, du déplacement dans un environnement fait pour des humains ou encore de la préhension d'objets... Dans ces cas, l'opérateur distant ne pilotera que des « ordres macro » (de type « Rends-toi à tel emplacement », « Ouvre cette porte », « Effectue la mission XDK74 », etc.). On parle alors de téléopération, car cela nécessite des fonctions de robotique avancée faisant appel à l'IA.
- **La collaboration** : la robotique collaborative définit les conditions grâce auxquelles on peut faire travailler sans risque corporel un opérateur à proximité immédiate d'un robot. On distingue 3 types de collaboration : la collaboration directe où l'opérateur et le robot travaillent simultanément sur une même pièce, la collaboration indirecte où l'opérateur et le robot travaillent alternativement sur la même pièce, et le partage d'espace de travail où opérateur et robot travaillent indépendamment dans un espace commun. Pour être sûres, ces collaborations imposent que le robot soit doté de facultés très pointues de perception des actions de son collaborateur humain. Ces facultés peuvent faire appel à de l'IA.

Exemples de cas d'usage

Pour illustrer ces domaines d'usage, deux exemples ont été présentés.

Usage d'un robot Spot (robot autonome ou téléopérable de la société Boston Dynamics⁴), permettant de remplacer ou d'éloigner un opérateur humain dans des situations de travail dangereuses.

Le projet est conduit par l'école des Mines de Nancy en collaboration avec l'Andra (Agence nationale pour la gestion des déchets radioactifs) et vise à remplacer les opérateurs dans des environnements dangereux ou difficilement accessibles.

Actuellement, le robot est testé sur le site de Bure. Spot a pour mission de vérifier à fréquence régulière la déformation de tunnels situés à 400 m sous terre destinés à accueillir, à terme, des déchets nucléaires. L'objectif recherché est de dispenser les opérateurs humains de ces missions en milieux confinés, difficilement accessibles et potentiellement pollués. L'IA est exploitée ici pour le niveau d'autonomie qu'elle apporte au robot, qui dispose ainsi d'une capacité d'adaptation à son environnement. Il existe deux modes de fonctionnement du robot : un mode autonome pour patrouiller dans un environnement, et un mode de pilotage asservi. Spot dispose de la même faculté de déplacement qu'un humain (prend l'ascenseur, ouvre les portes ou manipule des vannes). Il nécessite cependant d'avoir accès à un réseau WiFi, mais ne nécessite pas de supervision en continu.

Cog.Engines (solution visant à doter des robots collaboratifs de capacités de perception de leur environnement⁵), application permettant de seconder ou d'aider un opérateur humain à un poste de travail.

L'exemple présente un bras robotique collaboratif programmé et piloté grâce à une IA.

Ici, l'IA est utilisée pour permettre l'adaptation à une activité flexible, voire sur-mesure, qui va changer très régulièrement. Un boîtier de vision permet au robot de voir en 3 dimensions, et permet aux opérateurs de le piloter par gestes. Il a également la faculté de détecter les risques de collisions.

La gestion du danger provenant des collisions intervient à trois niveaux et est basée sur des algorithmes classiques, qui ne font pas appel à de l'IA : le cobot limite sa puissance et ses efforts en cas de collision ; le cobot ralentit lorsqu'un opérateur est détecté dans un périmètre défini ; le cobot s'arrête grâce à la caméra qui scrute la présence d'un opérateur.

Ces différentes caractéristiques lui permettent de travailler en collaboration avec les opérateurs, de reconnaître des produits et d'accomplir ses tâches. Le geste répétitif (dans le cas étudié, il s'agit de remplir des cartons avec des flacons) est laissé au robot.

L'opérateur peut alors prendre en charge de nouvelles tâches moins sollicitantes et plus qualifiées de supervision du robot. Un superviseur est en mesure de surveiller plusieurs robots sur différentes lignes.

4. <https://www.bostondynamics.com/products/spot>

5. <https://cogengines.com/>

Potentiel de développement

Ces deux types d'usages sont destinés à se déployer si les progrès suivants sont réalisés :

- Démocratisation de la technologie (coûts d'achat, d'intégration, d'exploitation et de maintenance)
- Extension des domaines de fonctionnement opérationnel
- Progression de l'acceptabilité et de la confiance par les collaborateurs humains
- Progression de la qualité de l'intégration de ces technologies dans les organisations de travail (transparence de l'objectif visé par cette intégration [augmentation de la productivité ou réduction de la pénibilité], préparation et concertation des équipes, cadence au poste de travail, accompagnement du changement, formation)

Pour ces deux domaines d'application, et à moyen terme, les potentiels de développement se concentrent autour de deux enjeux :

- L'augmentation de la flexibilité de ces équipements (variabilité des produits et des tâches)
- L'augmentation de la capacité de ces équipements à réaliser des tâches plus complexes

Atouts et points de vigilance

Le principal atout de ces technologies réside dans la réduction, voire la suppression, des risques auxquels les opérateurs sont soumis grâce au remplacement du salarié par le robot sur les sites confinés ou pollués, pour la réalisation de tâches répétitives, ou encore le port de charges lourdes.

Elles pourraient ainsi favoriser le maintien ou le retour à l'emploi d'opérateurs déchargés de la partie physique du poste. Par ailleurs, elles peuvent constituer une opportunité de requalification des opérateurs qui, une fois déchargés des tâches répétitives désormais réalisées par le robot, peuvent monter en compétences, notamment dans les domaines liés à la maintenance et la supervision de ces nouveaux outils (*cf. scénarios 2 et 3*).

Cependant, ces technologies peuvent également faire émerger certains risques. Une organisation du travail qui n'aurait pas impliqué une expérimentation en amont ni une réflexion collective sur l'intégration de ces nouveaux outils pourrait conduire à une augmentation délétère des cadences, du fait du maintien des objectifs de production auxquels s'ajouteraient les activités de maintenance et de supervision. Le remplacement du salarié par le robot peut aussi entraîner une perte du geste professionnel, conduisant à une progressive déqualification de la main-d'œuvre, une dévalorisation des opérateurs, une réduction de leurs marges de manœuvre, et une potentielle perte de sens au travail. Les risques psychosociaux pouvant en découler sont donc nombreux (*cf. scénario 1*).

Par ailleurs, des comportements inattendus de la machine, provoqués par une situation imprévue ou une cyberattaque, pourraient conduire à des réflexes de rattrapage ou à des interventions en situations dégradées et être la cause d'accidents du travail. La perte de confiance qui pourrait en résulter pourrait mener à un abandon de ces technologies, en dépit des atouts qu'elles représentent (*cf. scénario 4*).

Implications pour les acteurs de la S&ST

Les acteurs de la prévention, de par leur pratique pluridisciplinaire, leur expertise des risques mécaniques et organisationnels et des dispositifs juridiques, sont à même

d'accompagner l'émergence de ces technologies et d'intervenir en amont, lors des phases de développement. Cependant, ces acteurs pâtissent pour le moment d'un manque de compétences en matière d'intelligence artificielle et de robotique intelligente et de terrains d'expérimentation.

À l'avenir, une évaluation de la pertinence du recours à ces outils –réalisée autour d'un dialogue social au sein des entreprises – pourrait être favorisée par des méthodes adaptées. Celles-ci pourraient émerger de travaux de veille et de recherche concernant les risques spécifiques liés à l'IA, et de collaborations entre acteurs. Ces recherches pourraient également contribuer à la validation de la robustesse des systèmes, ainsi qu'à faire évoluer la réglementation. Des produits d'information et de sensibilisation pourraient également en découler.

Des efforts de formation – des préventeurs, des concepteurs, mais également des salariés – sont nécessaires pour permettre aux acteurs concernés de pouvoir débattre de ces sujets et instaurer confiance et acceptabilité.

Apport de la chémo-informatique à la prévention des risques associés aux usages de substances chimiques

L. Chedik

Contexte

L'évaluation de la toxicité des produits chimiques utilisés en milieu professionnel est essentielle pour la santé et la sécurité des travailleurs. Les agences sanitaires, les préventeurs et les chercheurs s'efforcent d'évaluer les risques chimiques des substances auxquelles sont exposés les travailleurs. En outre, les entreprises qui développent ces composés chimiques se doivent, depuis la réglementation REACH de 2006, de justifier de l'absence d'effets nocifs pour la santé humaine et l'environnement.

Pour obtenir un profil toxicologique d'un seul produit chimique, il est nécessaire d'effectuer de très nombreuses expérimentations *in vitro* et *in vivo* (expérimentation animale). Le coût et le temps considérables nécessaires constituent des obstacles majeurs à l'évaluation des risques des substances chimiques disponibles sur le marché. De fait, il existe des lacunes importantes dans la littérature concernant la toxicité de la grande majorité des produits chimiques utilisés. De plus, le contexte socio-politique pousse à la fois à une meilleure évaluation des risques de ces substances et à une réduction du recours à l'expérimentation animale.

Dans ce cadre, les approches *in silico* chémo/bio-informatique sont complémentaires aux expérimentations *in vitro* / *in vivo* dans la mesure où elles permettent de réduire les coûts et les délais de recherche, et de gagner en réactivité pour évaluer *a priori* les risques toxicologiques des substances chimiques. Elles ont également un intérêt pour la priorisation des travaux de recherche (criblage et sélection de molécules préoccupantes à tester expérimentalement). En cas d'interdiction ou de restriction d'utilisation de certains produits, elles permettent de rechercher des substituts potentiels sur la base de leur innocuité pour la santé humaine et ce, avant même leur synthèse. Ces nouvelles méthodes ont également un intérêt pour l'évaluation des effets de mélanges de substances chimiques, étant donné qu'il est impossible de tester toutes les combinaisons possibles.

Définition

La chémo-informatique regroupe un ensemble de domaines de recherche qui consistent en l'application d'outils et de méthodes informatiques (algorithmes, Big Data, apprentissage profond...) pour la résolution de problèmes relatifs à la chimie, à l'instar de la bio-informatique qui s'intéresse aux problèmes relatifs à la biologie.

Les champs de recherche en chémo/bio-informatique se sont développés avec les progrès opérés en sciences des données et en intelligence artificielle. Ce document sera axé essentiellement sur la modélisation QSAR et la modélisation moléculaire, qui ont montré leurs intérêts pour l'évaluation du risque chimique.

Modélisation QSAR

La modélisation QSAR (pour « Relation Structure Activité Quantitative ») consiste à corrélérer l'activité biologique d'un composé chimique à sa structure et ses propriétés physicochimiques. Cette méthode permet de créer un modèle numérique capable de prédire l'activité biologique de composés inconnus à partir de leurs structures chimiques.

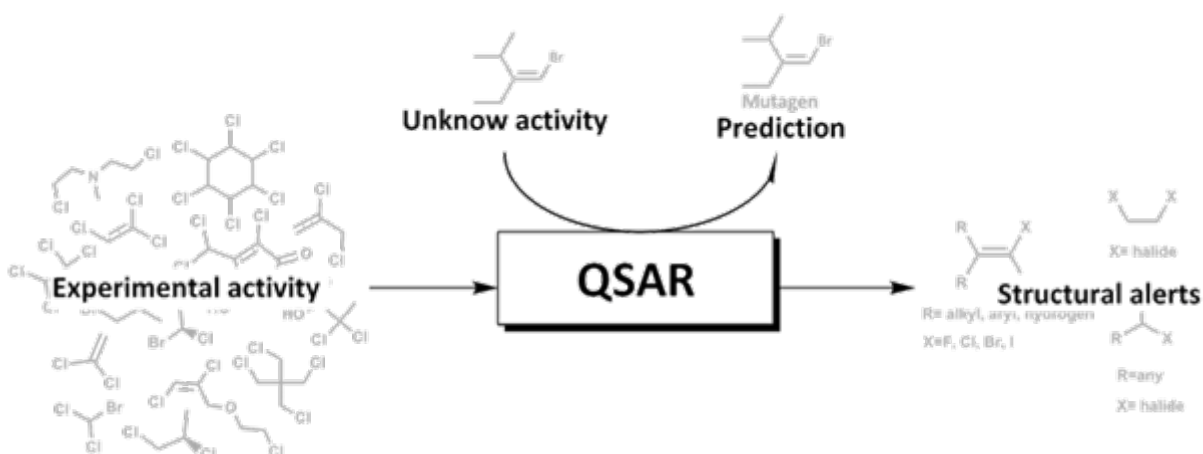


Figure 1. Principe de la modélisation QSAR

De nombreux modèles QSAR ont été développés depuis une trentaine d'années pour prédire des effets toxicologiques, tels que la cytotoxicité¹, ou la cancérogénicité². Les logiciels DEREK Nexus et ToxTree utilisent par exemple des alertes structurales (toxicophores) pour prédire le potentiel mutagène d'une substance. Citons également le module QSARToolBox de l'Agence européenne des produits chimiques (ECHA), qui permet par exemple la prédiction de la toxicité pour le développement embryonnaire ou encore de l'irritation oculaire³.

Certains modèles QSAR prédicteurs d'effets toxiques ont été implémentés dans des interfaces web afin de permettre à des personnes ne disposant pas de compétences particulières en programmation d'évaluer *a priori* la toxicité de composés pour lesquels on ne dispose pas de données expérimentales.

1. MouseTox, Varsou et al., 2017, <https://doi.org/10.1016/j.fct.2017.09.058>

2. CarcinoPred-EL, Zhang et al., 2017, <https://doi.org/10.1038/s41598-017-02365-0>

3. <https://echa.europa.eu/fr/support/oecd-qsar-toolbox>

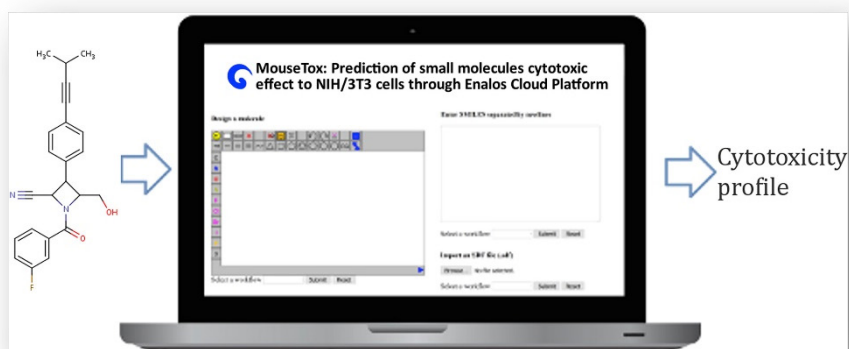


Figure 2. Principe du service web MouseTox permettant la prédiction des effets cytotoxiques des composés chimiques sur les cellules NIH/3T3 fibroblastes embryonnaires de souris.

Source : <http://www.enaloscloud.novamechanics.com/EnalosWebApps/MouseTox/>, Varsou et al. 2017.

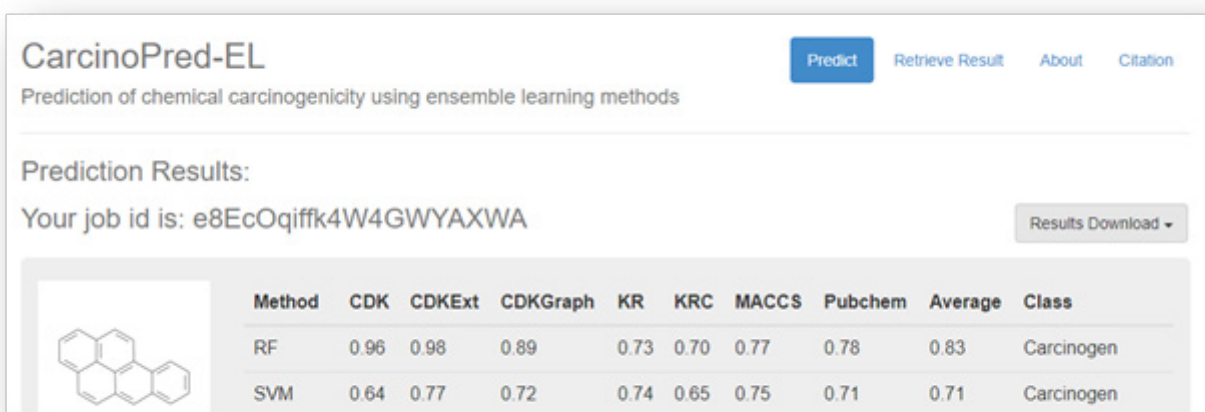


Figure 3. Exemple de prédiction du potentiel cancérigène du Benzo[a]pyrène via le service web CarcinoPred-EL.

Source : <http://112.126.70.33/toxicity/CarcinoPred-EL/index.html>

L'INRS développe actuellement des modèles QSAR pour prédire le potentiel sensibilisant des substances chimiques sur la base de leur structure chimique et des données expérimentales déjà générées au sein de l'Institut⁴. Ces modèles devraient permettre d'alerter sur les risques allergisants des substances chimiques manipulées par les professionnels.

En plus de ces effets toxicologiques, des paramètres toxicocinétiques renseignant notamment sur le devenir des substances chimiques dans l'organisme (potentialité d'être absorbées, distribuées, bio-transformées, excrétées) peuvent être prédits par ces mêmes approches QSAR. C'est l'objectif d'un second QSAR élaboré au sein de l'INRS, qui vise à prédire l'ampleur de l'absorption par voie cutanée des toxiques industriels.

4. Ces travaux sont menés par Lisa Chedik, du laboratoire Toxicocinétique, Inhalation et Passage Cutané (TIPC), en collaboration avec le laboratoire Immunologie, Sensibilisation et Allergie.

Il est prévisible, qu'à l'avenir, les agences sanitaires demanderont de plus en plus souvent des données issues de modèles QSAR dans les dossiers d'enregistrement réglementaire des composés chimiques. C'est déjà le cas pour l'évaluation du caractère mutagène des impuretés présentes dans les médicaments⁵.

Modélisation moléculaire

La modélisation moléculaire consiste à construire des modèles informatiques de molécules ou d'ensemble de molécules pour prédire leurs comportements ou leurs interactions. Ce domaine de recherche se base essentiellement sur des données de biologie structurale pour reconstruire la structure tridimensionnelle de molécules et étudier leurs fonctionnalités. Ces approches moins accessibles aux non-initiés (pas d'interface web) permettent de cribler des bibliothèques de molécules et de déterminer leur capacité à interagir avec des cibles biologiques. De nombreux modèles ont permis la prédiction du potentiel perturbateur endocrinien de composés chimiques en fonction de leur capacité à se lier à différents récepteurs d'hormones. De la même manière que pour la modélisation QSAR, ces approches permettent un gain de temps et d'argent, en priorisant les études et recherches *in vitro* et *in vivo* ciblant des molécules susceptibles de perturber l'homéostasie de l'organisme et de provoquer des pathologies chez le travailleur.

L'amarrage moléculaire (*docking* moléculaire) a par exemple montré son intérêt pour l'évaluation du potentiel sensibilisant cutané des substances chimiques⁶. Des produits chimiques sensibilisants pour la peau, sélectionnés par cette méthode, ont pu par la suite être testés sur des modèles *in vitro* pour valider leur capacité à stimuler la prolifération des lymphocytes T (une des caractéristiques des sensibilisants cutanés).

D'autres approches chémo/bio-informatiques sont intéressantes pour l'évaluation du risque chimique. Citons par exemple l'analyse de la chimie des réactions électrophiles pour prédire le potentiel sensibilisant respiratoire⁷ ou encore l'utilisation du QSPR (Quantitative Structure Property Relationship) – qui est le pendant du QSAR pour la prédiction des propriétés physico-chimiques (inflammabilité, explosivité...)⁸. Toutes ces approches devraient se développer à l'avenir, et apporter de nouveaux éléments utiles à l'évaluation des risques inhérents aux substances chimiques utilisées en milieu professionnel.

5. En l'absence de données expérimentales adéquates, la ligne directrice ICH M7 recommande l'utilisation de deux modèles QSAR complémentaires pour estimer une classification initiale du danger.

6. Schutte, et al. 2019, <https://doi.org/10.1111/cod.13283>

7. Steven J. Enoch, David W. Roberts, et Mark T. D. Cronin, 2009, <https://doi.org/10.1021/tx9001463>

8. G. Fayet et P. Rotureau, 2022, <https://doi.org/10.1002/minf.202000190>

Focus thématiques

Ces focus thématiques sont proposés en complément pour fournir au lecteur des synthèses sur certains domaines intéressant particulièrement les acteurs de la S&ST : aspects juridiques, éthiques et questions de normalisation. Ils reprennent pour partie des éléments abordés dans les fiches variables.

Encadrement juridique des usages des technologies de l'information et de la communication en milieu professionnel

L'introduction des technologies de l'information et de la communication (TIC) en milieu professionnel a notamment fait évoluer l'organisation du travail. Ces évolutions peuvent être à l'origine de la dégradation des conditions de travail et générer des risques professionnels. Ainsi, si l'employeur est tenu de respecter des réglementations spécifiques – qui encadrent l'usage des algorithmes au travail, la protection des données personnelles des salariés et la surveillance numérique de ces derniers –, il doit également assurer la prévention des éventuels risques associés au déploiement de ces technologies en vertu de son obligation générale de sécurité prévue par le Code du travail.

En effet, l'employeur doit prendre les mesures nécessaires pour assurer la sécurité et protéger la santé physique et mentale des travailleurs. Ces mesures comprennent notamment des actions de prévention des risques professionnels, ainsi que la mise en place d'une organisation et de moyens adaptés¹. L'employeur doit mettre en œuvre ces mesures sur le fondement des 9 principes généraux de prévention, notamment :

- L'évaluation des risques², l'employeur doit identifier les risques (par exemple, les risques psychosociaux, les risques liés au travail isolé), définir les actions de prévention les plus appropriées, et formaliser les résultats dans le document unique d'évaluation des risques professionnels (DUERP) ;
- L'adaptation du travail à l'homme, en vue de limiter le travail monotone et cadencé potentiellement favorisé par l'automatisation ;
- La prise en compte de l'état d'évolution de la technique, en adaptant la prévention aux innovations disponibles (systèmes de détection et d'alarme, équipements de protection connectés, systèmes téléopérés, automatisation de certaines tâches, etc.) ;
- La planification de la prévention en intégrant technique, organisation et conditions de travail, relations sociales et environnement ;
- La formation et l'information des salariés afin qu'ils connaissent les risques et les mesures de prévention (formation aux outils numériques, au télétravail, etc.)³.

1. Art. L. 4121-1 du Code du travail.

2. Art. L. 4121-3 et L. 4121-3-1 du Code du travail.

3. Art. L. 4121-2 du Code du travail.

Le rôle du comité social et économique (CSE), dans le cadre du développement des TIC en milieu professionnel, est également important. En effet, l'introduction de toute nouvelle technologie doit être soumise à la consultation préalable du CSE dans les entreprises d'au moins 50 salariés⁴. Par exemple, un projet d'introduction d'un système d'intelligence artificielle doit être soumis à l'avis préalable du CSE, même s'il n'a que des conséquences mineures sur les conditions de travail⁵.

Le CSE est également informé sur les méthodes d'aide au recrutement des candidats à un emploi. Il est informé sur les traitements automatisés de gestion du personnel, ainsi que sur les moyens ou les techniques permettant un contrôle de l'activité des salariés⁶.

L'usage des algorithmes au travail

Dans une proposition de directive du 19 décembre 2021 du Parlement européen et du Conseil relative à l'amélioration des conditions de travail dans le cadre du travail via une plateforme, la Commission européenne a souhaité renforcer la transparence des entreprises dans l'utilisation des algorithmes en :

- Imposant un suivi humain des conditions de travail
- Informant les travailleurs sur la manière dont les algorithmes sont utilisés
- Évaluant l'exposition des travailleurs aux RPS causés par des décisions algorithmiques
- Donnant la possibilité aux travailleurs de contester les décisions automatisées
- Obligeant les plateformes à consulter les représentants du personnel sur les changements apportés par les décisions algorithmiques

L'AI Act : une proposition de régulation européenne⁷

Dans une proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle du 21 avril 2021, la Commission européenne a également souhaité veiller à ce que les systèmes d'intelligence artificielle – dont les algorithmes – respectent les droits fondamentaux des citoyens de l'Union européenne⁸. L'AI Act est une première mondiale, en ce qu'aucune autre institution politique n'a encore tenté d'encadrer les usages de l'intelligence artificielle. Il se fixe pour objectif de fournir un cadre juridique dans le but de favoriser le développement d'une IA de confiance.

Pour ce faire, la Commission européenne a choisi une approche par risque plutôt qu'une approche sectorielle ou juridique. Quatre niveaux de risque sont ainsi distingués : les dispositifs d'intelligence artificielle présentant un risque inacceptable sont interdits⁹, ceux présentant un haut niveau de risque sont soumis à une évaluation de conformité détaillée¹⁰, ceux présentant un risque faible font l'objet d'obligations de transparence¹¹ et ceux présentant un risque minimal ne font l'objet d'aucune obligation particulière.

4. Art. L. 2312-8 du Code du travail.

5. [Cass. soc., 12 avr. 2018, n° 16-27.866.](#)

6. Art. L. 2312-38 du Code du travail.

7. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0020.02/DOC_1&format=PDF

8. https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_fr

9. Art. 5 de l'AI Act.

10. Art. 6 à 51 de l'AI Act.

11. Art 52 de l'AI Act.

Sont considérées comme inacceptables les technologies d'IA utilisées à des fins de surveillance de masse, celles qui visent à manipuler les comportements, les opinions ou les décisions, ou encore celles qui permettent la notation sociale par les États (la sécurité publique et la menace terroriste constituent cependant une exception).

Les IA à haut risque (énumérées dans l'annexe II du projet de règlement), doivent répondre à un certain nombre d'exigences de conformité (spécifications des données, documentation, traçabilité, transparence, surveillance humaine, exactitude et robustesse), qui peuvent faire l'objet d'une certification par l'entreprise elle-même (comme c'est le cas pour les machines dangereuses). Le non-respect de ces exigences, de même que l'infraction aux pratiques interdites, pourra donner lieu à des amendes pouvant aller jusqu'à 30 millions d'euros ou 6 % du chiffre d'affaires.

Les usages à faible risque font pour leur part l'objet de simples obligations de transparence.

Le projet prévoit parallèlement la création d'un Conseil européen de l'IA, amené à émettre des recommandations concernant les IA à interdire et celles à considérer comme à haut risque.

La protection des données personnelles des salariés

Le Règlement général sur la protection des données (RGPD) s'applique à tout organisme au sein de l'Union européenne traitant des données personnelles¹². Il s'applique donc aux employeurs, mais également aux CSE lorsqu'ils gèrent les activités sociales et culturelles, ainsi qu'aux syndicats.

Le traitement des données personnelles du salarié doit répondre à un objectif précis et être justifié au regard des activités de l'entreprise. L'employeur doit veiller à ne collecter que les données pertinentes et strictement nécessaires. Certaines catégories de données appellent une vigilance renforcée de par leur caractère particulièrement sensible (numéro de sécurité sociale, casier judiciaire, etc.).

Un traitement de gestion du personnel peut ainsi être mis en œuvre pour les finalités suivantes : le recrutement, la gestion administrative des salariés, les rémunérations, la mise à disposition d'outils professionnels, l'organisation du travail, le suivi des carrières et de la mobilité, la formation, la tenue des registres obligatoires, la communication interne, la gestion des aides sociales et la gestion de contentieux. Les informations recueillies pour l'une de ces finalités ne peuvent être réutilisées pour poursuivre un autre objectif qui serait incompatible avec la finalité initiale.

Lorsqu'un traitement poursuit plusieurs finalités, le responsable du traitement doit déterminer la base légale la plus appropriée pour chacune d'elles¹³. Dans le cadre de la gestion des ressources humaines, les bases légales les plus fréquemment mobilisables sont :

- Le respect d'une obligation légale incombant à l'employeur imposant la mise en œuvre d'un traitement (obligations liées à la déclaration sociale nominative, à la tenue d'un registre unique du personnel, à la gestion des élections professionnelles, etc.) ;
- L'exécution du contrat de travail (bulletins de salaire, gestion des demandes de formation, etc.) ou de mesures précontractuelles (traitement des candidatures) ;

12. Art. 2 et 3 du RGPD.

13. Art. 6.1 du RGPD.

- La réalisation de l'intérêt légitime poursuivi par l'employeur (constitution d'une CV-thèque, réalisation de statistiques à des fins de gestion administrative, gestion des dotations individuelles en fournitures, etc.), sous réserve de ne pas méconnaître les droits fondamentaux du salarié ;
- L'exécution d'une mission d'intérêt public dont est investi l'employeur.

Dans la majorité des cas, le consentement du salarié ne doit pas être requis pour le traitement courant de ses données personnelles. En effet, les salariés sont très rarement en mesure de donner, refuser ou révoquer librement leur consentement, de par leur lien de subordination juridique avec l'employeur.

Ils ne peuvent donner leur libre consentement que si l'acceptation ou le rejet d'une proposition n'entraîne aucune conséquence sur leur situation¹⁴.

L'employeur a différentes obligations, notamment informer individuellement les salariés de l'existence d'un traitement de leurs données personnelles, de ses caractéristiques et des droits dont ils disposent¹⁵, ainsi que veiller à la sécurité des systèmes d'information¹⁶.

Les salariés, quant à eux, disposent notamment des **droits** suivants :

- Le droit d'accès à leurs données personnelles¹⁷ ;
- Le droit de rectification des données inexactes¹⁸ ;
- Le droit à l'effacement des données¹⁹, dans certaines circonstances.

En cas de **violation des données personnelles**, l'employeur doit notifier cette violation :

- À la CNIL, dans les 72 h, si la violation est susceptible d'engendrer un risque pour les droits et libertés du salarié²⁰ ;
- Au salarié concerné dans les meilleurs délais, si la violation est susceptible d'engendrer un risque élevé pour ses droits et libertés²¹.

En cas de violation de leurs données personnelles, les salariés ont également divers recours : ils peuvent porter des réclamations à la CNIL²², exercer un recours juridictionnel contre une décision prise par la CNIL²³ ou contre une décision prise par l'employeur²⁴.

En cas de manquement à ses obligations, l'employeur peut être sanctionné²⁵ par la CNIL à travers des avertissements (rappel des obligations), des mises en demeure (qui peuvent comprendre une limitation temporaire ou définitive d'un traitement ou la satisfaction des demandes du salarié [rectification, effacement des données, etc.]) ou par des amendes administratives²⁶. Les décisions de la CNIL peuvent être publiques ou non.

14. Référentiel relatif aux traitements de données à caractère personnel mis en œuvre aux fins de gestion du personnel, adopté le 21 nov. 2019 par la CNIL.
15. Art. 12, 1. du RGPD.
16. Art. 32, 1. du RGPD.
17. Art. 15, 1. du RGPD.
18. Art. 16 du RGPD.

19. Art. 17, 1. du RGPD.
20. Art. 33, 1. du RGPD.
21. Art. 34, 1. du RGPD.
22. Art. 77, 1. du RGPD.
23. Art. 78, 1. du RGPD.
24. Art. 79, 1. du RGPD.
25. Art. 58 du RGPD.
26. Art. 83 du RGPD.

Enfin, l'employeur peut avoir à réaliser une analyse d'impact dès lors que le traitement qu'il met en œuvre est susceptible de présenter un risque élevé pour les droits des salariés²⁷. Dans une délibération du 11 octobre 2018, la CNIL liste les opérations de traitement pour lesquelles une analyse d'impact est exigée. En matière de ressources humaines, les traitements établissant des profils de personnes physiques à des fins de gestion des ressources humaines et les traitements ayant pour finalité de surveiller de manière constante l'activité des employés concernés sont considérés comme des traitements susceptibles de présenter un risque.

La surveillance numérique des salariés

En principe, l'employeur a le droit de surveiller ses salariés pendant le temps de travail, de manière temporaire. L'emploi de procédés clandestins de surveillance est illicite²⁸. Tout système de surveillance doit répondre à deux conditions :

- Il est proportionné au but recherché par l'employeur ;
- Il est justifié par la nature de la tâche à accomplir²⁹.

De plus, l'employeur est tenu de respecter des obligations d'information :

- Porter préalablement à la connaissance du salarié tout dispositif de collecte de ses informations personnelles³⁰ (par exemple, pour la géolocalisation, chaque salarié doit être informé de l'identité du responsable de traitement, des finalités poursuivies, de la base légale du dispositif, des destinataires des données, de son droit d'opposition pour motif légitime, de la durée de conservation des données, de ses droits d'accès et de rectification, de sa possibilité d'introduire une réclamation auprès de la CNIL).
- Informer et consulter le CSE sur la mise en œuvre des moyens de contrôle de l'activité des salariés³¹.

Surveillance informatique

- E-mails, messagerie, sms : en principe, le salarié a droit au respect de sa vie privée³², même quand le salarié est sous la subordination de l'employeur³³ :
 - Messagerie professionnelle : l'employeur a le droit de consulter les mails³⁴ et sms professionnels³⁵, en dehors de la présence du salarié, sauf ceux qui sont clairement identifiés comme « personnels » ou « privés » ;
 - Messagerie personnelle : l'employeur n'a pas le droit de consulter les mails personnels, même s'ils ont été envoyés depuis l'ordinateur professionnel³⁶.
- Sites Internet : la consultation de sites Internet pendant le temps de travail grâce aux outils informatiques professionnels est présumée avoir un caractère professionnel de sorte que l'employeur peut les identifier, hors la présence du salarié³⁷.
- Dossiers et fichiers :
 - Les dossiers et fichiers créés par un salarié grâce aux outils informatiques

27. Art. 35 du RGPD.

28. [Cass. soc., 14 mars 2000, n° 98-42.090.](#)

29. Art. L. 1121-1 du Code du travail.

30. Art. L. 1222-4 du Code du travail.

31. Art. L. 2312-37 du Code du travail.

32. Art. 9 du Code civil.

33. CEDH, 5 sept. 2017, aff. Barbulescu c. Roumanie, n° 61496/08.

34. [Cass. soc., 26 juin 2012, n° 11-15.310.](#)

35. [Cass. com., 10 févr. 2015, n° 13-14.779.](#)

36. [Cass. soc., 7 avr. 2017, n° 14-27.949.](#)

37. [Cass. soc., 9 juill. 2008, n° 06-45.800.](#)

professionnels sont présumés avoir un caractère professionnel de sorte que l'employeur peut y avoir accès hors sa présence³⁸ ;

- L'employeur peut consulter les fichiers présents sur l'ordinateur professionnel du salarié, sauf si ces fichiers sont identifiés comme « personnels » ou « privés »³⁹.
- Télétravail : l'employeur peut contrôler l'activité des salariés en télétravail si cela ne porte pas atteinte aux droits des salariés, notamment au droit du respect de leur vie privée. Ce contrôle ne saurait être exercé de manière excessive. Ainsi, l'employeur ne peut pas surveiller de manière permanente ses salariés en télétravail (interdiction par exemple de la surveillance constante du salarié en télétravail au moyen d'une webcam).

Vidéosurveillance

- Les caméras peuvent être utilisées à des fins de sécurité des biens et des personnes, ainsi que pour contrôler l'activité des salariés⁴⁰.
- L'employeur ne peut pas mettre en œuvre un dispositif de contrôle de l'activité professionnelle qui n'a pas été porté préalablement à la connaissance des salariés. Toutefois, il peut leur opposer les preuves recueillies par des systèmes de surveillance qui n'ont pas pour objet le contrôle de l'activité des salariés⁴¹.
- C'est à l'employeur de fixer la durée de conservation des images, qui doit être proportionnée à l'objectif poursuivi par les caméras. En principe, elle ne peut excéder 1 mois⁴².

Géolocalisation

- Si l'employeur a désigné un Délégué à la protection des données (DPO), celui-ci doit être associé à la mise en œuvre du dispositif. Le système de géolocalisation doit être inscrit au registre des activités de traitement tenu par l'employeur.
- Le recours à la géolocalisation n'est pas justifié s'il existe des dispositifs moins intrusifs⁴³, même si ces dispositifs sont moins efficaces que la géolocalisation⁴⁴.
- Son recours est interdit pour les salariés bénéficiant d'une liberté dans l'organisation de leur travail (salariés itinérants, VRP, visiteurs médicaux, etc.)⁴⁵. Par exemple, la mise en place d'un système de géolocalisation sur le véhicule d'un vendeur est illicite si le salarié dispose d'une liberté dans l'organisation de ses déplacements, à charge pour lui de respecter le programme fixé et de rédiger un compte rendu journalier précis et détaillé⁴⁶.

Biométrie

- Les employeurs qui souhaitent mettre en place des accès biométriques doivent respecter un règlement type adopté par la CNIL le 10 janvier 2019. À défaut de le

38. [Cass. soc., 18 oct. 2006, n° 04-48.025.](#)

39. [Cass. soc., 18 oct. 2006, n° 04-47.400.](#)

40. [Cass. soc., 10 janv. 2012, n° 10-23.482.](#)

41. [Cass. soc., 19 avr. 2005, n° 02-46.295.](#)

42. [Délibération CNIL n°2010-112 du 22 avril 2010.](#)

43. [Cass. soc., 16 déc. 2020, n° 19-10.007.](#)

44. [CE, 15 déc. 2017, n° 403776.](#)

45. [Cass. soc., 19 déc. 2018, n° 17-14.631.](#)

46. [Cass. soc., 3 nov. 2011, n° 10-18.036.](#)

respecter, un tel contrôle ne peut être mis en place.

- Le règlement type oblige l'employeur à justifier son besoin de mettre en œuvre un dispositif biométrique par rapport à d'autres solutions de contrôle moins intrusives, et à respecter un cahier des charges rigoureux⁴⁷.
- Les badges biométriques ne peuvent être utilisés que pour contrôler :
 - L'accès aux locaux limitativement identifiés par l'employeur devant faire l'objet d'une restriction de circulation ;
 - L'accès à des appareils informatiques limitativement identifiés par l'employeur⁴⁸.
- Sauf circonstances exceptionnelles, les données biométriques ne peuvent être utilisées par les employeurs pour contrôler les horaires des salariés⁴⁹.

47. Art. 3 du règlement type de la CNIL du 10 janv. 2019.

48. Art. 2 du règlement type de la CNIL du 10 janv. 2019.

49. [Délibération CNIL SAN-2018-009 du 6 sept. 2018.](#)

Un aperçu des initiatives d'encadrement éthique des usages de l'IA

L'intervention des régulateurs internationaux s'est rapidement avérée nécessaire pour assurer la confiance et l'acceptabilité de l'IA dans la société. Ainsi, la proposition de règlement « AI Act » par la Commission européenne, publiée le 21 avril 2021, s'inscrit dans une dynamique plus vaste, mobilisant toutes les organisations supranationales. Depuis 2017, encouragées par leurs États membres, elles se sont mobilisées pour accompagner cette transformation et répondre aux inquiétudes des opinions publiques en élaborant des chartes proposant un cadrage éthique des usages de l'IA. On peut en distinguer plusieurs types :

- **Les chartes et recommandations visant à instaurer la confiance dans le cadre du développement et la sécurité des économies et des marchés.**

Dans ce domaine, la recommandation du Conseil sur l'intelligence artificielle de l'OCDE⁵⁰ publiée en mai 2019 fait office de référence. Cette recommandation, qui vise essentiellement à encourager le développement de politiques nationales et internationales, identifie une série de « principes pour une gestion responsable de l'IA digne de confiance », tels que la croissance inclusive, le développement durable et le bien-être, les valeurs centrées sur l'humain et l'équité, la transparence et l'explicabilité, la robustesse, la sûreté et la sécurité, et enfin la responsabilité.

On peut également relever que le projet d'AI Act de la Commission européenne⁵¹ s'est largement appuyé sur les travaux du HLEG (*high level expert group*) sur l'IA. Hors d'Europe, on retiendra la Déclaration de Montréal⁵² pour une IA responsable pour le Canada, les « *social principles of human centric AI* » du Japon⁵³, et les Beijing AI Principles⁵⁴ de la Chine.

- **Les chartes et recommandations visant à établir une confiance pour les droits des individus et le développement durable des sociétés.**

Sur cet aspect, la recommandation sur l'éthique de l'intelligence artificielle de l'UNESCO⁵⁵ parue en 2021 est la plus significative. Elle s'appuie sur 5 valeurs principales (le

50. <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449>

51. <https://digital-strategy.ec.europa.eu/fr/policies/expert-group-ai>

52. <https://www.declarationmontreal-iaresponsable.com/>

53. <https://www8.cao.go.jp/cstp/english/humancentricai.pdf>

54. <https://www-pre.baai.ac.cn/news/beijing-ai-principles-en.html>

55. https://unesdoc.unesco.org/ark:/48223/pf0000380455_fre

respect, la protection et la promotion de la dignité humaine ; les droits de l'Homme et les libertés fondamentales ; la prospérité des environnements et des écosystèmes ; la diversité et l'inclusion ; la paix) et 10 grands principes (proportionnalité et innocuité, sûreté et sécurité, équité et non-discrimination, durabilité, respect de la vie privée et protection des données, surveillance et décision humaines, transparence et explicabilité, responsabilité et redevabilité, sensibilisation et éducation, gouvernance et collaboration multipartites et adaptatives). L'ensemble de ces recommandations est concrétisé par des actions stratégiques. On peut citer par exemple l'évaluation de l'impact éthique (EIE) tout au long du cycle de vie d'un système d'IA comme instrument visant à mettre en évidence « les répercussions sur les droits de l'Homme et les libertés fondamentales ».

- **D'autres chartes éthiques se sont fixées pour objectif d'adresser des champs spécifiques.** Le Conseil de l'Europe a ainsi adopté la première Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires⁵⁶. On peut aussi noter la charte éthique de Pôle Emploi⁵⁷. Aux États-Unis, on retiendra l'adoption de principes éthiques pour l'IA par le Département de la Défense, et l'édition d'un guide pour une IA responsable⁵⁸.
- **Parallèlement, l'adoption de chartes éthiques ou de codes de conduite s'est développée dans les entreprises, notamment dans le secteur du numérique.** Ces chartes et codes constituent un cadre pour les pratiques et comportements liés à l'intégration et l'usage de l'IA au sein de l'entreprise. Elles ont pour principal objectif d'anticiper l'entrée en vigueur de réglementations en cours d'élaboration et de faciliter l'acceptation de ces nouveaux systèmes technologiques, par les consommateurs comme par les salariés. On peut citer pour exemple les codes de conduite de Google (« AI at Google: our principles »⁵⁹) ou de Microsoft (« Microsoft AI principles »⁶⁰), et les chartes éthiques d'organisations professionnelles telles que l'IEEE-CS – Institute for Electrical and Electronics Engineers – (« The IEEE Global initiative for Ethical Considerations in AI and Autonomous Systems »⁶¹), ou de la MAIF (« Pour un monde numérique résolument humain et éthique »⁶²).

Face à la prolifération de ces initiatives aussi bien publiques que privées, le Conseil de l'Europe propose un outil de data visualisation qui répertorie et cartographie plus de 500 références de ce type⁶³.

56. [https://www.coe.int/fr/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment#:~:text=La%20Commission%20europ%C3%A9enne%20pour%20l,IA\)%20dans%20les%20syst%C3%A8mes%20judiciaires](https://www.coe.int/fr/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment#:~:text=La%20Commission%20europ%C3%A9enne%20pour%20l,IA)%20dans%20les%20syst%C3%A8mes%20judiciaires)

57. <https://www.actuia.com/actualite/focus-sur-la-charte-de-pole-emploi-pour-une-utilisation-ethique-de-lintelligence-artificielle/>

58. <https://www.diu.mil/responsible-ai-guidelines>

59. [AI at Google: our principles](#)

60. [Microsoft AI principles](#)

61. [The IEEE Global initiative for Ethical Considerations in AI and Autonomous Systems](#)

62. [Pour un monde numérique résolument humain et éthique](#)

63. <https://www.coe.int/fr/web/artificial-intelligence/national-initiatives>

Accord-cadre européen des partenaires sociaux sur la numérisation⁶⁴

Les partenaires sociaux ont conclu le 22 juin 2020 un accord-cadre européen sur la numérisation. Ce texte offre un cadre au dialogue social en cours pour accompagner une politique de régulation des évolutions technologiques au travail. Si ce document ne traite pas directement de l'IA, ces recommandations restent globalement structurantes pour l'usage de ce type de technologie dans un cadre professionnel.

L'accord définit 4 enjeux majeurs : le développement des compétences numériques par la formation dans le but de sécuriser les emplois ; la définition de modalités de connexion et de déconnexion ; la nécessité de limiter la collecte de données dans le cadre professionnel à un but précis et transparent afin que la surveillance numérique se fasse dans le respect de la dignité humaine ; et l'importance de maintenir la prédominance du contrôle humain sur les systèmes d'intelligence artificielle afin de s'assurer qu'ils ne viennent pas dégrader les conditions de travail, mais contribuent au contraire à leur amélioration. L'accord préconise pour cela de procéder à une évaluation des risques liés à l'intégration de ces systèmes avant déploiement, et d'analyser leur mode de fonctionnement afin de faciliter leur explicabilité et d'éviter les possibles biais discriminatoires qui pourraient en découler. La question de la transparence de ces nouveaux systèmes est particulièrement critique dans le domaine des ressources humaines (recrutement, évaluation, promotion et licenciement) et suppose le renforcement du droit des travailleurs à pouvoir contester toutes décisions qui seraient prises à leur encontre sur la base de systèmes d'intelligence artificielle et à accéder à l'analyse et aux données ayant donné lieu à ces décisions.

64. <http://www.irshare.eu/fr/ue-accord-cadre-des-partenaires-sociaux-europeens-sur-la-numerisation-fr-1277.html>

Les machines intégrant de l'IA et la réglementation

La réglementation sur les machines

Deux directives européennes traitent des machines :

- La directive 2006/42/CE, dite directive « Machines », qui a pour objectif de définir un cadre de conception et de construction des machines pour respecter des exigences essentielles de santé et de sécurité. Ces exigences sont listées et définies dans l'annexe I de cette directive qui décrit le « marquage CE ».
- La directive 2009/104/CE, dite directive « Utilisation », qui quant à elle fixe des prescriptions minimales de sécurité et de santé pour l'utilisation – par les travailleurs au travail – des machines.

Le passage de la réglementation conception/construction à la réglementation utilisation s'opère lors de la mise en service. Celle-ci est clairement identifiée pour les machines conventionnelles, c'est une période d'installation et de réglage durant laquelle un transfert de propriété (ou simplement de responsabilité) est réalisé entre le constructeur et l'utilisateur. Dans le cas des machines qui embarquent un algorithme d'apprentissage, les frontières de la mise en service deviennent moins claires. En effet, cette technologie permet à une machine d'apprendre son métier pendant ses propres cycles de production. Dès lors, ces machines pourraient passer alternativement de phases de production à des phases de mise en service plusieurs fois au cours de leur vie. Ainsi, la question de la remise en cause du marquage CE par un nouvel apprentissage sur une machine en service pourrait se poser.

Une procédure de certification « CE » des machines est décrite dans la norme EN ISO 12100. Cette norme, d'application volontaire, donne aux concepteurs qui l'appliquent une présomption de conformité aux exigences essentielles de santé et de sécurité : elle est dite « harmonisée » à la directive 2006/42/CE. L'ISO 12100 décrit un procédé de conception itératif qui amènera à réduire les risques d'atteintes aux personnes que peut engendrer la machine. Ainsi, son application amènera le concepteur au marquage CE de sa machine.

De plus, la directive 2009/104/CE renvoie la responsabilité du maintien de cette certification à l'employeur durant toute la vie de l'équipement.

- Art.4 §2 : « *L'employeur prend les mesures nécessaires afin que les équipements de travail, tout au long de leur utilisation, soient gardés [...] à un niveau tel qu'ils satisfassent*

[...] [§1.a).i.)] aux dispositions de toute directive communautaire pertinente applicable. »

Cependant, comme nous l'avons vu précédemment, dans le cas d'une machine embarquant de l'IA, la machine – et donc son comportement – peut être amenée à évoluer. Pour respecter la réglementation actuelle, la solution trouvée consiste à limiter l'usage de l'IA apprenante à des fonctions ne remettant pas en cause la santé et sécurité des utilisateurs. Ainsi, l'analyse de risques effectuée lors de la conception doit montrer que les actions ou décisions prises par le système d'IA n'interviennent pas dans la sécurité de la machine. Les évolutions permises au système resteront contenues dans le registre fonctionnel et ne pourront pas engendrer de risques. Les deux réglementations, de conception et d'usage, restent effectives, les concepteurs gardent la responsabilité de la conformité de la machine et l'employeur garde celle de l'utilisation et l'entretien.

Limites actuelles

Comment démontrer que le système d'IA n'aura aucune conséquence sur la sécurité des personnes ? Au-delà des procédures de certification, les normes harmonisées aident à la conception de machines sûres répondant à la réglementation. C'est le cas, notamment, des robots collaboratifs qui s'appuient sur un corpus de normes complet. Ce n'est pas encore le cas pour les systèmes d'IA : il est donc complexe de garantir que le système d'IA n'impactera pas la sécurité des personnes. Actuellement, face à ce problème, la sécurité des machines est toujours assurée par un système annexe, indépendant, qui n'intègre pas de fonction à base d'IA.

Les lacunes de la normalisation

Pour assurer l'innocuité d'un système d'IA, le concepteur devra démontrer que, d'une part, les résultats du système sont répétables et explicables et, d'autre part, que ses évolutions ne permettront pas au système de sortir de ses limites initiales. Ainsi, le système d'IA pourra être qualifié d'IA de confiance. Ce cas échéant, les normes seront en capacité de standardiser des pratiques et usages pour une diffusion plus large de ces technologies. Un projet de réglementation spécifique aux systèmes d'IA est à l'étude en Europe, et envisage de s'appuyer sur ces futures normes.

Annexes

Liste des variables

Intitulé	Auteur·e·s
Composante Offre en IA	
Fiche 1 : Avancées technologiques de l'IA	T. Silvestre, B. Braunschweig, M. Sarrey
Fiche 2 : Disponibilité des ressources pour la production et le fonctionnement des dispositifs d'IA	M. Malenfer
Fiche 3 : Disponibilité des ressources humaines (hypothèses seules)	M. Malenfer
Fiche 4 : Acteurs et dynamique de la diffusion de l'IA, processus de normalisation	N. Fatès, M. Sarrey, J.-F. Soupizet
Composante acceptabilité et usages	
Fiche 5 : Acceptabilité sociale générale (citoyens) de l'IA	V. Mandinaud
Fiche 6 : Règlementation sur l'IA et les données	M. Bieri
Fiche 7 : Acceptabilité par le monde du travail	S. Halluin
Fiche 8 : Questions éthique sur l'acceptabilité des usages de l'IA en prévention	A. Olympio
Composante Travail et prévention des risques professionnels	
Fiche 9 : Croissance économique et géographie de la production	M. Héry
Fiche 10 : Démographie du travail	M. Héry
Fiche 11 : Organisation du travail et statuts des travailleurs	J. Clerté
Fiche 12 : Automatisation, cobotisation et équipements du travailleur	M. Sarrey
Fiche 13 : Responsabilités et mode de gestion de la prévention	J. Munoz

Avancées technologiques de l'IA

T. Silvestre, B. Braunschweig, M. Sarrey

Il s'agit d'explorer l'historique, l'état de l'art et les évolutions possibles des techniques et des outils de l'IA.

Nous aborderons l'IA sous différents aspects :

- Les technologies électroniques liées au calcul (calcul quantique, microprocesseurs et architectures plus sobres [*edge computing*], etc.)
- Les algorithmes, y compris les méthodes holistiques, celles liées à la confiance
- Les technologies liées aux données (capteurs, stockage, préparation des données, interopérabilité)

Cette fiche s'attachera également à évaluer la maturité technologique des IA en termes de fiabilité et de sécurité (robustesse, processus de validation, qualification des résultats, cybersécurité, prévention des biais et erreurs).

Définitions

Il existe de très nombreuses définitions de l'intelligence artificielle (IA), aucune ne répondant totalement à la question, car la discipline a des contours flous.

On peut définir l'IA par intension, au moyen d'une phrase, comme l'ont fait la Commission d'enrichissement de la langue française et la publication au Journal officiel du 9 décembre 2018 : « *Champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines* ».

Un autre exemple en intension est celui de la Commission européenne, la même année : « *Même s'il n'y a pas de définition consensuelle, le terme « IA » fait référence à des systèmes qui exhibent des comportements intelligents en analysant leur environnement et en effectuant des actions avec un certain degré d'autonomie, pour atteindre des objectifs spécifiques. Les systèmes d'IA peuvent être purement logiciels, agissant sur le monde virtuel. Des exemples sont les agents conversationnels, les logiciels d'analyse d'images, les moteurs de recherche et les systèmes de reconnaissance de visages. D'autres systèmes d'IA sont ceux où l'IA est embarquée dans des appareils comme les robots avancés, les voitures autonomes, les drones ou les applications de*

l'Internet des Objets. En contraste avec les systèmes d'IA logiciels, de tels systèmes perçoivent leur environnement grâce à des capteurs, et agissent dessus ou s'y déplacent, exigeant donc des mesures de sécurité plus robustes. »

Une autre approche est de définir l'IA en extension, par la somme des diverses sous-disciplines qui la composent : apprentissage automatique, raisonnement automatique, traitement de la langue naturelle, vision artificielle, représentation des connaissances, etc. C'est le choix fait par l'Académie des Technologies dans son rapport « Renouveau de l'Intelligence artificielle et de l'apprentissage automatique » paru également en 2018 :

- Perception, un domaine très vaste contenant la reconnaissance des images, des formes, et des sons (et de toutes formes de signaux produits par des « senseurs »). La reconnaissance des images est le domaine le plus important (on parle de *machine vision*) à cause de ses applications et de par les progrès spectaculaires de ces dernières années cités en introduction, et liés au *deep learning* ;
- Traitement du langage naturel, sous sa forme écrite ou orale (ce qui implique de le coupler avec la perception de la parole) ;
- Planning et navigation, qui pourraient être étendus aux domaines de la résolution de problèmes, formalisés dans l'univers de la recherche opérationnelle ;
- Représentation de connaissances, qu'il s'agisse de les manipuler, de les rechercher ou de les combiner. Ce domaine a fortement évolué avec le développement des mégadonnées (Big Data) ;
- Raisonnement logique lié à la forme symbolique de l'intelligence artificielle, rendue célèbre dans les années 80 par les « systèmes experts ».

Pour les besoins de cette fiche, nous retiendrons ces trois définitions, compatibles entre elles. On y retrouve notamment les capacités données à des machines (matériel et logiciel) de réaliser des tâches demandant de l'intelligence quand elles sont exercées par des humains, des exemples d'applications de ces capacités (véhicule autonome, agent conversationnel, reconnaissance d'images, etc.), et des technologies au service de ces fonctions (représentation des connaissances, raisonnement, apprentissage, planification, etc.).

Historique de l'IA, une histoire de 80 ans

Les prémices

Elles datent des années 1940 et de la Seconde Guerre mondiale, avec la conception, par Warren McCulloch¹ et Walter Pitts², du premier modèle de neurone formel en 1943³ : une unité simple capable de faire la somme pondérée de données en entrée, d'activer une fonction seuil et de donner une sortie. Aujourd'hui, cette représentation reste un élément de base des réseaux de neurones artificiels.

1. Warren McCulloch (1898-1969), chercheur américain en neurophysiologie et cybernétique.

2. Walter Pitts (1923-1969), chercheur américain en psychologie cognitive.

3. W.S. McCulloch, W. Pitts, « A Logical Calculus of the Ideas Immanent in Nervous Activity », Bulletin of Mathematical Biophysics, 1943.

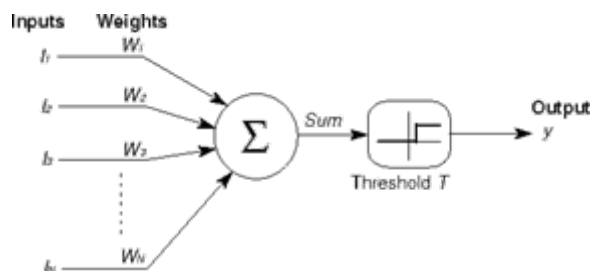


Figure 1. Représentation d'un neurone artificiel.

Dans le même contexte, le mathématicien et cryptologue britannique Alan Turing joue un rôle essentiel en décryptant le code Enigma utilisé par les armées allemandes dans leurs transmissions. Considéré comme un des pionniers de l'intelligence artificielle, il travaille sur l'un des tout premiers ordinateurs, et participe à la réflexion sur le potentiel de l'intelligence artificielle. À cet effet, il crée le test de Turing⁴ en 1950, un test d'intelligence artificielle fondé sur la faculté d'une machine à imiter la conversation humaine⁵.

La recherche en intelligence artificielle a été conceptualisée lors du congrès de Dartmouth, qui réunit en 1956 les meilleurs spécialistes internationaux de l'époque. Parmi eux étaient notamment présents John McCarthy, Marvin Minsky, Claude Shannon, Alan Newell, Arthur Samuel, Herbert Simon ou encore Nathan Rochester d'IBM. C'est à cette occasion qu'est utilisé pour la première fois le terme d'« intelligence artificielle » : « Tout aspect de l'intelligence, et notamment de l'apprentissage, peut être décrit de manière tellement précise qu'une machine peut le simuler »⁶. Cet événement est considéré aujourd'hui comme le moment fondateur de la discipline.

L'âge d'or

Juste après ce congrès, les scientifiques s'attaquent à une grande diversité de recherches. La résolution de problèmes par des systèmes artificiels, le traitement du langage naturel, la traduction automatique et la résolution de jeux sont autant de sujets de recherche qui passionnent toute la communauté. C'est d'ailleurs dans le courant des années 1950 qu'apparaît le premier programme informatique d'auto-apprentissage sur le jeu de dames, développé par le chercheur américain Arthur Samuel. Il réalise alors avec sa machine une des premières démonstrations du concept fondamental de l'intelligence artificielle. À peu près au même moment, le principe de résolution par la logique, défini par Robinson en 1965, lance une ère de l'IA symbolique.



Arthur Samuel
devant son programme
d'apprentissage automatique,
baptisé Chinook.

4. A. Turing, « Computing machinery and intelligence », 1950.

5. [Wikipedia.org/wiki/Test_de_Turing](https://fr.wikipedia.org/wiki/Test_de_Turing)

6. Proposal of the Dartmouth Conference.

Les années de doute

Au début des années 1970, les capacités des programmes d'IA sont limitées. Les plus performants peinent à manipuler des versions simplistes de problèmes qu'ils sont supposés résoudre. La puissance et la mémoire de l'époque étaient considérées à juste titre comme un véritable frein aux applications pratiques. Par exemple, le travail de Ross Quinlan sur le langage naturel est limité à un vocabulaire de vingt mots, car la mémoire ne peut pas en contenir plus. La recherche en vision par ordinateur et en robotique a fait peu de progrès. La publication du livre de Minsky et Papert en 1969, intitulé *Perceptrons*, donnera un coup d'arrêt à la recherche connexionniste en montrant les limites de ce que les perceptrons peuvent résoudre. Il y est démontré qu'un réseau de neurones tel qu'on le concevait à l'époque – appelé perceptron – ne pouvait pas appréhender les phénomènes non linéaires. Si l'on considère l'exemple du XOR – le « ou exclusif » –, un problème non séparable linéairement, il était impossible aux perceptrons de l'époque d'apprendre à le traiter en une seule couche.

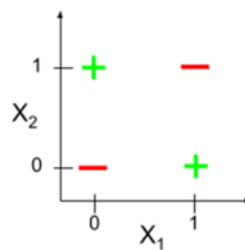


Figure 2. Représentation graphique du XOR ou « ou exclusif ».

Beaucoup de fonds et de subventions destinés à l'innovation en matière d'intelligence artificielle se voient lourdement réduits tandis que plusieurs projets sont aussitôt abandonnés. Le pessimisme quant aux possibilités réellement offertes par l'intelligence artificielle se répand, et l'intérêt pour le domaine de recherche s'estompe. C'est ce que la communauté nomme le « premier hiver » de l'intelligence artificielle.

Le nouveau souffle

Les années 80 voient le boom des systèmes experts : des programmes répondant à des questions ou résolvant des problèmes dans un domaine de connaissance donné, à l'aide de règles logiques dérivées de la connaissance des experts humains de ce domaine. C'est le développement de langages dédiés, tels que le PROLOG et le LISP, qui permettent d'optimiser des processus dans les entreprises et ainsi de les faire gagner en efficacité. Durant cette décennie, on dépense des millions pour l'ordinateur 5^e génération, qui est censé traduire et dialoguer comme un être humain. Les investissements repartent à la hausse au profit des systèmes experts et, en parallèle, le logiciel Chinook devient le premier programme informatique à résoudre entièrement le jeu de dames, c'est-à-dire qu'il est certain de gagner la partie, quelle que soit la situation de départ.

L'hiver des systèmes experts... printemps de l'apprentissage profond

L'un des principaux goulets d'étranglement de l'IA symbolique est la quantité et la haute qualité du travail humain nécessaires pour modéliser un domaine de connaissance, aussi étroitement circonscrit soit-il. Certaines connaissances humaines sont très difficiles à extraire et à modéliser : les systèmes experts se sont révélés utiles, mais uniquement dans des contextes très spécifiques.

Par ailleurs, les systèmes experts ont un coût de maintenance trop élevé : ils sont difficiles à mettre à jour et ne peuvent pas apprendre pour évoluer. De plus, ils peuvent faire des erreurs grossières quand les paramètres sortent des valeurs habituelles.

L'IA symbolique fonctionne bien dans les micro-mondes fermés des jeux ou des laboratoires, mais se trouve rapidement dépassée dans les environnements ouverts qui ne répondent pas à un petit nombre de règles strictes.

Encore une fois, les résultats obtenus sont loin d'être à la hauteur des promesses formulées. C'est le deuxième hiver de l'IA, celui des systèmes experts.

La discipline va pourtant bénéficier d'une phase d'expansion sans précédent : la puissance de calcul des ordinateurs augmente de manière gigantesque ; la plupart des espaces de l'activité humaine (des supermarchés aux banques, en passant par les hôpitaux et l'école) s'appuient sur les ordinateurs, ce qui fournit des données abondantes. Répondant à cette augmentation de la matière (les données) et des moyens (les capacités de calcul), les algorithmes émergent.

En 1986, le connexionnisme renaît avec une nouvelle méthode : la rétropropagation du gradient (Rumelhart, 1986). C'est aussi l'époque où de nouvelles avancées sur les réseaux neuronaux – notamment la convolution, qui permet le partage des poids des connexions et qui se révèle particulièrement efficace pour le traitement de signaux et d'images (par exemple, sur la reconnaissance de caractères manuscrits) – sont expérimentées par Yann Le Cun, qui verra ses travaux récompensés par un prix Turing (partagé avec Yoshua Bengio et Geoffrey Hinton). Il faudra attendre l'avènement du Web et des masses de données qu'il rend disponibles, combinées à l'augmentation toujours exponentielle des capacités de calcul, pour que ces travaux débouchent enfin sur des applications importantes au début des années 2000.

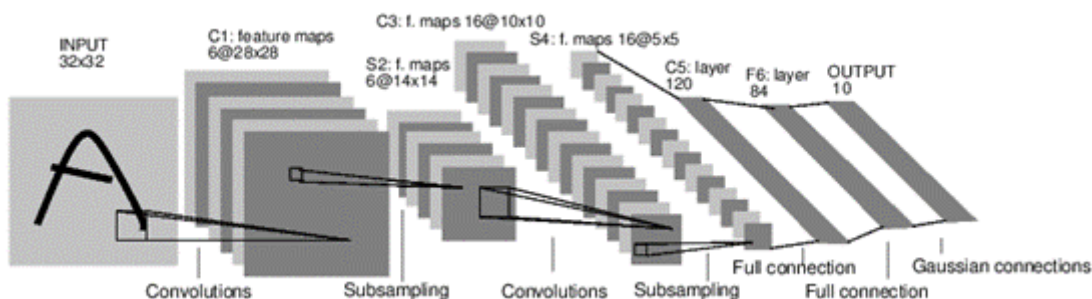


Figure 3. Architecture mise au point par Yann Le Cun pour la reconnaissance de caractères manuscrits.

À la même époque, le logiciel DeepBlue, développé par la firme américaine IBM, bat le champion du monde Garry Kasparov aux échecs en 1997.

Le virage du Web dans les années 2000

L'une des étapes clés du développement de l'intelligence artificielle telle que nous la connaissons aujourd'hui est l'usage du Web et l'apparition de ses premières applications. L'intérêt économique était tel, que tous les algorithmes, les méthodes ou les techniques développés pendant les cinquante précédentes années (et qui étaient clairement tombés en désuétude) pouvaient être réexploités pour ce secteur.

Un constat est d'ailleurs sans appel : aujourd'hui, quand on parle des géants de l'IA, de qui parle-t-on réellement ? Google, Meta, Microsoft, Amazon, Baidu, Alibaba... les géants du Web !

A partir des années 2010, l'apprentissage profond envahit la discipline. Cette technique se développe dans à peu près tous les domaines (reconnaissance visuelle, robotique, bio-informatique, etc.) et beaucoup d'espoirs sont à nouveau formulés.

L'histoire de l'IA, et notamment de ses branches « IA basée sur les données » (numérique) et « IA basée sur les modèles » (symbolique) est bien résumée dans la figure ci-dessous, extraite d'un article de D. Cardon et coll.

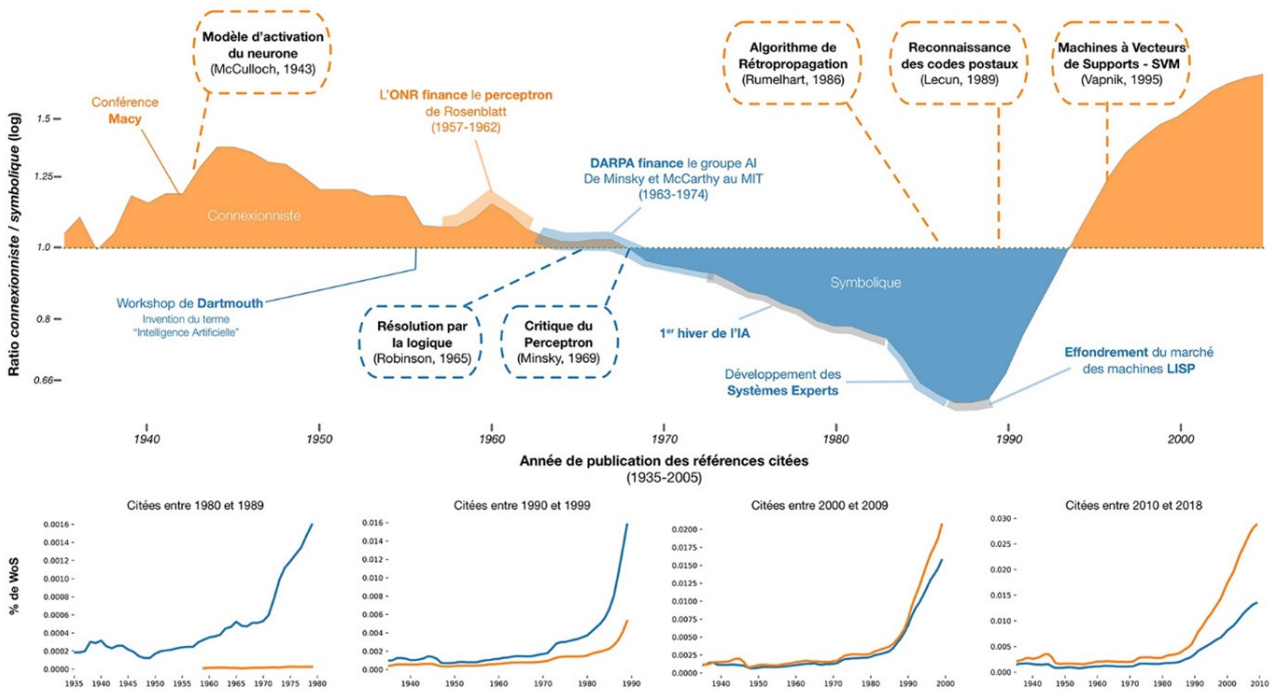


Figure 4. « La revanche des neurones », D. Cardon, J.P. Cointet, A. Mazières

Les avancées algorithmiques, combinées à la disponibilité de moyens de calcul de très grande ampleur et de données d'entraînement en masse, ont permis à l'apprentissage profond (ou *deep learning*) d'obtenir des résultats jusqu'alors inespérés dans de nombreux domaines d'application divers : jeux, systèmes de recommandation, reconnaissance d'images, de la parole, traitement de la langue naturelle (génération, traduction, résumés automatiques, questions-réponses...), diagnostic – en particulier médical, prévision de séries temporelles, etc. Les exemples sont innombrables et font l'objet de fréquentes communications et publications. Les trois graphiques suivants, tirés du AI index 2021⁷, montrent les progrès effectués dans trois domaines exemplaires : reconnaissance d'images (taux de succès des campagnes ImageNet) ; traitement du langage naturel (taux de

7. <http://AIindex.org>

réponses justes pour le Stanford Question Answering Dataset SQuAD) ; découverte scientifique (compétition de repliement de protéines CASP – *Critical Assessment of Protein Structure competition*).

ImageNet: Top-5 Accuracy

Top-5 accuracy asks whether the correct label is in at least the classifier's top five predictions. Figure 2.1.2 shows that the error rate has improved from around 85% in 2013 to almost 99% in 2020.²

IMAGENET CHALLENGE: TOP-5 ACCURACY

Source: Papers with Code, 2020; AI Index, 2021 | Chart: 2021 AI Index Report

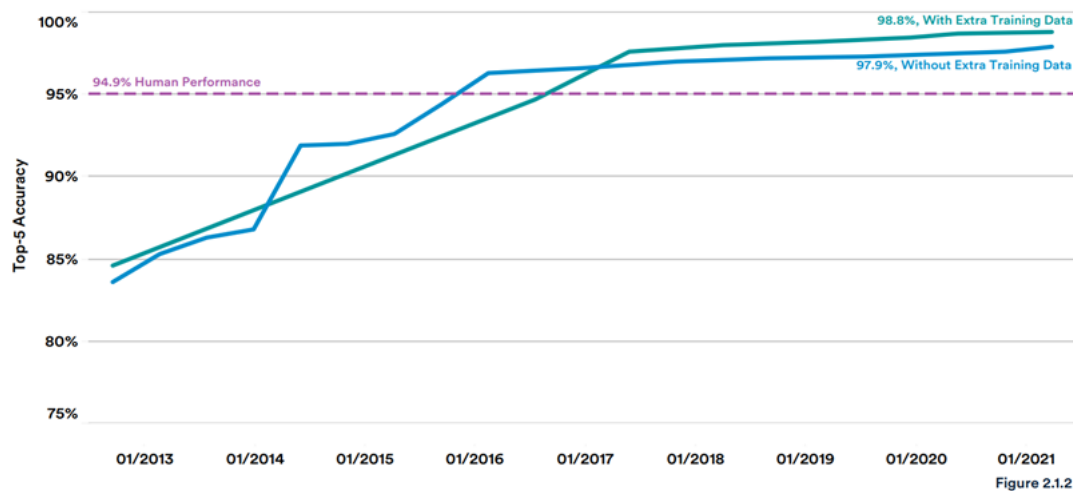
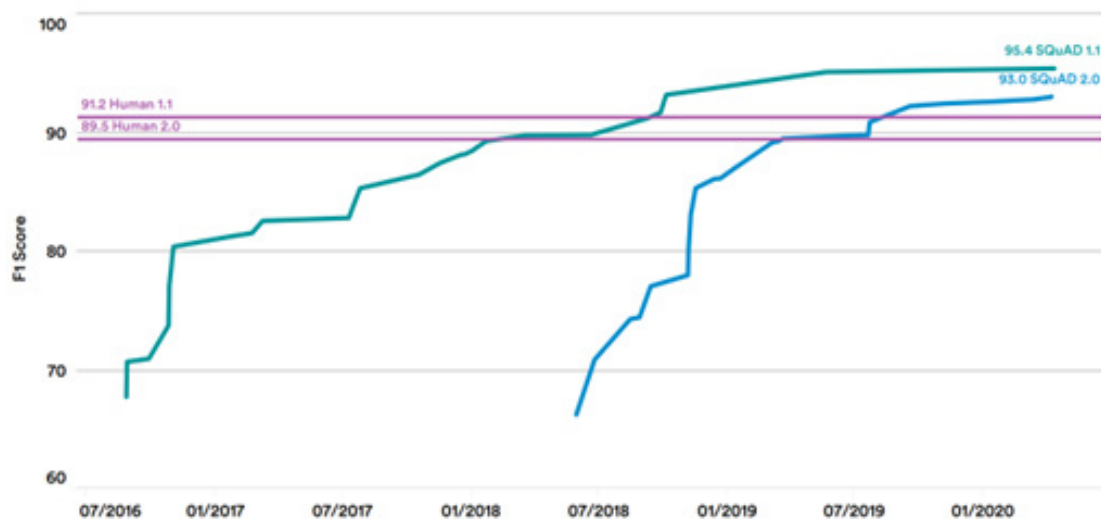


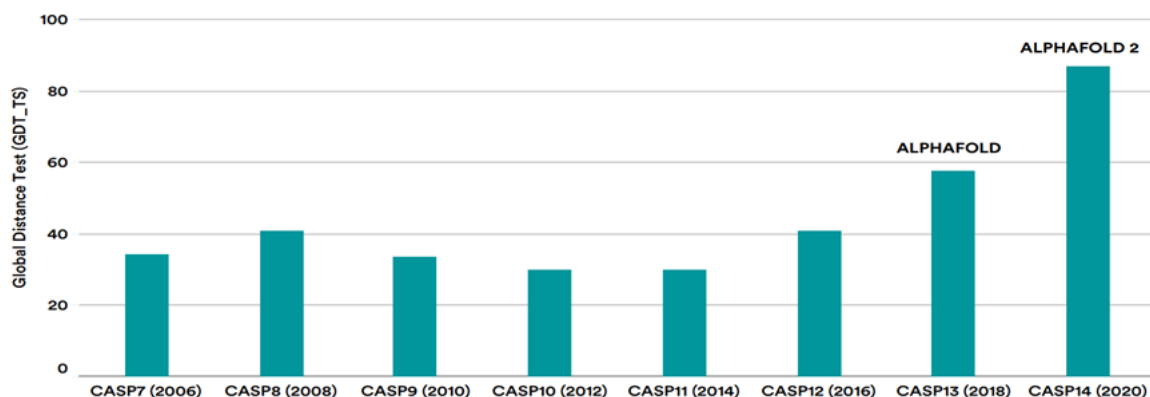
Figure 5. Taux de précision dans la reconnaissance d'image du challenge ImageNet.

SQUAD 1.1 and SQUAD 2.0: F1 SCORE

Source: CodaLab Worksheets, 2020 | Chart: 2021 AI Index Report



CASP: MEDIAN ACCURACY of PREDICTIONS in FREE-MODELING by THE BEST TEAM, 2006-20
Source: DeepMind, 2020 | Chart: 2021 AI Index Report



L'essor du *deep learning* (DL) à partir des données massives et d'une nouvelle puissance de calcul dévoile de nombreux biais et limites actuels. Après les avancées foudroyantes des années 2010 en matière d'apprentissage automatique par les réseaux neuronaux (AlexNet remporte la compétition ImageNet en 2012 et, en 2016, Alphago bat Lee Sedol au jeu de go), les progrès semblent marquer le pas depuis quelques années. Ces applications restent hyperspécialisées dans des domaines comme les mathématiques, les jeux, la recherche opérationnelle... Aujourd'hui, de nouvelles attentes et prophéties sont de plus en plus envisagées – celles-ci pourraient provoquer des désillusions... En effet, pour obtenir des performances marginalement meilleures, il faut désormais multiplier par plusieurs ordres de grandeur la taille des ensembles de données et la puissance de calcul utilisée pour entraîner les modèles.

On s'aperçoit que les réseaux de neurones sont soumis à de forts biais d'apprentissage (intrinsèques à l'humain) et au manque de données qualifiées, car le DL nécessite des quantités colossales de données souvent pré-traitées par des humains. « L'IA dans le domaine de la santé n'est pas encore robuste : chaque réussite l'est sur un cas particulier » (cf. rapport OECD « Trust in Health AI »). Le 22 juin 2021, le NIST a publié un nouveau document proposant une méthodologie pour l'identification et la gestion des biais dans les algorithmes d'intelligence artificielle.

En termes de développement durable, il y a également une prise de conscience de l'impact énergétique de ces systèmes d'apprentissages : l'ordinateur le plus puissant du CEA (Tera 1000) doit fonctionner pendant 300 jours pour réapprendre le modèle GPT3 – on parle de zettaflops de calculs et de millions de kWh^{8,9}. Il existe aussi des problèmes imprévisibles qui feront « planter » la machine, comme le cas du véhicule autonome Uber qui reconnaît un cycliste, un piéton, mais pas un piéton qui pousse un vélo ! Puisque leurs résultats ne peuvent être ni expliqués ni justifiés de manière conceptuelle – c'est-à-dire sur un mode compréhensible par les humains –, il est difficile de faire confiance à ces

8. E. Strubell, A. Ganesh, A. McCallum, « Energy and Policy Considerations for Deep Learning in NL », 2019, <https://arxiv.org/abs/1906.02243v1>.

9. N. Thompson et coll., « The Computational Limits of Deep Learning », 2020, <https://arxiv:2007.05558v1>

modèles pour le moment.

Aujourd'hui, on distingue deux systèmes de pensée, nommés « système 1 » et « système 2 » par le professeur Daniel Kahneman. Le système 1 correspond aux automatismes : par exemple, faire des tâches de la vie de tous les jours, comme conduire sur une route connue. Le système 2 est plus raisonné et réfléchi, il demande un effort de concentration : par exemple, conduire de nuit dans une ville où l'on cherche sa route. Avec l'apprentissage machine profond, l'IA est actuellement au niveau du système 1. Pour pouvoir dépasser cela, représenter les connaissances de sens commun, faire de la planification, des raisonnements élaborés, il faudra coder le niveau du système 2. Mais les avis diffèrent sur la manière d'y arriver : certains pensent qu'il suffit d'empiler des couches de neurones artificiels en organisant très finement leur architecture et leur communication ; d'autres pensent que des paradigmes différents de la représentation de l'intelligence humaine seront nécessaires (par exemple, en hybridant l'apprentissage machine avec le raisonnement symbolique utilisant des règles, des faits, des connaissances).

Critères d'impacts de l'IA aujourd'hui

Les critères permettant de mesurer les limites/faiblesses ou forces de l'IA aujourd'hui sont détaillés ci-dessous, autour de ce que l'on appelle l'IA de confiance. Il y a 13 critères :

Technologie (T)

- Robustesse & fiabilité (taux d'erreurs, instabilités...)
- Précision & performance (latence, exactitude et pertinence des résultats)
- Cybersécurité (risques de dévoiement, résistance aux malveillances, attaques...)
- Transparence & traçabilité (accès aux données, enregistrements des activités, notifications, véracité des interlocuteurs...)
- Explicabilité (pouvoir expliquer un choix ou un résultat)

Métiers (M)

- Responsabilité & redevabilité (à qui incombe les conséquences d'une erreur)
- Maîtrise (avoir la main, *monitoring*, réponse exacte à ce que l'on a demandé, et ne pas perdre le contrôle à tout moment du processus)
- Légalité & conformité (ISO, LNE, AFNOR, normes, certificats, autorisations...)
- Sûreté & *safety* (dangers physiques ou psychologiques pour les usagers ou les tiers)

Social (S)

- Équité (risques de discrimination dans les résultats...)
- *Privacy* (respect de la vie privée et des données personnelles)
- Inclusion (s'adresser à tout type de public quel que soit l'âge, l'origine ethnique...)

■ Durabilité (impact sur l'environnement, consommation énergétique)

Pour simplifier l'exercice par la suite, nous regrouperons ces différents critères en 4 indicateurs :

- **Performance** : niveau de performance atteint par les meilleurs systèmes d'IA sur les problèmes *benchmarks*, reconnaissance d'images, compréhension de textes (questions-réponses), reconnaissance d'actions dans les vidéos. Mesuré annuellement par AI Index, par exemple.
- **Énergie** : performance énergétique, consommation en kWh nécessaire à l'entraînement du plus gros système d'IA (*a priori* transformé avec optimisation d'architecture). Souvent déclaré dans les publications sur les systèmes concernés.
- **Sécurité** : résistance aux attaques adverses, capacité de maintenir une performance équivalente lorsqu'il y a des attaques, ou capacité de détecter l'attaque et de ne pas produire de sortie. Mesuré dans les publications sur les systèmes concernés.
- **Confiance** : niveau de confiance accordé par les professionnels aux systèmes d'IA avec lesquels ils collaborent. Mesuré par une enquête sociologique menée auprès des professionnels.

Les limites de l'IA

Nous allons illustrer par quelques exemples les limites de l'IA actuelle en utilisant les indicateurs cités précédemment.

Performance

Les réseaux de neurones plafonnent en termes de performance, et la quantité de données qu'il faut fournir à un réseau augmente toujours beaucoup plus que l'augmentation du niveau de performance. On atteint une asymptote (exemple de la reconnaissance d'images en figure 5).

Énergie

« Quand la quantité de calcul nécessaire aux applications les plus avancées explose d'un facteur 10 chaque année, l'efficacité énergétique des ordinateurs s'améliore d'un facteur 10... en gros tous les huit ans. »¹⁰

Sécurité

Il existe deux types de sécurité pour les systèmes IA :

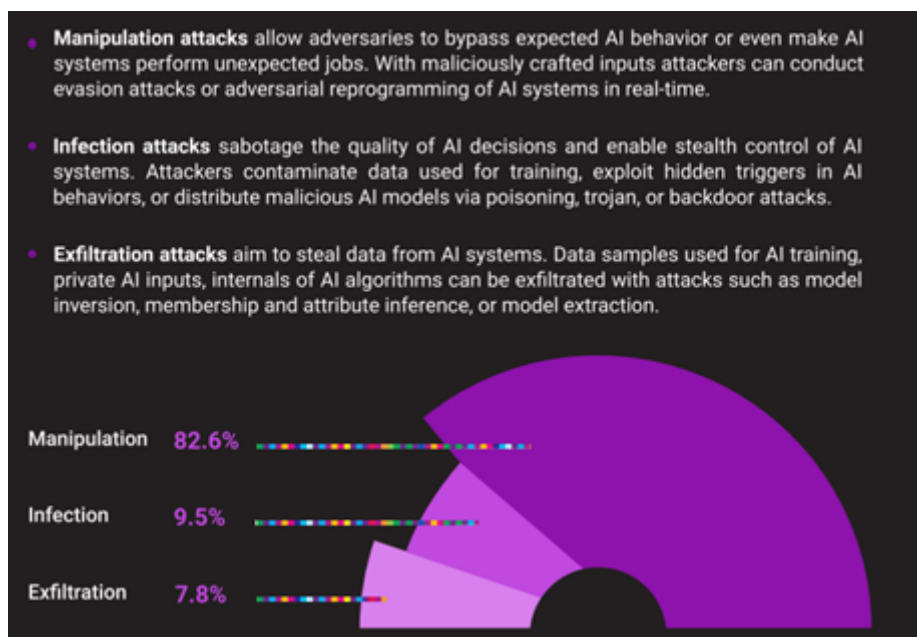
10. <https://blog.link-value.fr/quelle-est-lempreinte-%C3%A9cologique-des-entra%C3%AEnements-de-mod%C3%A8les-en-data-science-25aa07beb7a3>

- Les vulnérabilités intrinsèques des algorithmes et systèmes ;
- Les vulnérabilités que des systèmes IA peuvent générer elles-mêmes (réseaux antagonistes génératifs, ou « GAN »).

Vulnérabilités intrinsèques

Les systèmes d'IA d'aujourd'hui souffrent d'un certain nombre de nouvelles vulnérabilités non résolues. Celles-ci sont distinctes des vulnérabilités logicielles traditionnelles (par exemple, les débordements de mémoire tampon) et démontrent que, si les systèmes d'IA peuvent dépasser les performances humaines à bien des égards, ils peuvent également échouer de beaucoup d'autres manières. On distingue alors :

- Les manipulations (82 % des cas recensés) : on fausse les comportements ou les résultats d'un système d'IA, pouvant même être reprogrammés en temps réel.
- Les infections (10 %) : les jeux de données peuvent être contaminés et changent le comportement de l'IA, ce qui modifie la qualité des décisions ou permet de prendre le contrôle du programme.
- Les exfiltrations (8 %) : les jeux de données ou leurs modèles sont volés ou



remplacés.

Le MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) répertorie toutes les types d'attaques identifiées.

Vulnérabilités générées

Les IA peuvent servir à attaquer : c'est le cas des attaques GAN.

Les GAN servent également à générer les *deepfake*, qui sont de fausses vidéos détournant les propos d'une personne – généralement connue – de manière réaliste.

On peut voir des exemples de ces deux types d'attaques ci-après.

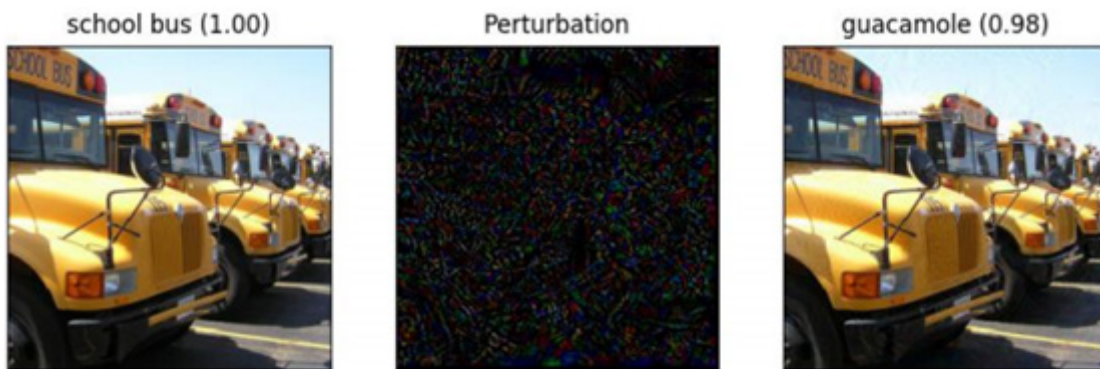


Figure 4: Illustration of an adversarial example using Basic Iterative Method [69]. The classifier used is Inception v3 [70]. The image comes from the validation of the ImageNet dataset [71] (left) Original image, correctly classified as a school bus. (middle) Perturbation added to the image, with a 10x amplification (right) Adversarial example, wrongly classified with high confidence.



On peut imaginer une multitudes d'autres attaques faites par les IA :

- Automatisation des attaques d'ingénierie sociale
- Automatisation de la découverte des vulnérabilités
- Déni de service à caractère humain
- Priorisation des cibles des cyberattaques à l'aide de l'apprentissage automatique
- Etc.

Confiance

Robustesse et fiabilité : soit on représente mal les données, soit les comportements humains changent.

Par exemple, le modèle pendant la crise de la Covid s'est trompé pour la consommation électrique (aspects logistiques, prévisions 1 à 7 jours sur l'énergie pour faire des échanges sur les marchés et relancer les bons modes de production énergétique...).

Sûreté : le lundi 19 mars 2018, un véhicule expérimental de la société Uber percutait une femme poussant un vélo, conduisant à son décès. Pourtant, une conductrice était présente derrière le volant afin de réagir rapidement en cas de situation dangereuse, mais elle a détourné les yeux à un moment critique. Le système aurait bien détecté le piéton, mais l'a classé comme un faux positif, c'est-à-dire un objet tel qu'un carton poussé par le vent ou un sac plastique. Cet évènement tragique montre comment une erreur de l'IA peut avoir des conséquences critiques sur la sûreté des systèmes dans lesquels ils sont embarqués¹¹.

Les tendances

Tendance 1

Les GAFAM (USA) et BATX (Chine) dominent l'évolution technologique. En 2020, l'investissement national de recherche et développement expérimental (R&D) en IA aux États-Unis a été de 1,5 Md\$, alors que l'investissement privé atteignait 80 Md\$. Les modèles d'apprentissage de plus en plus gros nécessitent des ressources gigantesques de DATA et CPU, dont seuls disposent les GAFAM et BATX.

Tendance 2

L'architecture des *Transformers* devient la norme pour toutes les applications de texte, et s'étend vers d'autres applications. Une de ses caractéristiques est de ne pas nécessiter d'annotation des données, ce qui permet de traiter des bases d'entraînement de plus en plus grandes (dizaines de milliards de caractères).

Tendance 3

Avec la diffusion très large des applications, initialement orientées vers le grand public (B2C), mais allant de plus en plus vers l'industrie et les systèmes critiques (B2B), les questions sociétales – qui ont des implications technologiques majeures – deviennent prépondérantes (confiance, responsabilité, biais, éthique de manière générale).

Tendance 4

Approches distribuées d'apprentissage (*federative learning, swarm AI...*) qui permettent plus de frugalité dans les quantités de données à fournir pour l'apprentissage et plus de *privacy* dans l'utilisation des jeux de données (notamment médicaux).

11. https://www.sciencesetavenir.fr/high-tech/transports/uber-le-vehicule-autonome-aurait-bien-identifie-la-cycliste-avant-l-accident-mais-n-a-pas-freine-ignorant-cette-information_123806

Tendance 5

Progrès du *hardware* (*in-memory computing*, *3D stacking*, *neuromorphic chips*...) et de l'IA embarquée (IA proche capteur, EDGE, FOG...) moins consommatrice d'énergie et de données (moins de latence et plus de *privacy*, moins de partage de données dans le Cloud).

Tendance 6

Émergence d'une IA hybride (neurosymbolique) qui pourrait permettre de réduire les erreurs liées au contexte et d'améliorer la robustesse, la fiabilité et la sûreté. Le fait qu'il y ait besoin de moins de données étiquetées permet de réduire le jeu de données.

Tendance 7

Augmentation du nombre de cyberattaques dans le monde (IA comprise), augmentant la multiplicité des failles (sécurité), mais aussi le développement de parades en réaction (robustesse).

Tendance 8

On arrive à une asymptote dans les performances de l'IA (quantité de données à fournir et problème de surparamétrisation) et il existe toujours des problèmes que l'on ne peut résoudre avec un bon niveau de confiance¹².

Tendance 9

Approches *open-source* poussées par les chercheurs, mais limitées par les risques d'applications duales (militaire, terrorisme). En effet, cela concentre la technologie dans les mains de gros acteurs (le modèle GPT3 de OpenAI n'est pas en *open-source*).

Incertitudes majeures

Incertitude 1

Les performances des systèmes d'IA atteindront-elles le niveau espéré et nécessaire pour des utilisations massives sur des systèmes à risques (hors B2C) ? C'est déjà le cas sur les jeux, la traduction automatique, le repliement de protéines, mais cela ne l'est pas sur la plupart des applications.

12. <https://spectrum.ieee.org/deep-learning-computational-cost>

Incertitude 2

La robustesse et la fiabilité des systèmes d'IA seront-elles atteintes dans tous les cas, en particulier pour les systèmes à risques, tels qu'identifiés par la Commission européenne ? Si ce n'est pas le cas, quelles mesures (politiques, économiques, technologiques...) seront prises pour garantir l'application de ces systèmes auprès des utilisateurs ?

Incertitude 3

Les demandes de capacité de calcul, de volume de données, et donc la consommation d'énergie des systèmes d'IA, continueront-elles de croître exponentiellement ? Si c'est le cas, quelles mesures (technologiques, économiques...) seront prises pour limiter ces facteurs ?

Incertitude 4

Verra-t-on l'émergence de nouvelles typologies d'attaques sur l'IA (et par l'IA), qui limiteraient son déploiement ? La « course aux armements » de la sécurité de l'IA sera-t-elle compensée par des actions externes dissuasives (de type social, politique, économique) ?

Incertitude 5

Le calcul quantique sera-t-il au point et mis au service des systèmes d'IA ? Si un ordinateur quantique voit le jour, ses capacités de calcul seront supérieures de plusieurs ordres de grandeur à celles des systèmes actuels, ce qui pourrait permettre une nouvelle croissance exponentielle des demandes de calcul des systèmes d'IA, augmentant en conséquence leurs performances.

Incertitude 6

Restera-t-on au niveau du système 1, ou bien arrivera-t-on à concevoir des IA (hybrides ou non) capables d'un raisonnement de système 2 ?

Incertitude 7

Arrivera-t-on à embarquer des systèmes IA performants dans les objets ? Et par qui ? GAFAM, BATX, NVIDIA, ARM, opérateurs spécialisés ? L'IA, pour être pleinement efficace, doit être embarquée dans les dispositifs que nous utilisons quotidiennement, ce qui pose de nombreux problèmes, dus aux contraintes spécifiques des systèmes embarqués (taille, poids, capacité de calcul, de mémoire, rapidité, consommation, connectivité).

Incertitude 8

Utilisations duales des technologies d'IA : application contrôlées par le gouvernement et non diffusées, car les conséquences de telles technologies peuvent être malhonnêtement détournées.

Hypothèses

Hypothèse 1. Trop de bugs, trop d'erreurs, pas de sécurité...

Défiance : 3^e hiver de l'IA

Malgré les recherches et développements sur l'IA de confiance et l'IA éthique (robustesse, équité, sécurité, explicabilité, certification...), mais qui ne sont applicables que sur un sous-ensemble très réduit des systèmes d'IA, de nombreuses applications sont mises en œuvre sans précautions suffisantes. Cela conduit à des accidents, des injustices, des mauvaises décisions non justifiées. Le grand public, mais aussi les professionnels, rejette la technologie, ce qui cause le troisième hiver de l'IA, un hiver social, contrairement aux deux précédents qui étaient technologiques.

Hypothèse 2. La technologie limite l'application à des secteurs non critiques

Les progrès en matière de confiance et d'éthique sont réels, mais ne s'appliquent qu'à un sous-ensemble réduit des systèmes d'IA (reconnaissance d'images, langage...), ce qui ne permet pas d'étendre les résultats, surtout lorsque les applications d'IA embarquées sont critiques et en temps réel (VA, machines-outils). Les développements technologiques en IA sont de plus fortement encadrés par la réglementation (principalement d'origine européenne), qui interdit de mettre en œuvre des systèmes sans prendre suffisamment de précautions sur leur applicabilité.

Le monde du travail et le grand public se satisfont de cette situation qui limite l'IA à des applications non critiques et sous supervision humaine. L'IA de confiance est toujours en développement et reste dans les mains d'un nombre limité de fabricants.

Hypothèse 3. Exubérance technologique malgré les problèmes résiduels

Les niveaux de robustesse et de fiabilité atteints par les solutions IA mises sur le marché sont satisfaisants pour la plupart des applications, avec un taux d'erreur encadré et normé. Les systèmes d'apprentissage profond continuent d'accroître leurs performances en intégrant de plus en plus de données, en consommant de plus en plus de ressources de calcul, mais les résultats sont tellement importants que leur utilisation se diffuse très largement dans tous les domaines d'application, critiques ou non, y compris au moyen de plateformes *open-source*. Des accidents se produisent, des décisions injustifiées sont prises, mais le bénéfice est tel pour les milliards de personnes utilisant au quotidien l'IA embarquée dans leurs objets numériques que l'exubérance se poursuit.

Hypothèse 4. L'hybridation conduit à des IA « système 2 »

Les recherches sur l'hybridation (neuro-symbolique) de l'IA aboutissent à des systèmes capables de dépasser le niveau perception, pour arriver au niveau système 2 de Kahneman, capables de conduire des raisonnements à partir des données et des connaissances dont ils disposent, dotés d'une certaine dose de sens commun, et en mesure d'expliquer leur

fonctionnement à leurs concepteurs, utilisateurs et certificateurs. Ces systèmes demandent beaucoup de développement et de ressources humaines pour les programmer, ils sont donc mis en œuvre avec parcimonie, mais donnent totalement satisfaction dans ce cadre.

Hypothèse 5. Rupture entre le quantique et l'IA générale

Ces deux ruptures sont mentionnées pour mémoire, mais ne sont pas envisageables à l'horizon de prospective de l'étude (2035-2040) :

1. L'ordinateur quantique devient une réalité, ses performances sont telles que l'on peut construire des applications d'apprentissage machine aux dimensions de plusieurs ordres de grandeur supérieures à ce qui est fait en 2022. L'IA peut s'attaquer à des domaines bien plus vastes et complexes : compréhension de scènes dynamiques complexes (vidéo), dialogue naturel avec les humains, pilotage d'équipements industriels lourds, véhicule totalement autonome, médecin artificiel... C'est un pas significatif pour la deuxième partie de ce scénario, celui de l'IAG.

2. Malgré ce que l'on pense aujourd'hui en 2022, une IA générale (donc supérieure à l'intelligence humaine) voit le jour suite à un énorme effort de développement pour préparer les données, faire l'apprentissage, réaliser l'interface humain-machine. Cette IA est donc capable de se reproduire et de s'améliorer. À partir de là, la prospective est impossible¹³.

13. Voir les écrits sur le sujet : Bostrom, Kurzveil, etc.

Disponibilité des ressources pour la production et le fonctionnement des dispositifs d'IA

M. Malenfer

Cette fiche vise à explorer ce dont l'IA a (ou aurait) besoin pour se développer sur le plan des équipements (processeurs, *data centers*, capteurs, ressources minérales, énergie...) et d'envisager la disponibilité et les éventuelles pénuries susceptibles d'entraver l'essor de l'IA.

La répartition de ces ressources entre acteurs sera prise en considération.

Définitions

Gisement : concentration naturelle de minéraux supérieure à la concentration moyenne permettant d'en envisager l'exploitation. Seuls certains gisements peuvent faire l'objet d'une exploitation à un moment donné. Si certains sont connus, d'autres restent à découvrir.

Ressources (minières) : ensemble des gisements connus susceptibles de faire l'objet d'une exploitation si les conditions techniques et économiques du moment le permettent. Elles sont définies sur la base de caractéristiques essentiellement géologiques.

Réserves : part des ressources qui peut techniquement et économiquement être exploitée au moment où elles sont déclarées. Elles sont estimées sur la base de critères géologiques et socio-économiques.

Les ressources peuvent donc être supérieures aux réserves.

Origine et qualité des données

L'estimation des ressources et des réserves est une démarche coûteuse qui est assurée par les sociétés minières. Bien que des standards internationaux aient été définis, les données sont plus ou moins précises et fiables en fonction de leur nature et de leur

origine. On considère que les sociétés minières cotées en bourse publient des données de qualité. En revanche, les sociétés totalement privées (« *private equity* ») ne publient généralement pas leurs données. Enfin, certains pays dans lesquels le secteur minier est très étatisé peuvent refuser de publier ces données, ou utiliser des normes différentes. De plus, ces données varient dans le temps en fonction des découvertes de nouveaux gisements et de la révision des réserves. La fiabilité et l'exhaustivité des données disponibles ne sont donc pas parfaites, ce qui rend les comparaisons et prévisions complexes.

Il est à noter qu'il n'existe pas pour les marchés des minéraux d'organismes comparables à ce qu'est l'OPEP pour le pétrole.

Indicateurs

Le traitement de cette problématique impose de manipuler différents indicateurs.

L'estimation des réserves est généralement exprimée en tonnage – mise en regard avec les tonnages extraits annuellement –, ce qui permet d'estimer une durée de disponibilité de la ressource en années.

Cet indicateur présente deux principales limites :

- Si la demande et les prix augmentent dans le futur, les investissements en exploration vont augmenter également, entraînant une augmentation des réserves estimées.
- Si la consommation de la ressource augmente de manière exponentielle, les quantités extraites annuellement vont augmenter également, et l'épuisement en ressources sera plus rapide que prévu.

Pour être le plus précis possible, il convient également de prendre en compte la teneur en matière recherchée du minerai extrait. Plus cette teneur sera faible, plus le processus sera consommateur d'énergie, coûteux et nuisible à l'environnement.

Rétrospective

Des ressources non renouvelables et inégalement réparties

Les ressources minérales sont le fruit de processus géologiques très longs (plusieurs millions d'années) et ne sont donc pas renouvelables à l'échelle de l'humanité. Leur genèse étant notamment liée à la tectonique des plaques, leur répartition sur la surface du globe est inégale : certaines régions, voire certains États, disposent de proportions importantes des réserves de certains minéraux, alors que d'autres en sont totalement dépourvues.

Une activité à fort impact environnemental

L'extraction et la production de métaux est fortement consommatrice d'énergie, on considère que 10 % de la consommation mondiale d'énergie y est consacrée. Durant la période allant de la moitié du XX^e siècle jusqu'à 2011, la consommation mondiale d'énergie a été multipliée par deux, alors que, sur cette même période, la consommation d'énergie consacrée à l'exploitation minière a été multipliée par quatre.

De plus, cette activité est très consommatrice d'eau, notamment les phases de broyage et de concentration du minerai. Ce sujet devient de plus en plus problématique pour l'industrie minière, car de nombreux projets sont situés dans des zones pauvres en eau où se posent des questions de concurrence entre les usages. Dans un contexte d'augmentation des tonnages extraits et de baisse des teneurs, le besoin d'eau continuera de croître même si des gains d'efficacité sont réalisés. Dans certains pays, comme le Chili, les besoins en eau sont en partie couverts par de l'eau de mer désalinisée, mais ce dessalement est lui-même consommateur d'énergie et augmente les coûts d'exploitation.

Au-delà du problème de l'extraction des matières premières, la consommation d'eau dans les processus de fabrication des composants informatiques nécessaires au développement d'IA, tels que les microprocesseurs ou les cartes graphiques, est très importante. À ce titre, la sécheresse à Taïwan est un facteur aggravant de la pénurie actuelle de semi-conducteurs¹.

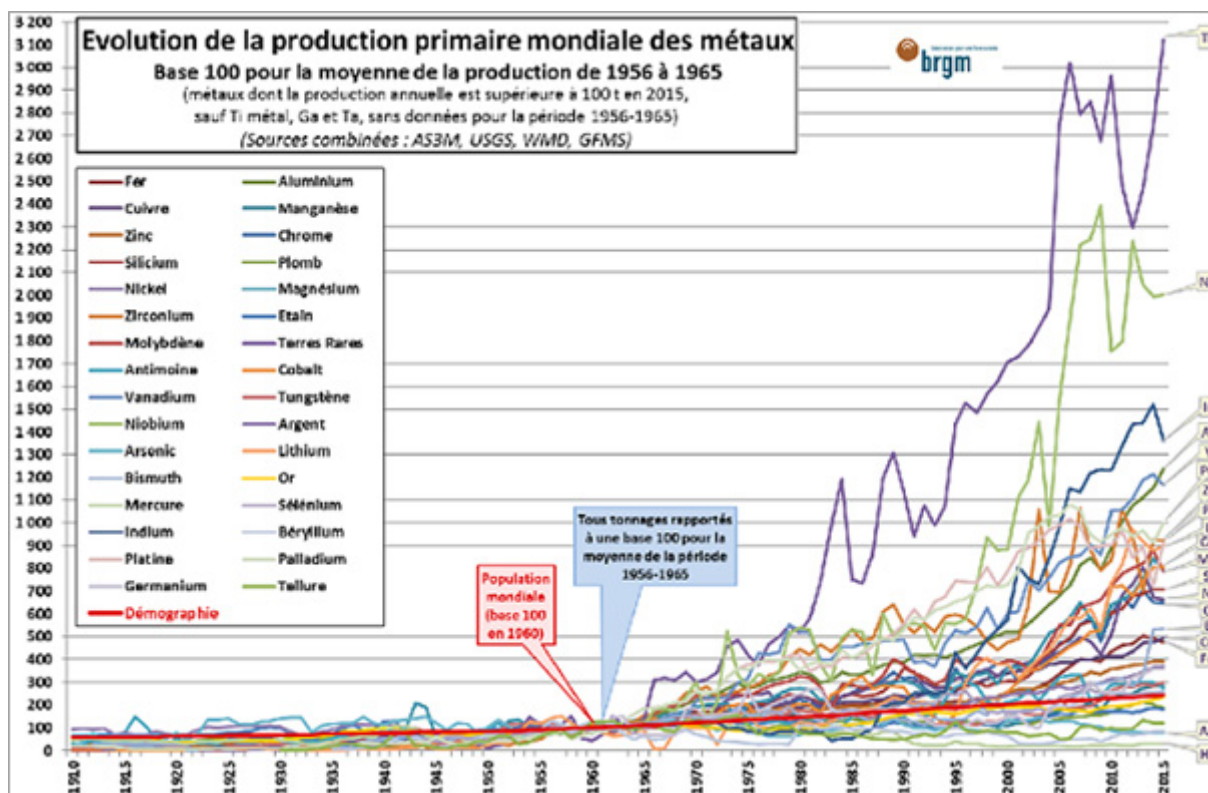
Enfin, les activités d'extraction ont un impact sur les écosystèmes et les paysages pouvant être lourd de conséquences et conduisant souvent les États à interdire ou encadrer fortement certaines activités.

Une demande qui explose

Entre 1900 et 2009, la population mondiale a été multipliée par 4. Sur cette même période, la consommation de minéraux a été multipliée par 30.

La consommation de ressources minérales connaît une croissance exponentielle de l'ordre de 3 % par an pour des métaux comme le fer ou le cuivre, de l'ordre de 6 % par an pour certains métaux rares. Cette croissance est largement corrélée à la croissance mondiale (PIB), qui est notamment tirée par le phénomène de rattrapage des économies occidentales par les nouvelles puissances industrielles, au premier rang desquelles se trouve la Chine.

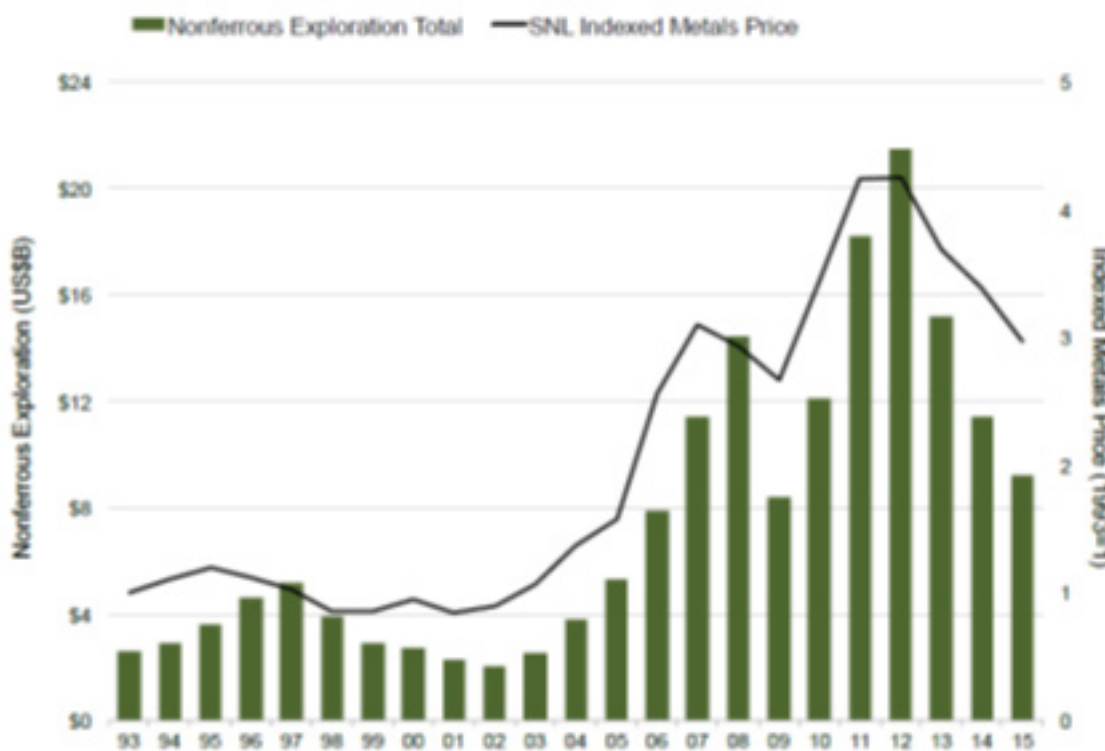
1. <https://www.O1net.com/actualites/pourquoi-la-secheresse-qui-frappe-taiwan-pourrait-aggraver-la-crise-des-semi-conducteurs-2037299.html>



Dans une fiche technique sur ce sujet publiée par l'ADEME en 2017, Alain Geldron estime que « sans évolution politique et sociétale majeure au niveau mondial, on peut s'attendre à ce que d'ici 2050 nombre de métaux et de minéraux de base poursuivent une croissance similaire de leur consommation, et donc de la demande en ressources minérales avec *a minima* un doublement d'ici 2050. Celui-ci nécessitera la découverte et l'exploitation de nouveaux gisements importants à un rythme soutenu ».

Exploration et exploitation minière sous influence des lois du marché

Alors que les délais entre les découvertes de gisements potentiels et leur exploitation effective sont longs (de l'ordre de plusieurs décennies en général), les investissements en exploration sont fortement corrélés aux prix des matières premières, comme le montre le graphique ci-contre.



Estimation globale des budgets d'exploration des métaux non ferreux entre 1993 et 2015 (SNL Metals & Mining, 2016).

De plus, les gisements connus, accessibles et facilement exploitables sont naturellement exploités en premiers. Le coût moyen des découvertes augmente donc au fil du temps. L'exploitation de gisements à teneurs plus faibles ou plus difficilement atteignables revient également plus chère.

La notion de pic de ressources ou courbe de Hubbert

Ce modèle défini par Marion King Hubbert, un géophysicien pétrolier américain, représente l'exploitation d'un gisement sous la forme d'une courbe en cloche dite « courbe de Hubbert ».

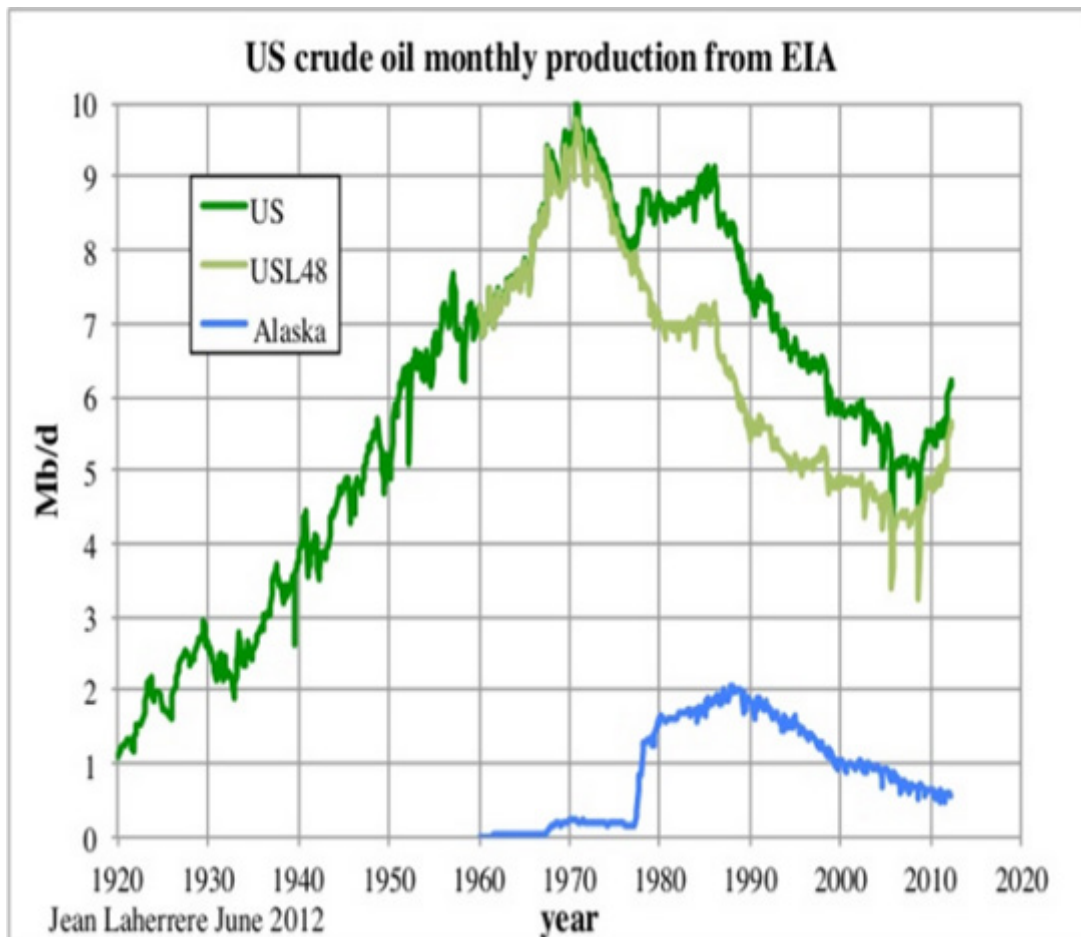
Le pic de production est atteint lorsque la moitié de la ressource a été extraite. Durant la phase ascendante de la courbe, les coûts d'exploitation sont faibles, mais ils augmentent lorsque la courbe s'inverse. En additionnant les données de plusieurs gisements, on peut établir une courbe à une échelle plus large, celle d'un territoire, par exemple. Ce modèle permet de prédire les futurs pics de production à partir des données d'exploration.

Bien que défini et validé sur le cas particulier des ressources pétrolières américaines, ce modèle a été utilisé pour prédire les pics de ressources de nombreux minéraux. Cependant, il fait l'objet de critiques en raison de plusieurs limites liées à la nature des ressources étudiées, aux innovations technologiques et aux règles de marché :

- Contrairement aux hydrocarbures, les métaux extraits ne sont pas détruits, ils peuvent être réutilisés après recyclage.
- Les ressources ultimes n'étant pas connues, de nouvelles découvertes interviennent régulièrement et invalident les prévisions.

- Des phénomènes de substitution peuvent faire baisser la demande sur certaines ressources.
- Une augmentation des prix peut entraîner des investissements en exploration et la découverte de nouvelles réserves.

Le graphique ci-dessous illustre bien certaines limites de ce modèle. Au moment où il avait établi sa prévision d'un pic de la production de pétrole aux États-Unis, dans les années 50, Hubbert disposait d'informations précises sur les réserves connues qui lui ont permis de situer le « *peak oil* » à 1970. La courbe ci-dessous montre que cette prévision était tout à fait exacte. Cependant, les investissements en exploration et les progrès techniques motivés par la hausse des prix du pétrole ont entraîné deux phénomènes qui viennent sensiblement modifier la courbe. D'une part, la découverte et l'exploitation d'importants gisements de pétrole en Alaska qui retardent le pic d'une quinzaine d'années et, d'autre part, le recours à la technique de fracturation hydraulique qui interrompt le déclin de la production et réoriente la courbe à la hausse à partir des années 2010.

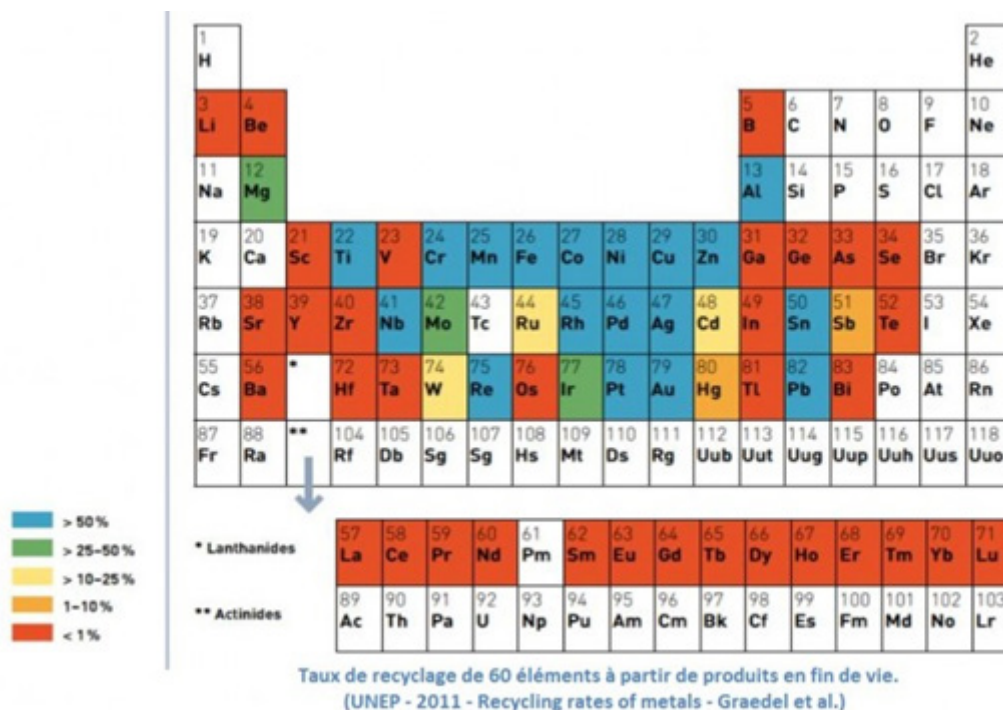


Le recyclage permettra-t-il d'atténuer le phénomène de raréfaction des ressources minérales ?

Face au phénomène de raréfaction des ressources naturelles, le recyclage est souvent une solution avancée. Mais un certain nombre d'éléments atténuent les effets potentiels du recyclage sur la raréfaction des métaux.

Les taux de recyclage des métaux de base peuvent être particulièrement élevés (de l'ordre de 83 % pour l'acier), mais, pour les métaux utilisés en faible quantité dans des alliages, les

taux sont plus faibles. Certains alliages sont réalisés pour des usages bien spécifiques, avec des dosages précis permettant d'obtenir des propriétés recherchées, et il est quasiment impossible d'atteindre les mêmes résultats à partir de matière première recyclée, moins pure.



Une part de la matière est perdue à chaque boucle : par exemple, avec une efficacité théorique de 80 %, il ne reste que 33 % de la matière initiale après cinq cycles de recyclage.

De plus, le recyclage des métaux n'est possible qu'à partir de déchets qui doivent être collectés, les matières consommées dans des usages dispersifs (incorporées dans des produits cosmétiques, par exemple) sont donc perdues. Le traitement de ces déchets doit ensuite être techniquement possible. Certains métaux rares sont utilisés en petites quantités dans de très nombreux petits équipements électroniques (notamment les smartphones), pour lesquels il est difficile d'organiser une filière de recyclage viable.

Le recyclage n'offre pas non plus la même souplesse d'adaptation à la demande qu'un gisement naturel. Car, si les déchets issus directement des processus de fabrication (dits « *new scrap* ») peuvent rapidement être réinjectés dans une filière de recyclage, ce n'est pas le cas des déchets issus des biens de consommation ou des constructions (dits « *old scrap* »), qui ne seront disponibles qu'à l'issue des cycles de vie des produits en question, moins prévisibles et qui peuvent être de plusieurs semaines, mois, années, voire décennies pour les bâtiments.

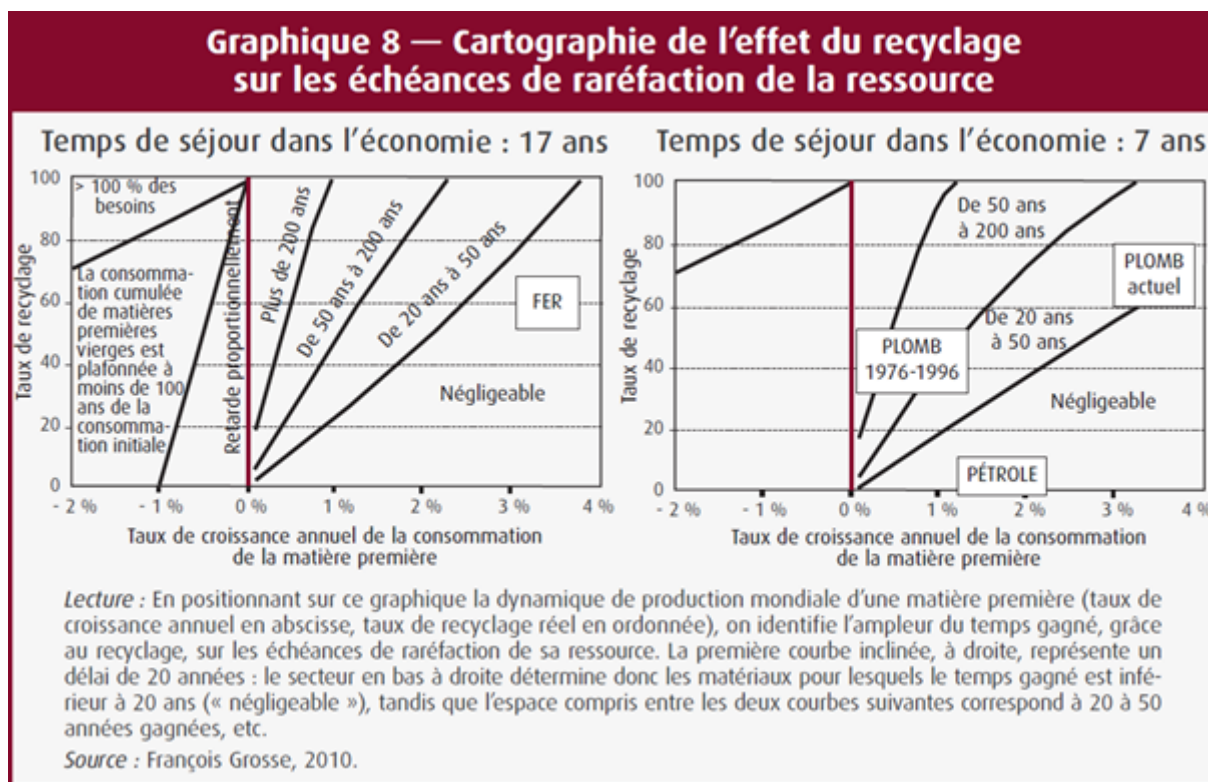
Enfin, et surtout, tous les efforts qui seront faits en matière de recyclage n'auront qu'un effet marginal si les niveaux actuels de croissance de la demande se maintiennent (plus de 3 % par an).

En effet, cette croissance soutenue de la demande impose nécessairement une augmentation des volumes extraits dans une proportion quasiment équivalente. C'est uniquement à des niveaux de croissance faibles (moins de 1 %) que le recyclage peut avoir un réel impact.

Si on prend l'exemple de l'acier, qui est un des métaux qui se recycle le mieux avec un

taux de 83 % (part de l'acier récupéré et introduit dans une filière de recyclage), dans la consommation mondiale d'acier, la part d'acier recyclé à partir de produits en fin de vie ne représente que 16 % de la consommation totale d'acier, 22 % provient des déchets de fabrication et 62 % sont issus de minerais de fer (BIR 2015).

La figure ci-dessous permet de représenter l'effet du taux de croissance de la consommation d'une ressource sur les échéances de raréfaction, en tenant compte du taux de recyclage. Pour le fer, par exemple, si le taux de croissance de la consommation se maintient à 3 % l'effet du recyclage sera négligeable, même avec un taux de recyclage élevé.



Dans un article paru en 2010 dans *Futuribles*, François Grosse explique que, pour enrayer le phénomène de raréfaction des ressources minérales, il faudra deux découplages qui devront s'exercer à l'échelle mondiale :

- Un « découplage fondamental entre développement économique et consommation totale de la matière première, qu'elle soit primaire ou secondaire (vierge ou recyclée) ».
- Puis un « découplage relatif entre consommation totale de matières premières et consommation de matières vierges, par le développement du recyclage ».

« Le découplage fondamental, c'est faire en sorte que la croissance des consommations de matières premières soit sensiblement inférieure à la croissance de la richesse économique, et s'abaisse tendanciellement au-dessous de 1 % par an. Le découplage relatif, c'est faire en sorte de satisfaire les consommations de matières premières en majorité grâce au recyclage, et subsidiairement grâce aux matières vierges (les minerais). Le second découplage, on l'a montré, n'a aucun impact significatif tant qu'il n'est pas précédé du premier. Le premier, on le verra ci-après, n'a qu'un impact décevant s'il n'est pas suivi du second. En englobant généralement sous un concept unique de découplage toute une palette d'outils, certes pertinents, mais juxtaposés sans ordre de priorité, on suggère que les effets de ces outils s'additionneraient, alors qu'ils se

multiplient : c'est-à-dire que l'effet du tout est nul ou presque si l'un des deux découplages est insuffisamment réalisé. »

De nouveaux horizons d'exploration qui n'offrent pas de solution à court terme

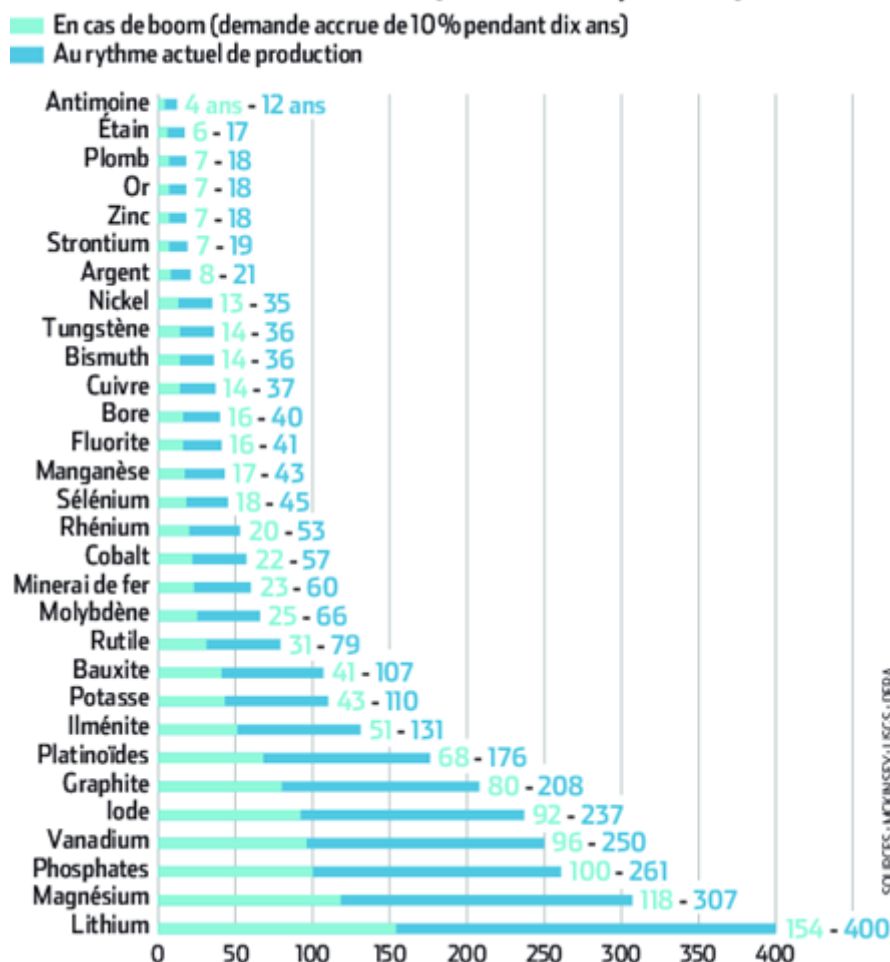
Face à la diminution des réserves et à l'augmentation des prix, des recherches sur de potentielles ressources jusqu'ici inenvisageables sont engagées : zones d'extraction aux conditions climatiques extrêmes, gisements en grande profondeur, fonds marins, astéroïdes.

Cependant, les conditions dans lesquelles l'exploitation de ces ressources pourrait modifier les perspectives d'approvisionnement ne sont pas rassemblées aujourd'hui et ne le seront pas à l'horizon auquel nous nous projetons pour cet exercice. Par exemple, il n'existe qu'un seul cas concret de projet d'exploitation minière off-shore par une société canadienne concernant l'exploitation de sulfures massifs pour en extraire du cuivre, au large de la Papouasie-Nouvelle-Guinée. Il était prévu de démarrer l'exploitation en 2019, mais les ressources totales de ce gisement représentent moins de 1 % de la production annuelle de cuivre actuelle.

Transition énergétique et ressources minérales

L'épuisement annoncé des réserves de pétrole et de gaz conduit à une accentuation des investissements dans les énergies renouvelables. Dans l'état des technologies actuelles, celles-ci induisent des besoins importants de certaines ressources minérales spécifiques.

Durée de vie des réserves rentables (en années d'exploitation)



Les énergies renouvelables posent notamment des problèmes de stockage d'énergie afin de pouvoir gérer des besoins de consommation qui sont irréguliers. En effet, parmi les énergies renouvelables, seule l'hydroélectricité est stockable grâce aux barrages. Le développement de l'éolien et du photoélectrique impose donc des avancées en matière de stockage. De même que le développement du marché des véhicules électriques, qui entraîne une forte augmentation de la production de batteries Li-ion.

Aujourd'hui, les systèmes de batterie les plus performants utilisent notamment des ressources minérales pour lesquelles des questions d'approvisionnement risquent de se poser rapidement : le cobalt et le lithium.

Le lithium est le métal le plus léger et ses caractéristiques physiques en font un excellent conducteur. Cette ressource est relativement abondante, mais peu de gisements présentent un intérêt commercial, la majorité étant trop petits ou pas assez concentrés. Aujourd'hui, 200 Kt de lithium sont extraites par an, quatre pays assurant à eux seuls 95 % de la production (Chili, Australie, Chine et Argentine). La satisfaction des besoins à venir pourrait en nécessiter 1 400 Kt en 2050. Face à ces besoins, les investissements en exploration ont entraîné la découverte de nombreux gisements : entre 2009 et 2011, les réserves ont triplé et les ressources estimées sont passées de 14 à 33 millions de tonnes. Il existe trois sources de lithium : les saumures, les minéraux et l'eau de mer.

Extraction de lithium

Aujourd'hui, l'extraction se fait essentiellement à partir de saumures, car le procédé qui recourt à l'évaporation naturelle est le moins coûteux. Les saumures les plus étendues sont concentrées en Amérique latine (Chili, Bolivie, Argentine...), ainsi qu'en Chine et au Tibet. Le plus grand gisement exploité est le salar d'Atacama au Chili qui fournit à lui seul 40 % de la production mondiale. Il existe un gisement encore plus important en Bolivie, le salar d'Uyuni, mais il n'est pas exploité à ce jour en raison de considérations géographiques (haute altitude, faible potentiel d'évaporation), techniques (forte concentration de magnésium) et environnementales (site unique et concurrence sur l'usage de l'eau avec les fermes voisines).

L'extraction de lithium à partir de minerais ou d'eau salée en reste pour le moment au stade projet en raison de coûts plus élevés. Mais l'exploitation de minerais de pegmatites pourrait se développer, car ils contiennent, outre du lithium, d'autres éléments rares comme le béryllium, le tantale ou le niobium.

Le lithium est un bon exemple des questions que pose cette variable, notamment la nécessité de prendre en compte de nombreux paramètres. Cette ressource ne présente absolument pas de risque d'épuisement, mais les données économiques, géopolitiques et techniques ne mettent pas à l'abri d'un risque de pénurie dans les années à venir.

Le cas du cobalt en est une autre illustration. La fabrication d'une batterie de véhicule électrique peut nécessiter jusqu'à 22 kg de cobalt. Cette industrie en fort développement capte 40 % de la production actuelle, et on en trouve également dans les smartphones. La moitié des réserves connues et de la production actuelle de cobalt est localisée en République Démocratique du Congo (RDC), dont 95 % des recettes d'exportation proviennent de la mine. Le raffinage se fait majoritairement en Chine, mais le prix est fixé à Londres, sur le London Metal Exchange (LME), où il a été multiplié par 3,7 en deux ans. La RDC, dont le poids dans la production mondiale va encore augmenter dans les années à venir, a récemment entrepris une modernisation de son code minier qui pourrait se traduire par une multiplication par cinq de la redevance des compagnies minières étrangères sur le cobalt, qui serait désormais considéré comme un métal « stratégique ».

Selon Amnesty international, 20 % du cobalt congolais est extrait à la main, en partie par de jeunes enfants. Cette situation conduit à une réaction du LME qui demande la mise en place d'une traçabilité éthique : celle-ci pourrait recourir notamment à la technologie de la *blockchain*.

Ici encore, la question qui se pose n'est pas tant celle d'un épuisement de la ressource que celle des conditions économiques, géopolitiques, éthiques et environnementales qui ne permettront peut-être pas un approvisionnement à la hauteur de la demande.

Vulnérabilité des chaînes d'approvisionnement en équipements

La crise actuelle des semi-conducteurs met en évidence la grande vulnérabilité des chaînes d'approvisionnement en équipements de pointe nécessaires au développement de l'IA.

Il n'existe que deux fabricants de semi-conducteurs (pour les plus performants) : TSMC à Taiwan, et Samsung en Corée du Sud. Ces deux acteurs sont les fournisseurs uniques de très nombreuses industries : automobiles, smartphones, ordinateurs, consoles de jeux vidéo... Ils sont eux-mêmes dépendants de fournisseurs souvent éloignés de leurs sites de production : par exemple, certains produits chimiques indispensables aux bains proviennent du Japon. Et sont fortement consommateurs de ressources naturelles, comme le silicium et l'eau. TSMC consomme chaque jour 156 millions de litres d'eau ! Ils sont surtout tous les deux dépendants d'un fournisseur unique de machines de fabrication de puces, ASML, une entreprise néerlandaise dont la capacité de production est limitée (environ 50 machines par an).

La crise pandémique qui a fortement perturbé les chaînes logistiques en amont et en aval de ces goulots a mis en évidence la très grande vulnérabilité de cette chaîne de valeur. Elle aura également mis en relief l'importance stratégique de certains composants et provoqué les réactions des puissances économiques qui en ont la capacité. Le Sénat américain (Républicains et Démocrates ensemble) a ainsi voté le *Endless Frontier Act*², un plan qui prévoit 170 milliards de dollars d'investissements dans les technologies pour réduire la dépendance des États-Unis et de rattraper le retard sur la Chine. Ce plan prévoit 52 milliards sur 5 ans pour fabriquer localement les puces et semi-conducteurs dont l'industrie américaine a besoin, et d'injecter 120 milliards dans la National Science Foundation. En la matière, l'argent ne semble pas être un problème : rien que pour l'année 2021, ASML prévoit 28 milliards de dollars d'investissement pour développer ses capacités de production, dont 12 milliards seront investis sur le sol américain.

La question du Cloud

Le Cloud se développant à grande vitesse joue un double rôle en matière de développement de l'IA, et notamment des modèles de *deep learning*. Les fournisseurs offrent d'une part une solution de stockage des quantités de données nécessaires à l'entraînement des modèles, ce qui pose la question de la consommation d'énergie des *data centers* que nous développons ci-dessous. Mais les *cloud providers* fournissent également de la capacité de calcul à leurs clients, ce qui tend à rendre accessibles au plus grand nombre (y compris des petites entreprises) des capacités computationnelles nécessaires aux solutions de *deep learning*.

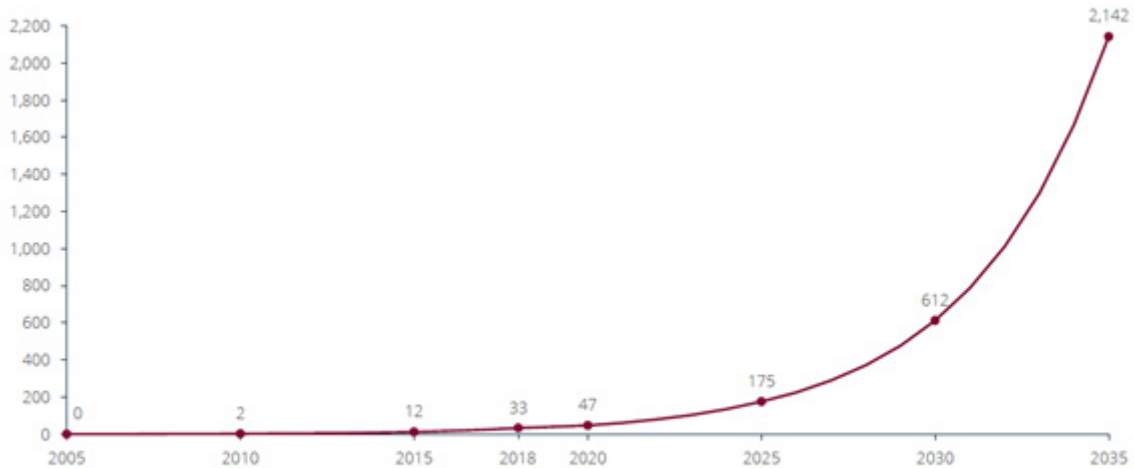
Ici se pose donc la question des *data centers* et de leur consommation d'énergie.

2. <https://www.france24.com/fr/am%C3%A9riques/20210609-le-s%C3%A9nat-am%C3%A9ricain-approuve-un-plan-d-investissement-historique-pour-contrer-la-chine>

L'explosion des données

Il est prévu, dans les quinze prochaines années, une explosion du nombre de données à stocker. Ce phénomène est lié entre autres à la multiplication des objets connectés et à l'accroissement des interactions entre machines sans intervention humaine.

Worldwide amount of data created per year in zettabytes¹



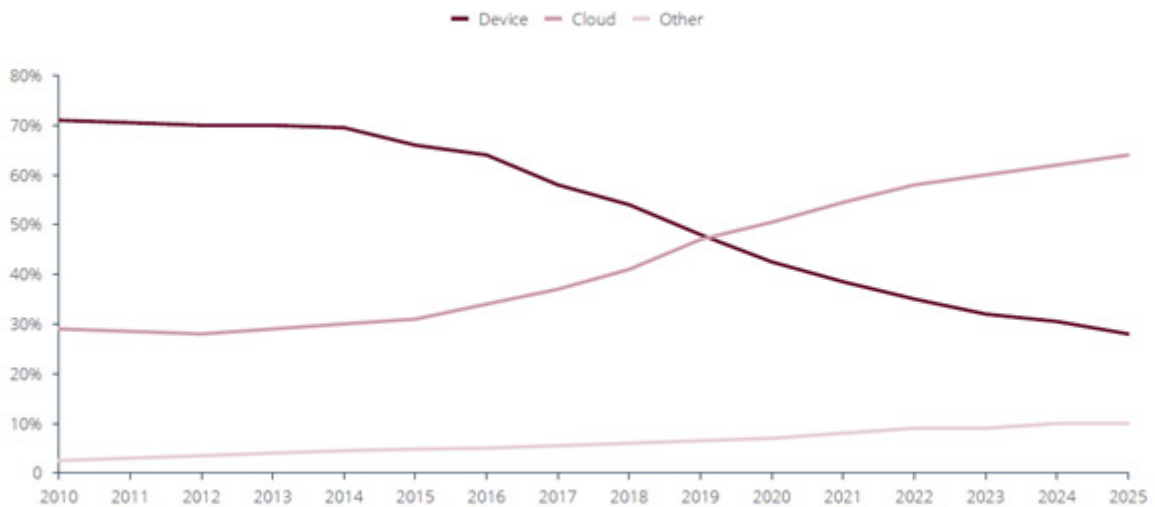
6 ¹: One zettabyte equals to 1 billion terabytes
Source: IDC, Kleiner Perkins. [Statista Digital Market Outlook](#)

statista

[Un zettaoctet correspond à mille milliards de gigaoctets]

En dix ans, les données ont massivement transité, depuis les équipements vers le Cloud, qui présente de nombreux avantages.

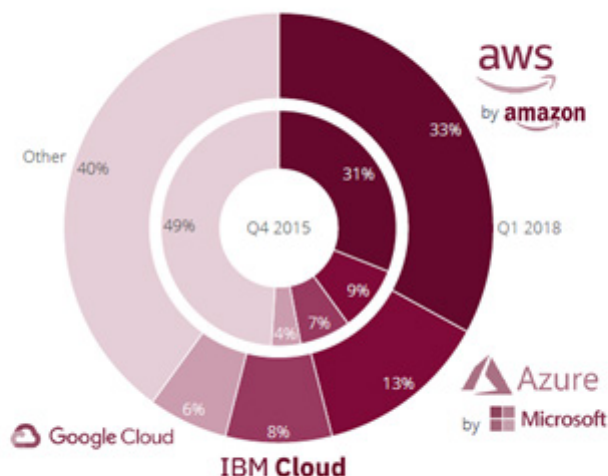
Share of data storage solutions



9 Source: IDC's Data Age 2025 study

statista

60 % des revenus générés par le Cloud sont captés par quatre opérateurs, tous américains, dont les parts de marchés augmentent. La tendance est donc à la concentration, mais aussi à l'interconnexion des *data centers*, entre ceux situés au plus près des utilisateurs (en milieu urbain) et les *data centers hyperscales* en zone rurale.

Worldwide market share of Cloud services¹ revenue in Q4 2015 and Q1 2018

1: Includes Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Cloud Services for end-users
Source: Synergy Research Group

statista

L'ADEME a consacré un rapport à l'impact spatial et énergétique des *data centers*³. Il est estimé que les *data centers* consomment actuellement 2 % de l'électricité mondiale (l'équivalent de 60 tranches nucléaires). Selon le Shift Project, ce chiffre devrait monter à 5 % en 2025 : la question de la soutenabilité de ce rythme de croissance de la consommation électrique se pose donc.

Dans son rapport « Transition(s) 2050 », l'ADEME propose plusieurs scénarios intégrant des projections sur la consommation énergétique des *data centers*⁴.

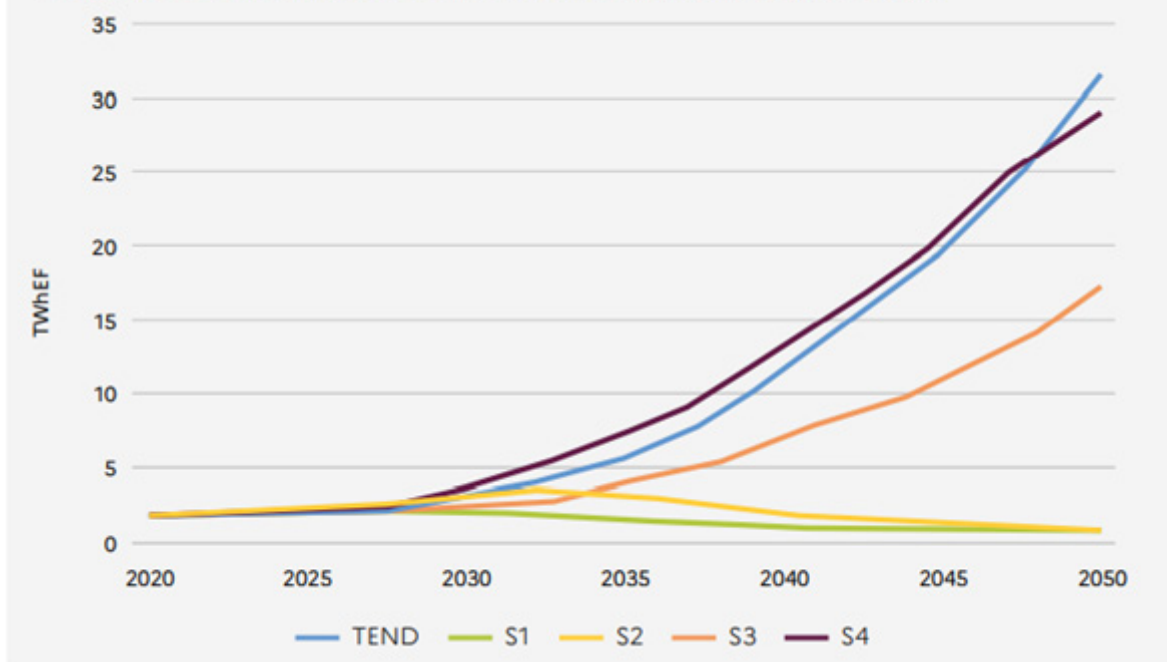
Data centers hébergeurs	<p>Très forte augmentation de la consommation énergétique. Le volume de données hébergées croît fortement, particulièrement dans les petits data centers (sous l'influence de la 5G et de l'Internet des objets). L'efficacité énergétique s'accroît, mais sans compenser la hausse du volume de données.</p> <ul style="list-style-type: none"> • Volume de données. Le volume de données des petits data centers (5G, Internet des objets...) est de 4,21 exabytes (EB) en 2020. Il croît de 50% par an jusqu'en 2025, puis de 40% de 2026 à 2040. Cette croissance diminue ensuite par palier pour atteindre 30% à partir de 2043 jusqu'en 2050. Le volume de données des gros data centers (hyperscale, gros hébergeurs...) est de 107,65 EB en 2020. Il croît de 15% par an jusqu'en 2050. • Type de data centers. Développement du edge computing, croissance plus importante des petits data centers pour l'Internet des objets et la 5G. Les gros data centers continuent leur croissance avec des besoins de stockage de plus en plus importants et l'externalisation des applications d'entreprise. • Efficacité énergétique de la partie informatique. L'efficacité de traitement des données des petits data centers et des gros augmente de 13% par an jusqu'en 2050, à partir de respectivement 27,38 GWh/EB et 8,66 GWh/EB. • Power Usage Effectiveness (PUE)[®]. Le PUE des petits data centers est de 2,10 en 2020 et diminue de 2% tous les ans (= 1,15). Le PUE des gros data centers est de 1,80 en 2020 et diminue de 2% par an tous les ans pour atteindre le palier de 1,02 à partir de 2048.
--------------------------------	---

Hypothèse tendancielle de l'ADEME concernant les *data centers*.

3. C. Diguët et F. Lopez, « L'impact spatial et énergétique des data centers sur les territoires », rapport ADEME, 2019.

4. [transitions2050-rapport-compresse.pdf](#), p.114 et 153

Graphique 10 Consommation des data centers hébergeurs dans tous les scénarios



Projections des consommations énergétiques des data centers selon les scénarios Transition de l'ADEME. [Tend = tendanciel ; S1 = le scénario le plus frugal ; S4 = le plus technologique].

L'efficacité énergétique des *data centers* (PUE, pour *Power Usage Effectiveness*) a beaucoup progressé ces 20 dernières années, grâce à des évolutions structurelles.

En terme d'émissions de gaz à effet de serre (GES), il est important de tenir compte de la localisation des *data centers*. Car, selon la composition du mix énergétique en vigueur, les rejets ne seront pas du tout les mêmes. Ainsi, les conséquences climatiques de l'implantation d'un *data center* ne sont pas les mêmes selon que celui-ci est situé en France, dont l'électricité est largement décarbonnée, ou en Chine, qui recourt encore massivement à des énergies fossiles.

Hypothèses

Hypothèse 1. Pénuries et arbitrages

Les difficultés d'approvisionnement en composants performants, l'augmentation de leurs coûts, et une meilleure mise en avant de l'impact environnemental de modèles très gourmands, comme le *deep learning*, dissuadent certains usages. De nombreuses entreprises renoncent à des modèles qui les exposent trop à des risques de pénuries, de pannes, et nuisent à leur image. L'usage de l'IA est cantonné à certains domaines stratégiques.

Hypothèse 2. La rareté des ressources impose une IA sobre

D'importants progrès techniques sont réalisés pour permettre une croissance des usages de l'IA. Cela passe par le recours à des composants moins gourmands en énergie et nécessitant moins de métaux rares, d'importants progrès en matière de récupération et de recyclage des métaux rares, une optimisation du partage des capacités de calcul, le développement d'un Cloud souverain faiblement émissif en GES, le renoncement à des niveaux de fiabilité très élevés pour certains usages.

Hypothèse 3. L'IA quoi qu'il en coûte

Dans le cadre d'une course à l'IA mondiale, la France ne veut renoncer à rien. Tous les arbitrages sont faits en faveur du développement technologique, soutenu par un argumentaire économique et présentant l'IA comme une condition de la réussite de la transition énergétique. Des efforts sont faits pour réduire la consommation énergétique des dispositifs et s'approvisionner en énergies renouvelables, mais aucune concession n'est faite sur la puissance des dispositifs.

Disponibilité des ressources humaines nécessaires au déploiement de l'IA (hypothèses seules)

M. Malenfer

Le développement de l'IA nécessite également des ressources humaines, avec des compétences particulières. Le besoin concerne aussi bien les capacités de développement des systèmes en amont que les aptitudes des entreprises à intégrer ces dispositifs dans leurs process en aval.

Ces hypothèses viennent compléter celles de la Fiche 2 sur la disponibilité des ressources naturelles. Les éléments qui ont permis de les formuler ne sont pas développés ici.

Hypothèse 1. Les géants monopolisent la main-d'œuvre

Les compétences en IA sont rares et chères. Les géants de la tech dont le business modèle repose en grande partie sur l'IA, et qui disposent des moyens nécessaires, captent les compétences, y compris en Europe. L'IA se diffuse via les solutions qu'ils développent et mettent sur le marché, et dont ils assurent également la maintenance.

Hypothèse 2. Externalisation et plateformes des compétences en IA

Les compétences en IA sont rares et chères en France. Les entreprises ne parviennent pas à constituer ni fidéliser des équipes internes. Elles font appel à des prestataires externes, qui sont soit des SSII, soit des freelances qu'elles font travailler via des plateformes. Dans les deux cas, une part importante de la main-d'œuvre est étrangère et travaille à distance (notamment depuis l'Inde). Cette organisation pose de nombreux problèmes d'adaptation des solutions aux caractéristiques des entreprises et de leurs marchés.

Hypothèse 3. Création d'un vivier de compétences

Des investissements publics et privés importants permettent le développement de filières

de formation initiale et continue en IA. De grandes entreprises jouent un rôle actif en créant des établissements et des filières d'apprentissage efficaces. Une politique est mise en place pour accueillir et former une main-d'œuvre étrangère dans ces domaines (immigration choisie). Un vivier de main-d'œuvre se développe en France, limitant l'augmentation des salaires : cela permet des développements de solutions IA plus distribués, y compris dans de petites entreprises. Par conséquent, ces solutions sont moins standardisées et plus proches des besoins spécifiques des entreprises.

Acteurs et dynamique de la diffusion de l'IA, processus de normalisation

N. Fatès, M. Sarrey, J.-F. Soupizet

Présentation

Cette fiche présente une cartographie des acteurs de l'offre dans le domaine de l'intelligence artificielle (IA). Elle analyse les dynamiques de la diffusion de cette offre et porte sur :

- Les acteurs étatiques qui interviennent de multiples manières : en amont pour financer la recherche fondamentale et le développement (programmes européens et nationaux), les développements spécifiques par des actions pilotes (démonstration à échelle réelle) ; en aval par la commande publique, et notamment militaire, et enfin en agissant auprès des autres acteurs par des politiques incitatives, y compris dans le domaine des données ;
- Les géants de l'Internet offrant des services d'IA, en particulier Microsoft, Amazon et Facebook (60 % du marché mondial à eux trois) ;
- Les autres acteurs proposant des services d'IA et les acteurs intermédiaires comme les sociétés de services.

Par souci de concision, nous avons laissé de côté la question des utilisateurs directs des techniques d'intelligence artificielle (professionnels et grand public), car ces utilisateurs sont ceux à qui les États et les entreprises destinent leurs services. La deuxième partie du document présentera quatre exemples de secteurs d'utilisation de ces techniques.

Nous aborderons ensuite les dynamiques de diffusion sous l'angle des modèles technique et économique de l'offre (par exemple, quand elle est intégrée à une prestation « cloud »), ainsi que les questions de normes et de certification industrielle, dont l'élaboration relève de processus internationaux (ISO, par exemple).

Définitions et contexte

Le fait que nous ne disposons pas de définition claire du terme « intelligence artificielle »

est un point méthodologique important. Nous sommes donc amenés à mettre sous cette étiquette tout ce que les auteurs ou les institutions désignent par ce terme. Cela pose un problème de taille sur divers points, notamment l'évaluation des budgets alloués. On peut supposer que ces budgets sont artificiellement gonflés par des opérations de « labellisation IA », lesquelles enrôlent sous la bannière de l'IA des projets de recherche ou de développement qui étaient en cours ou qui auraient été approuvés même sans une politique spécifique.

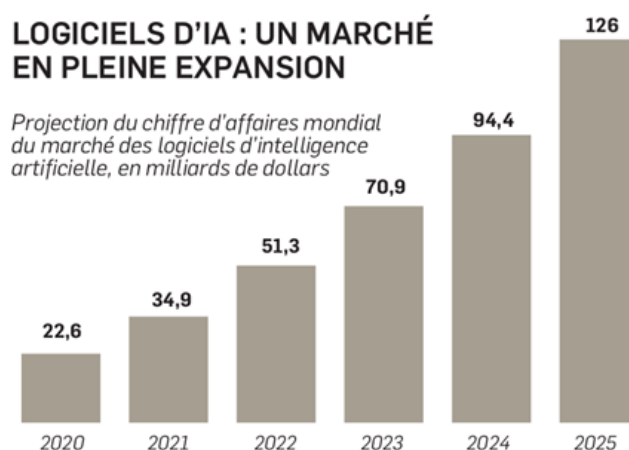
Rôle des États dans le financement du développement de l'intelligence artificielle

Nous commençons par présenter le rôle joué par les États dans le déploiement des techniques d'intelligence artificielle.

Situation à l'échelle mondiale

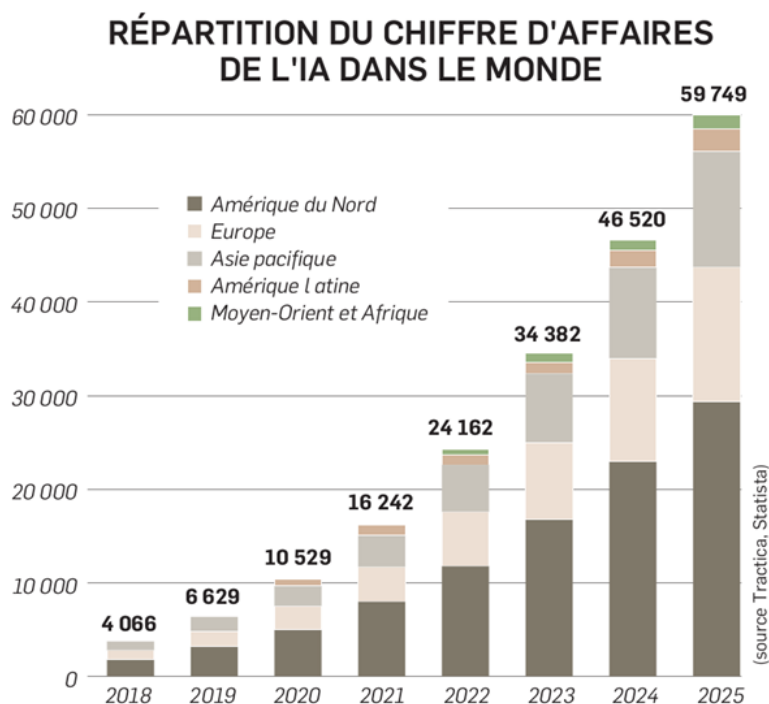
Il est difficile d'estimer les investissements étatiques à l'échelle mondiale. Thomas Campbell estime à 152 milliards de dollars l'investissement des États dans le domaine de l'intelligence artificielle¹, mais ce chiffre agrège l'ensemble des projets dont l'auteur a connaissance et rassemble donc des intervalles de temps potentiellement disparates (mais rarement au-delà de cinq ans). Cet auteur indique qu'en 2019, il existe 20 États ayant des structures d'investissement spécifiques.

Le dossier de la revue *SAY*² donne un chiffre de 55 milliards de dollars entre 2015 et 2020 pour les « investissements annuels mondiaux des grandes organisations dans l'IA ». Le marché mondial de l'IA est estimé à 327,5 milliards de dollars. Dans tous les cas, les courbes montrent une croissance du domaine de plus en plus rapide.



1. Thomas Campbell, « Artificial intelligence: an overview of state initiatives », 2019.

2. « Le marché de l'IA et son périmètre », revue *SAY*, n° 5, pages 144 à 145, 2021.



Il est manifeste, qu'à l'échelle planétaire, trois grands acteurs se dégagent : les États-Unis, la Chine et l'Europe. Un rapport comparant les effectifs de ces trois ensembles dans le domaine de l'intelligence artificielle constate, pour 2017 :

- **Nombre de chercheurs en IA** : Chine, 18 232 ; Europe, 43 064 ; États-Unis, 28 536 ;
- **Nombre d'articles** : Chine, 15 199 ; Europe, 14 776 ; États-Unis, 10 287.

Selon la société de conseil en affaires et en stratégie McKinsey, l'IA pourrait générer d'ici 2030 un produit économique supplémentaire d'environ 13 milliards de dollars US, soit une hausse du PIB mondial d'à peu près 1,2 % par an³.

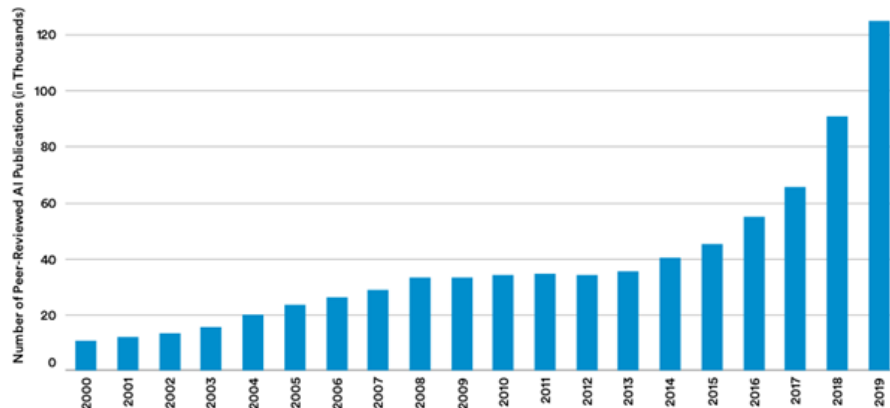
Au niveau académique, le nombre de publications liées à l'intelligence artificielle augmente également de manière soutenue : ce nombre représentait 3,8 % de toutes les publications en 2019, contre 1,3 % en 2011⁴.

3. Fabien Merz, « L'Europe et la course à l'intelligence artificielle », Politique de sécurité, analyses du CSS 247, 2019.

4. « Le marché de l'IA et son périmètre », revue SAY, n° 5, pages 144 à 145, 2021.

NUMBER of PEER-REVIEWED AI PUBLICATIONS, 2000-19

Source: Elsevier/Scopus, 2020 | Chart: 2021 AI Index Report



Initiatives de la France

Dès mars 2018, le président de la République Emmanuel Macron a annoncé son ambition de faire de la France un leader mondial de l'intelligence artificielle⁵. Se fondant sur les travaux du rapport du député Cédric Villani, une première phase de développement – soutenue par des investissements de 1,5 milliard d'euros – a notamment conduit à la création et au développement d'un réseau d'instituts interdisciplinaires d'intelligence artificielle (3IA), le soutien à des chaires d'excellence en IA, le financement de programmes doctoraux, et l'investissement dans les capacités de calcul de la recherche publique. En novembre 2021, le gouvernement a présenté une nouvelle phase de la stratégie française pour l'IA, qui mobilisera 2 milliards d'euros de cofinancements publics-privés. La contribution publique à la stratégie est principalement financée par le Programme d'investissements d'avenir (577 M€) et France 2030 (700 M€)⁶.

Sur le plan industriel, mentionnons entre autres le fait qu'un partenariat a été noué entre Google et Thalès pour l'hébergement en ligne de logiciels à l'usage des professionnels et des particuliers. La société S3ns a été créée en 2021 pour bénéficier de la puissance de la plateforme Google, tout en respectant les normes pour un *cloud* de confiance. Un schéma du même type a déjà été mis en place entre Cap Gemini, Orange et Microsoft.

Au niveau militaire, le rapport du ministère des Armées de 2019 déclare : « Afin de préparer les futures applications de l'IA à horizon 10 ans, le ministère des Armées investira massivement dans les études et la recherche. Durant la période de la présente Loi de programmation militaire 2019-2025, le ministère investira près de 430 M€ d'études en amont au profit de l'intelligence artificielle⁷ ». Ce rapport marque une prise de conscience de la nécessité pour l'armée de mener une réflexion spécifique par rapport à l'intelligence artificielle (cyber-guerre, mise en place de comités d'éthique, risque de détournement des technologies), ce qui marque une rupture avec une approche où l'IA est considérée comme une « technique de plus » dans un panel de possibilités. Les chiffres à retenir sont 100 millions d'euros par an en moyenne sur la période LPM (Loi de programmation militaire) 2019-2025, pour les études et la recherche sur l'IA⁸.

5. <https://www.intelligence-artificielle.gouv.fr/fr>

6. <https://www.gouvernement.fr/investissements-d-avenir-nouvelle-phase-de-la-strategie-nationale-d-intelligence-artificielle-le>

7. « L'intelligence artificielle au service de la défense », rapport de la Task Force IA du ministère des Armées, septembre 2019.

8. <https://www.intelligence-artificielle.gouv.fr/fr/secteurs-prioritaires/l-intelligence-artificielle-et-monde-de-la-defense>

Atos et Thales ont annoncé la création d'Athea, une société commune destinée à développer une plateforme associant traitement de données massives et intelligence artificielle pour les secteurs de la défense, du renseignement et de la sécurité intérieure et qui s'adresse tant aux acteurs publics que privés. Athea bénéficie de l'expérience acquise par les deux entreprises dans la phase de démonstration du programme « ARTEMIS », la plateforme Big Data du ministère des Armées. Les deux sociétés sont titulaires du marché d'optimisation et de préparation de l'industrialisation du programme, notifié par la Direction générale de l'armement (DGA) le 30 avril 2021. La société commune doit servir le marché français, puis, à terme, les marchés européens⁹.

Initiatives de l'Union européenne

L'Union européenne (UE) a lancé un programme « conçu pour combler l'écart entre la recherche sur les technologies numériques et leur déploiement sur le marché ». Doté de 7,5 milliards d'euros, ce programme numérique fait partie du prochain budget à long terme de l'UE sur la période 2021-2027. Il financera des projets dans cinq domaines : « calcul à haute performance, intelligence artificielle, cybersécurité, compétences numériques avancées, et large utilisation des technologies numériques dans l'ensemble de l'économie et de la société », détaille le communiqué de l'UE¹⁰. « [Le programme] renforcera et soutiendra les tests et les expérimentations en matière d'intelligence artificielle dans des domaines comme la santé et la mobilité, et encouragera les États membres à coopérer, avec pour ambition de mettre en place un espace européen des données et d'accroître l'utilisation de l'intelligence artificielle dans les entreprises et les administrations publiques (2 milliards d'euros). »

Sur la période 2018-2020, le programme de recherche et d'innovation Horizon 2020 consacrait 1,5 milliard d'euros à l'intelligence artificielle. Par ailleurs, une stratégie pour les données a été élaborée dans l'Union européenne¹¹. L'objectif porte sur la réglementation et le développement d'un « marché unique européen des données ». Elle comporte plusieurs volets, dont un appui à l'initiative privée GAIA-X¹², qui prévoit le développement d'une infrastructure fédérée pour l'hébergement des données et les services connexes (informatique en nuage, ou *cloud*) entre les acteurs européens. À ce jour, l'initiative revendique 1800 contributeurs et 350 membres, et les travaux pour élaborer des normes communes ont été engagés dans une dizaine de secteurs.

Initiatives des États-Unis

En 2021, le budget fédéral consacré à l'intelligence artificielle aurait été de 1,5 milliard de dollars, soit une augmentation de 974 millions de dollars par rapport à 2020. Le plan Artificial Intelligence Initiative Act (AI-IA) prévoyait 2,2 milliards de dollars d'investissement au niveau fédéral sur cinq ans. Cependant, aux États-Unis, c'est le secteur privé qui pèse le plus dans le développement de l'intelligence artificielle.

Le secteur étatique possède plusieurs agences, dont chacune développe des projets d'intelligence artificielle de manière autonome. Par exemple, le DARPA a annoncé un plan

9. <https://www.thalesgroup.com/fr/group/journaliste/press-release/thales-et-atos-creent-le-champion-europeen-du-big-data-et>

10. <https://www.europarl.europa.eu/news/fr/press-room/20201211IPR93656/programme-pour-une-europe-numerique-accord-avec-le-conseil>

11. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr

12. <https://gaia-x.eu/>

de 2 milliards de dollars sur plusieurs années pour développer des techniques d'intelligence artificielle dites de « troisième vague », c'est-à-dire des techniques de construction de systèmes robustes capables de faire preuve de « bon sens »¹³.

Constatant que les forces armées avaient cinq à dix ans de retard sur le secteur privé américain dans les nouvelles technologies, le Pentagone a commencé dès 2017 une série de projets reposant sur l'IA. Parmi ceux-ci figure le projet MAVEN, qui utilise l'apprentissage automatique pour exploiter les données générées par les plateformes de renseignement américaines. En quinze ans, le Département de la défense (DoD) a investi plusieurs milliards de dollars pour équiper ses forces de moyens de surveillance et de reconnaissance¹⁴.

Initiatives de la Chine

Un article d'Aifang Ma¹⁵ résume les ambitions chinoises en la matière :

« En juillet 2017, le Conseil des affaires d'État a publié son Plan de développement de la prochaine génération d'intelligence artificielle pour la période de 2016 à 2030. Dans ce document, la Chine se fixe l'objectif de devenir le leader mondial dans ce domaine d'ici à 2030, en termes de théorie fondamentale, de technologies et d'applications.

Elle envisage de réaliser son but en trois étapes :

- Pour 2020, le pays doit avoir réalisé une découverte majeure dans les théories fondamentales et les technologies clés liées au Big Data, à l'automatisation et à l'application de l'IA aux médias de masse. Elle aura ainsi participé au développement des normes et des standards internationaux de cette technologie. Le secteur pèsera alors en Chine environ 23 milliards de dollars, et les secteurs liés 156 milliards de dollars.
- Pour 2025, l'IA devra être la dynamique principale de croissance économique et de mise à niveau de l'économie. Elle sera déjà utilisée sur des sujets clés comme la santé, les villes intelligentes, l'agriculture intelligente et la défense. Un cadre juridique et éthique ainsi qu'un système réglementaire de l'IA prendront forme. Le secteur vaudra alors 63 milliards de dollars, et les secteurs liés 781 milliards de dollars.
- Pour 2030, la Chine devra être leader mondial dans le secteur de l'IA. Elle aura fait une percée significative dans la recherche sur le cerveau. Elle sera devenue le centre mondial d'innovation de l'IA et une référence sur la réglementation. Le cadre juridique et éthique de l'IA aura atteint sa maturité. L'industrie chinoise de ce secteur atteindra 156 milliards de dollars, et 1 562 milliards de dollars pour les secteurs liés. »

La commande publique en France

L'État contribue à impulser une dynamique de développement, notamment par les appels d'offre, la mise en place de normes, de procédures d'évaluation (*benchmarking*), et par le soutien à des initiatives privées.

En 2021, le Laboratoire de métrologie a lancé « une nouvelle certification volontaire

13. « AI next campaign », <https://www.darpa.mil/work-with-us/ai-next-campaign>

14. Arnaud Gary, « Intelligence artificielle et armées françaises : une technologie du présent à mettre en œuvre immédiatement », revue *Défense Nationale*, HS4, p. 200-213, DOI : 10.3917/rdna.hs06.0200, 2021. <https://www.cairn.info/revue-defense-nationale-2021-HS4-page-200.htm>

15. Aifang Ma, « L'intelligence artificielle en Chine : un état des lieux », hal-02186714, 2018.

portant sur tous les systèmes d'IA, afin de favoriser la confiance accordée en ceux-ci en apportant des garanties délivrées par un tiers de confiance »¹⁶.

Exemple 1 : Services publics

L'État a lancé un robot conversationnel (*chatbot*) pour répondre aux questions des usagers des administrations¹⁷. En 2018, un premier appel à manifestation d'intérêt (AMI) a été lancé pour expérimenter l'intelligence artificielle dans les services publics¹⁸ et, en 2019, un second appel a été lancé¹⁹. Les projets retenus fin 2019 touchaient à la gestion de la relation usagers, au ciblage de contrôles et à l'analyse et la synthèse d'informations²⁰. Récemment, la presse a relayé les plaintes de personnes touchées par les baisses des aides sociales liées à l'emploi de techniques d'IA²¹.

Au sein du département Etalab de la Direction interministérielle du numérique (DINUM), un laboratoire pour l'intelligence artificielle (Lab IA) a été créé afin d'« accompagner les administrations dans le déploiement de leurs projets IA et renforcer leurs capacités en *data science* »²².

L'intelligence artificielle est utilisée par les administrations lors des contrôles fiscaux, principalement dans la sélection des dossiers et la détection des cas de fraude. Les administrations fiscales cherchent à augmenter les chances de succès du redressement. La France a ainsi annoncé, dans le rapport de la Direction générale des Finances publiques (DGFiP) 2019, avoir programmé en 2019 22 % des contrôles fiscaux grâce à la fouille de données, ce qui a représenté plus de 700 millions d'euros collectés. La DGFiP a annoncé vouloir porter la part de contrôles orientés par l'intelligence artificielle à 35 % en 2020, et à 50 % en 2022²³.

Exemple 2 : Justice

L'État encourage également le développement de la « justice prédictive ». Par exemple, un décret²⁴ datant de mars 2020 crée un traitement automatisé de données à caractère personnel, DataJust, qui vise à développer, pour une durée de deux ans, un dispositif algorithmique permettant de recenser, par type de préjudice :

- Les montants demandés et offerts par les parties à un litige ;
- Les montants alloués aux victimes en indemnisation de leur préjudice corporel dans les décisions de justice rendues en appel par les juridictions administratives et les formations civiles des juridictions judiciaires.

16. <https://www.lne.fr/fr/communiqués-de-presse/intelligence-artificielle-nouvelle-certification>

17. « MIA, le « chatbot » qui vous guide dans vos démarches administratives », communiqué du 14/09/2021, <https://www.gouvernement.fr/mia-le-chatbot-qui-vous-guide-dans-vos-demarches-administratives>

18. <https://www.modernisation.gouv.fr/presse/lancement-dun-appel-manifestation-dinteret-ami-pour-experimenter-lintelligence-artificielle>

19. <https://www.modernisation.gouv.fr/actualites/ministeres-et-operateurs-publics-2eme-appel-manifestation-dinteret-pour-tester-le>

20. <https://artimon.fr/perspectives/lintelligence-artificielle-au-service-de-ladministration-publique-quelles-techniques-et-quelles-specificites/> et <https://www.numerique.gouv.fr/espace-presse/intelligence-artificielle-15-nouveaux-projets-vont-etre-experimentes-dans-les-administrations/>

21. Olivier Tesquet, « Comment l'intelligence artificielle cible les plus précaires », *Télérama*, 14/10/2021, <https://www.telerama.fr/debats-reportages/comment-lintelligence-artificielle-cible-les-plus-precaires-6984905.php>

22. <https://www.etalab.gouv.fr/lab-ia>

23. Intelligence artificielle et fiscalité, <https://www.affiches-parisiennes.com/intelligence-artificielle-et-fiscalite-12259.html>

24. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000041763205>

La loi pour une République numérique du 7 octobre 2016 a autorisé la publication en *open data* des décisions de justice anonymisées. La loi du 23 mars 2019 de programmation et de réforme pour la justice, et le décret du 29 juin 2020 ont apporté des précisions sur les conditions de mise à disposition du public des décisions de justice, le renforcement de l'anonymat et le calendrier de mise en place des différents stades de diffusion de l'information²⁵.

Exemple 3 : Sécurité et sûreté du territoire

L'expérience prédictive de la gendarmerie est initiée par un petit noyau de gendarmes, scientifiques de formation, et férus d'intelligence artificielle. En 2015, une première démarche prédictive est officiellement lancée avec la mission Etalab, placée au sein de la Direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC).

Le déploiement des caméras de surveillance est conséquent, tant dans le domaine privé (plus d'un million) que dans le domaine public (150 000). Certaines de ces caméras peuvent accueillir des applications de reconnaissance faciale ou de « détection d'attitude ».

En Chine, on parle de 200 millions de caméras en circuit fermé (CCTV).

En France, nombre d'entreprises (par exemple, Sécuritas) proposent des techniques d'intelligence artificielle pour aider à la détection d'événements inhabituels dans un contexte de surveillance par caméra.

Exemple 4 : Santé

Le rapport de l'INSERM sur l'IA et la santé²⁷ recense les grandes utilisations de l'IA et en retrace le contexte historique. Dans les années 1980, l'approche dite « symbolique » a permis le développement d'outils capables de reproduire les mécanismes cognitifs d'un expert. C'est pourquoi on les a baptisés « systèmes experts ». Les plus célèbres, Mycin (identification d'infections bactériennes) ou Sphinx (détection d'ictères), s'appuient sur l'ensemble des connaissances médicales dans un domaine donné. Contrairement à l'approche symbolique, l'approche dite « numérique » raisonne sur les données en établissant des corrélations. Le système cherche des régularités dans les données disponibles pour extraire des connaissances, sans modèle préétabli. Cette méthode, née avec le connexionnisme et les réseaux de neurones artificiels dans les années 1980, se développe aujourd'hui grâce à l'augmentation de la puissance des ordinateurs et à l'accumulation des gigantesques quantités de données.

C'est dans ce contexte qu'en décembre 2019 le Health Data Hub (HDH) a été créé comme plateforme pour héberger les données de santé françaises²⁸. Il s'agissait de regrouper la quasi-totalité des données médicales détenues jusqu'ici par les différents acteurs, et en particulier les données des patients hospitalisés, celles de la Caisse nationale de l'assurance maladie (l'une des plus grandes bases du monde) et celles d'autres bases

25. <https://www.vie-publique.fr/eclairage/277098-lintelligence-artificielle-ia-dans-les-decisions-de-justice14>

26. <https://technologie.securitas.fr/decryptage/videosurveillance-intelligence-artificielle-service-surete>

27. INSERM, « Intelligence artificielle et santé - Des algorithmes au service de la médecine », <https://www.inserm.fr/dossier/intelligence-artificielle-et-sante/>

28. Article 41, avec entrée en vigueur le 31 décembre 2019 de la loi relative à la transformation du système de santé, <https://www.irdes.fr/documentation/syntheses/projet-de-loi-relatif-a-l-organisation-et-a-la-transformation-du-systeme-de-sante.pdf>

(dont Oscour, mise en place par les urgences). L'objectif étant d'offrir aux chercheurs une base riche et sécurisée afin de leur permettre d'exploiter des données de santé élargies aux données cliniques, et plus seulement médico-administratives. En juin 2020, le HDH a confié l'hébergement des données de la plateforme à Microsoft. Ce choix a suscité des critiques portant principalement sur la protection des données transférées à Microsoft, invoquant les risques de transfert de données vers les États-Unis²⁹. Le Health Data Hub s'est engagé à faire passer l'hébergement de Microsoft à un acteur de droit européen, promesse repoussée récemment à 2025. De nombreux projets liés à l'exploitation des données du HDH font explicitement appel au domaine de l'IA : TARPON (traumatisme), AIDA (maladie du foie), SEDAAR (ophtalmologie) ou encore APRIORICS (cancer du sein).

Au niveau multilatéral, l'Organisation mondiale de la Santé a publié un rapport visant à établir une référence mondiale sur l'éthique à respecter, et les modes de gouvernance à mettre en place dans le cadre des applications de l'IA dans le domaine de la santé³⁰.

Le rôle des géants du Net

Les acteurs nord-américains

Google est créé en 1998 par Larry Page et Sergey Brin. À l'origine, Google est un moteur de recherche : en 23 ans, ses activités se sont diversifiées vers d'autres services en ligne, comme le *cloud computing*, la cartographie numérique, la réalité augmentée, etc. mais également la robotique et l'intelligence artificielle.

En 2014, Google rachetait la start-up londonienne DeepMind³¹, aujourd'hui fer de lance dans le domaine de l'IA. Cette société crée notamment des briques logicielles destinées aux différents produits Google (Maps, Translate, Assistant, etc.).

Dévoilé en 2021, le système LaMDA (*Language Model for Dialogue Applications*) a pour but de tenir une conversation de manière naturelle sur de nombreux sujets. On peut y voir une référence au projet fondateur des pionniers de l'IA : rendre indiscernables l'humain de la machine dans une conversation (test de Turing). Ce système pourrait être prochainement ajouté à l'assistant numérique Google Assistant (déployé en 2016).

Aujourd'hui, **Google** revendique l'utilisation et le développement d'IA ouverte, et fournit des données accessibles à tous³². L'offre de Google en IA est principalement intégrée à ses produits, majoritairement en B2C (vers les consommateurs).

Facebook (Meta) emploie des milliers de chercheurs en IA. L'investissement de Facebook dans l'IA est stratégique pour son activité, notamment parce que Facebook ne peut pas modérer ses contenus sans l'apport de l'IA. Pour rappel, Facebook possède 3 milliards de comptes actifs et un chiffre d'affaires de 86 milliards de dollars. La pression réglementaire sur les réseaux sociaux s'intensifie³³ : d'abord dans l'Union européenne, où après des

29. Lettre du 10 juin 2020 de Marie-Laure Denis, présidente de la CNIL.

30. ISBN 9789240029200, <https://www.who.int/publications/i/item/9789240029200>

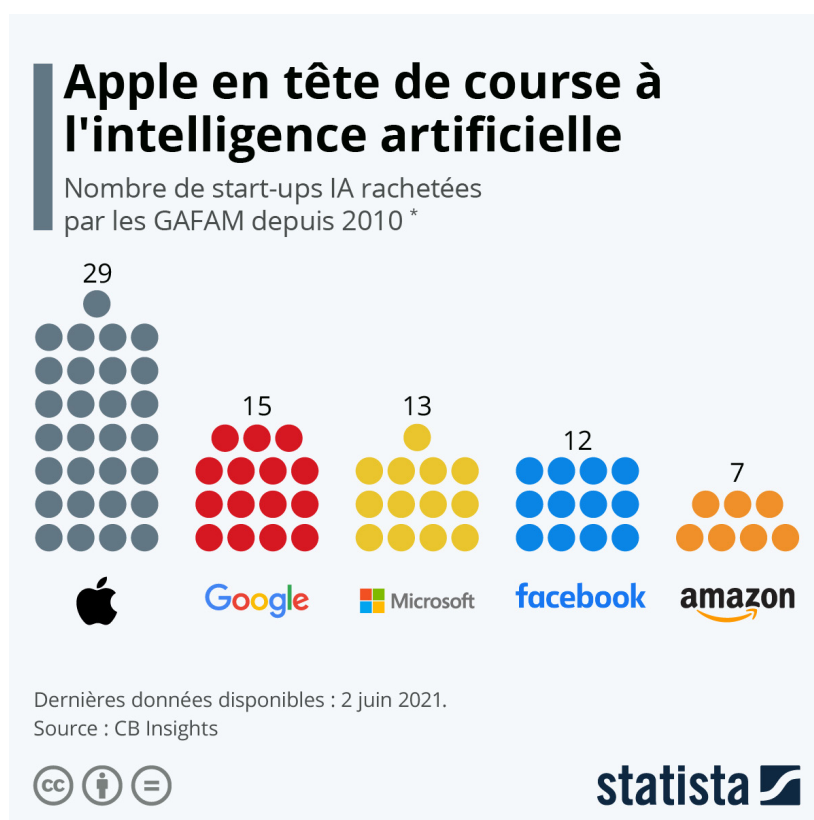
31. <https://www.lesechos.fr/tech-medias/hightech/deepmind-le-bras-arme-de-google-sur-lia-devient-rentable-pour-la-premiere-fois-1352937>

32. <https://www.lesechos.fr/tech-medias/hightech/google-veut-faire-de-lintelligence-artificielle-son-allie-1318661>

33. <https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-20210325-SD002.pdf>

mesures spécifiques (appels à la haine ou à la violence, pédopornographie, etc.), le projet de Data Service Act en cours d'adoption prévoit la mise en place de réseaux de lanceurs d'alerte de confiance, et leur articulation dans les systèmes judiciaires des États membres de l'UE. De nombreuses initiatives étatiques vont dans ce sens, y compris dans certains États des États-Unis. Par ailleurs, les techniques d'IA permettent à Facebook de fournir différents services à ses clients, par exemple la traduction de contenus ou la sélection des conversations proposées à la lecture.

Amazon Web Services (AWS), qui est la partie *cloud computing* de la compagnie d'e-commerce Amazon, fournit des services en ligne (API) à destination des entreprises. Cette division représente 7 % du CA d'Amazon, et sa première source de profit. AWS a été officiellement lancé en 2006 et, selon ses données commerciales, traite 89 % des projets *deep learning* hébergés sur un *cloud* dans le monde.



Microsoft est une multinationale américaine créée par Bill Gates et Paul Allen. Fondée en 1975 à Albuquerque, elle se concentre sur l'informatique et la micro-informatique.

L'entreprise mène depuis plusieurs années une stratégie d'acquisition : Skype, Nokia mobile OY ou encore LinkedIn ont été par exemple rachetées par Microsoft. La multinationale est également à la pointe dans le domaine de l'intelligence artificielle, notamment à travers son laboratoire Microsoft Research de Cambridge.

Avec Azure, Microsoft détient la seconde place sur le marché du *cloud* avec 22 % de parts de marché, derrière Amazon (AWS) qui en détient 33 %. Ces services incluent l'hébergement et des services de logiciels, y compris les algorithmes de l'intelligence artificielle. En outre, Microsoft ouvre des écoles « alternatives » de formation aux techniques de l'IA : après Brest et Rennes, une école en partenariat avec Simplon a été ouverte à Lannion en mars 2022.

Les acteurs chinois (BATX)

Si nous n'avons pas le détail des projets en cours, il est à noter que les géants du Net chinois ont tous annoncé des investissements significatifs sur l'IA ces dernières années.

BAIDU est le Google chinois, fondé en 2000 par Robin Li et quatrième site le plus visité au monde. Il reprend les mêmes fonctionnalités que Google (mails, agenda, plateforme de vidéos) et commence à le concurrencer sur le terrain de l'intelligence artificielle, ou la conduite autonome. En 2021, Baidu a vu ses bénéfices augmenter significativement, notamment grâce à l'IA³⁴. Sa capitalisation boursière est de l'ordre de 52 milliards en juillet 2022.

ALIBABA est le cousin d'Amazon. Même s'il s'est diversifié, ce géant tire la grande majorité de ses revenus du commerce en ligne. Fondé en 1999 par Jack Ma (lequel a désormais quitté ses fonctions suite aux pressions des autorités chinoises). Le groupe Alibaba a annoncé en mai 2020 des investissements de 1,4 milliard de dollars dans le domaine de l'intelligence artificielle et de l'Internet des objets (IoT)³⁵. Le groupe est l'une des plus grosses capitalisations boursières d'Asie, avec 311 milliards d'euros à la bourse de New York.

TENCENT est le parent de l'application WeChat, dont les multiples fonctionnalités rappellent largement Facebook. Fin 2018, alors que la capitalisation boursière de Facebook culminait à 404 milliards de dollars, celle de Tencent atteignait 357 milliards. En 2022, elle est la septième capitalisation mondiale avec une valeur de 438 milliards de dollars, et la seule non américaine du Top 10.

XIAOMI³⁶ est le numéro 2 mondial des ventes de smartphones, grâce à des produits haut de gamme moins chers que ceux des marchés occidentaux³⁷. L'entreprise se développe également dans les technologies d'IA et la robotique. En 2020, son PDG Lei Jun a annoncé vouloir dépenser plus de 7 milliards de dollars sur les cinq prochaines années dans différents domaines, tels que l'intelligence artificielle, l'Internet des objets, ou la 5G.

Les autres acteurs internationaux

Hormis les géants du Net, de nombreuses entreprises multinationales interviennent comme acteurs dans le domaine de l'intelligence artificielle. Citons ici les sociétés de services comme IBM, OVHcloud³⁸, Cisco, Oracle, les prestataires comme Accenture³⁹, Capgemini⁴⁰, Sopra bank⁴¹, ainsi que les constructeurs comme Nvidia ou Intel.

34. <https://www.lefigaro.fr/secteur/high-tech/l-intelligence-artificielle-fait-bondir-les-benefices-du-moteur-de-recherche-baidu-20210518>

35. <https://siecledigital.fr/2020/05/26/alibaba-investit-14-milliard-de-dollars-dans-lintelligence-artificielle-et-liot/>

36. <https://www.franceculture.fr/numerique/lexpansion-des-batx-les-gafam-chinois>

37. <https://intelligence-artificielle.com/xiaomi-guide-complet/>

38. <https://www.ovhcloud.com/fr/public-cloud/ai-machine-learning/>

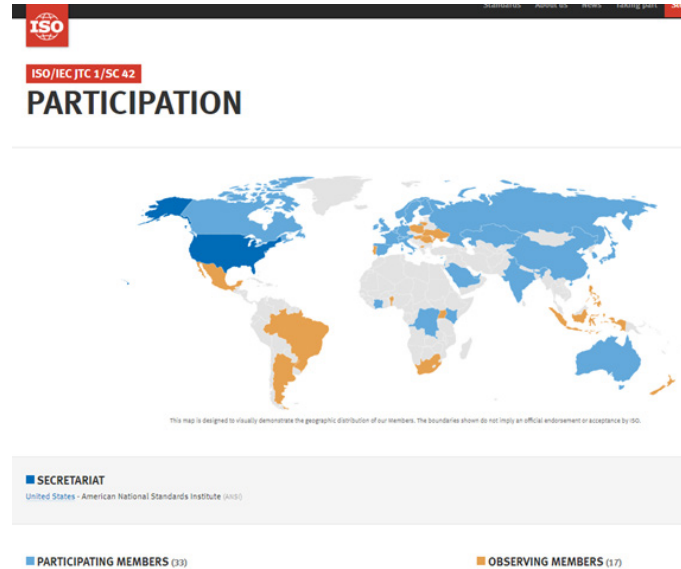
39. <https://www.accenture.com/fr-fr/insights/artificial-intelligence-index>

40. https://www.capgemini.com/wp-content/uploads/2020/07/AI-in-CX_CRI-Report_16072020_V4.pdf

41. <https://www.soprabanking.com/fr/par-technologie/intelligence-artificielle/>

Les instances de normalisation

Le travail de normalisation internationale sur l'IA est confié à un comité technique joint ISO & IEC⁴² : c'est le sous-comité 42 [ISO/IEC JTC 1/SC 42], dont le secrétariat est assuré par l'ANSI. Ce sous-comité comprend 50 membres (dont 33 ont une participation active).



À l'heure où nous écrivons, seuls trois rapports techniques ont été publiés ; de nombreux projets sont en cours de rédaction.

On trouvera par ailleurs en annexe la liste des projets de développement dans ce cadre. Cette longue énumération montre l'intérêt que suscite le sujet, et la diversité des thèmes abordés par l'IA.

Tendances et incertitudes majeures

Investissements R&D

Comme vu précédemment, nous assistons à une augmentation massive des investissements des États (lors des 15 dernières années, il y a eu une accélération des investissements). Cette augmentation devrait encore s'accroître lors de la prochaine décennie. Dans le même temps, le secteur privé investit massivement. La question est celle du fléchage de ces investissements.

L'intelligence artificielle est devenue un enjeu géopolitique majeur : elle intervient dans les différentes formes de « pouvoir » (économique, militaire, d'information et d'influence, etc.) et cette tendance devrait se renforcer avec le temps.

Équilibre public-privé

Si, du côté étatique, la pandémie de coronavirus a pu ralentir certains projets liés à l'intelligence artificielle, en revanche, les acteurs privés ont accru leur emprise et ont

42. <https://www.iso.org/committee/6794475/x/catalogue/p/1/u/1/w/1/d/1>

profité du mouvement de « digitalisation » de la société.

Une crise planétaire de grande ampleur pourrait engendrer des conséquences similaires : il y aurait un retrait des États, qui se trouveraient limités à des rôles de régulation et qui, pour lutter contre les effets de la crise, laisseraient le champ libre aux sociétés privées pour augmenter la digitalisation des échanges économiques et sociaux.

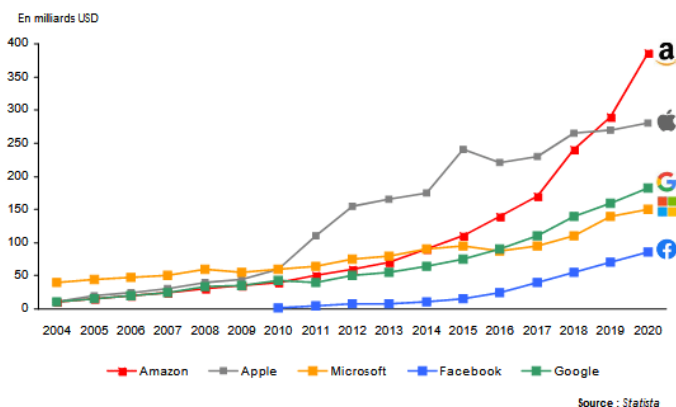
En France, de nombreuses initiatives de digitalisation de l'accès aux services publics se sont développées depuis 15 ans. Par exemple :

- Des algorithmes de détection de fraude fiscale, une plateforme d'accès aux services publics (ex : caisse primaire d'assurance maladie, déclaration de revenus pour entreprises et particuliers).
- Nombre de guichets ont été transformés en « plateformes ».
- Mise en place de services de « coaching numérique » (sorte d'écrivain public du XXI^e siècle), qui permet d'étendre cette « plateformes ».

Dans le secteur de la **justice**, certaines initiatives sont en cours : par exemple, la création de bases de données permettant d'estimer les réparations liées à des différends. L'estimation des dossiers à instruire par les cabinets d'avocats en est également à un stade expérimental, mais devrait se renforcer dans les années qui suivent. Nous pensons donc que la progression sera modérée sur les 10 prochaines années.

Du côté de la **sécurité et sûreté du territoire**, le nombre d'installations de caméras de surveillance ne cesse de croître. L'intelligence artificielle jouera un rôle prépondérant pour les « augmenter » avec différents dispositifs prédictifs (incidents, accidents et actes délictueux, etc.). La dernière décennie a vu l'apparition de la « cyberguerre » (cas en 2007 de l'Estonie paralysée par des pirates russes).

Pour les **géants du numérique**, si l'on suit la tendance depuis leur naissance dans les années 2000 jusqu'à devenir les plus grosses sociétés mondiales, la progression est fulgurante (cf. graphique ci-dessous). Toutes sont parties prenantes dans le développement de techniques de l'IA. Nous nous attendons donc à une croissance soutenue dans les deux prochaines décennies. L'inconnue majeure est la possibilité de limiter leur croissance par la régulation (droit à la concurrence, fiscalité) ou leur partition en entités moins grandes pour répondre aux problèmes d'abus de position dominante.⁴⁴



44. <https://www.atlas-mag.net/article/la-capitalisation-boursiere-des-gafam>

Hypothèses

Le maintien d'un *statu quo* est assez peu probable.

Hypothèse 1. Percée des solutions open-source

Les solutions *open-source* unifient les efforts de développement et la diffusion de l'IA.

Les grands acteurs privés et les acteurs publics « offrent » leurs algorithmes aux communautés de l'*open-source*. Cela rend les algorithmes plus transparents, mais ne signifie pas nécessairement un progrès vers l'explicabilité.

L'*open-source* apparaît comme un facteur favorable à la diffusion de l'IA (accès à des communautés de développeurs et à des viviers de compétences).

Pour expliquer la raison pour laquelle des entreprises ouvrent leurs techniques à un grand nombre d'acteurs, notons que la multiplication des « cas d'usages » peut être également utilisée comme une source précieuse pour des études de marché (à travers la mesure des activités des utilisateurs).

La multiplication des acteurs permet de dynamiser le développement et le déploiement des techniques d'IA (réalisation de projets *proof-of-concept* sans engager de gros moyens financiers).

Hypothèse 2. « Business as usual »

Dans cette hypothèse, le développement des systèmes d'intelligence artificielle se poursuit sur la lancée actuelle, avec un développement à bon rythme et le prolongement de l'équilibre (instable) public-privé. Les systèmes déployés se cantonnent à l'aide à la prise de décision et à ce qui est utile au plus grand nombre, mais sans engendrer de modifications profondes dans la relation aux techniques du numérique. Le principal obstacle à envisager est celui de fortes réactions sociales de la part de personnes qui se sentiraient « dépassées » ou « envahies » par ces techniques.

Hypothèse 3. Reprise en main des États

Les États imposent un certain contrôle du secteur privé et affirment leur position, non seulement en tant que régulateur, mais également en tant que moteur de la recherche. Dans un contexte géostratégique tendu, l'intelligence artificielle occuperait une place centrale et deviendrait l'enjeu des luttes des puissances. Cela ne signifierait pas automatiquement une démocratisation du contrôle de l'IA, les forces en présence étant concentrées dans les mains de quelques grands acteurs leaders du déploiement de l'intelligence artificielle. Cette reprise en main peut s'inscrire dans un contexte conflictuel (une guerre froide numérique durable) ou, à l'inverse, permettre des initiatives multilatérales, par exemple pour bannir certaines pratiques.

Hypothèse 4. Déploiement effréné

Les entreprises deviennent le moteur principal du développement de l'IA et imposent leur rythme aux États et aux peuples. Elles supplantent peu à peu les fonctions régaliennes

en devenant également des fournisseurs de « services publics ». L'influence croissante des entreprises conduirait à des changements profonds de la société, de manière à faire accepter les techniques d'intelligence artificielle comme « la norme » (par exemple, dans le domaine de la santé, de l'éducation, de l'orientation professionnelle ou encore de la gestion des budgets sociaux).

Enfin, nous avons tenu à ajouter une dernière hypothèse, qui est peut-être plus un scénario.

Hypothèse 5 / Scénario : Imprévu majeur (« wild card »)

Une brutale aggravation de la situation climatique ou d'un conflit (tel que l'Ukraine) entraîne des migrations massives, des conflits sur l'alimentation et les ressources en eau potable, un repli nationaliste et des tensions internationales insurmontables. L'intelligence artificielle se déploie en premier lieu sur le terrain militaire et sécuritaire.

Annexe

Liste des standards et normes en cours de développement par le comité ISO 42 [ISO/IEC jTC1 / SC42]

- ISO/IEC 4213 Information technology – Artificial Intelligence – Assessment of machine learning classification performance
- ISO/IEC 5259-1 Data quality for analytics and ML – Part 1: Overview, terminology, and examples
- ISO/IEC 5259-2 Data quality for analytics and ML – Part 2: Data quality measures
- ISO/IEC 5259-3 Data quality for analytics and ML – Part 3: Data quality management requirements and guidelines
- ISO/IEC 5259-4 Data quality for analytics and ML – Part 4: Data quality process framework
- ISO/IEC 5338 Information technology – Artificial intelligence – AI system life cycle processes
- ISO/IEC 5339 Information Technology – Artificial Intelligence – Guidelines for AI applications
- ISO/IEC 5392 Information technology – Artificial intelligence – Reference architecture of knowledge engineering
- ISO/IEC TR 5469 Artificial intelligence – Functional safety and AI systems
- ISO/IEC TS 5471 Artificial intelligence – Quality evaluation guidelines for AI systems
- ISO/IEC TS 6254 Information technology – Artificial intelligence – Objectives and approaches for explainability of ML models and AI systems
- ISO/IEC TS 8200 Information technology – Artificial intelligence – Controllability of automated artificial intelligence systems
- ISO/IEC 22989 Information technology – Artificial intelligence – Artificial Intelligence Concepts and Terminology
- ISO/IEC 23053 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
- ISO/IEC 23894 Information technology – Artificial Intelligence – Risk Management
- ISO/IEC TR 24027 Information technology – Artificial Intelligence (AI) – Bias in AI systems and AI aided decision making
- ISO/IEC 24029-2 Artificial Intelligence – Assessment of the robustness of neural networks – Part 2: Methodology for the use of formal methods
- ISO/IEC TR 24368 Information technology – Artificial Intelligence (AI) – Overview of ethical and societal concerns
- ISO/IEC TR 24372 Information technology – Artificial Intelligence – Overview of computational approaches for AI systems

- ISO/IEC 24668 Information technology – Artificial intelligence – Process management framework for big data analytics
- ISO/IEC 25059 Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality model for AI-based systems
- ISO/IEC 38507 Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations
- ISO/IEC 42001 Information technology – Artificial intelligence – Management system

Acceptabilité sociale générale (citoyen) de l'intelligence artificielle

V. Mandinaud

Cette fiche vise à identifier les différents accueils de l'IA en fonction de ses applications possibles, des domaines concernés, des retombées concrètes sur la vie quotidienne, ainsi que des différentes « catégories » de population, d'éventuelles éruptions de technophobie (ou de technophilie)... Quelles seront les grandes questions autour desquelles vont se structurer les débats et se construire les opinions ? Quels types de mobilisation pour les mouvements anti-IA ? Quels acteurs et espaces de discussion sur ces questions ? Des connexions peuvent être envisagées avec l'approche spécifique au monde du travail, traitée dans les fiches 7 et 7bis.

Contexte

Les problématiques d'acceptabilité sociale des projets technologiques – tout comme les tentatives d'institutionnalisation de leur traitement – ne sont pas nouvelles. On pourrait les faire remonter aux luddites, et plus encore. Plus proches de nous, plusieurs générations de travaux regroupés sous le vocable de « Technology Assessment » (TA)¹ ont émergé dans les pays de l'OCDE à partir des années 60. Tournés vers une logique de maîtrise « sociale » des choix technologiques ou de participation des publics à la conception et à l'orientation de projets scientifiques, techniques ou industriels, ces travaux de recherche, mais aussi ces pratiques d'évaluation sociale des technologies, constituent des références inspirantes pour les défis contemporains, notamment ceux liés au développement sans précédent des outils numériques, et de l'intelligence artificielle.

S'agissant de cet objet, en France, on peut considérer que les travaux réalisés par la Commission nationale de l'Informatique et des Libertés (CNIL), le Comité national du

1. Valenduc Gérard, « La technologie, un jeu de société : au-delà du déterminisme technologique et du constructivisme social », éditions Académia, 2005.

Numérique (CNUM) ou le Comité national pilote d'Éthique du Numérique (CNPEN) sous l'égide du Comité consultatif national d'Éthique (CCNE), au-delà de leurs différences, participent à ce travail institutionnalisé que nous qualifions d'« évaluation sociale » des technologies.

En deçà de ces pratiques institutionnelles, les problématiques d'acceptabilité sociale renvoient de façon plus ordinaire à des situations où des groupes concernés (bien souvent regroupés en associations et appuyés par des ONG) contestent le bien fondé, les modalités et les conséquences de projets technologiques ou d'assemblages sociotechniques². On parle alors de controverses, de polémiques, d'alertes, de débats publics, et parfois de scandales ou d'affaires³. Par-delà leur différences et spécificités, ces configurations renvoient à des dynamiques à la fois cognitives et politiques. Elles portent sur des dimensions tant technologiques que sociales, et donnent lieu à des formes publiques de critiques des assemblages sociotechniques remis en question, et parfois du cadre juridico-politique dans lequel ces assemblages prennent forme. Et elles participent à la régulation des projets technologiques : les arènes qu'elles construisent permettent de mettre à l'épreuve les grammaires de justification de l'acceptable et de l'inacceptable.

Définitions et concepts

Au flou qui entoure l'IA, pour l'exercice, il faut aussi reconnaître le flou qui entoure la notion d'acceptabilité sociale.

La notion d'**acceptabilité sociale** fait l'objet de diverses définitions tournant autour des enjeux d'assentiment ou de consentement libre et éclairé de la part des populations concernées, fondé par un jugement collectif. Elle se distingue donc de l'acceptabilité individuelle, et ne saurait se confondre avec la seule somme des acceptabilités individuelles. La notion n'est toutefois pas stabilisée.

Elle fait aussi l'objet de critiques, dans la mesure où elle véhicule une représentation « essentialiste » des technologies et des publics, alors que de nombreux travaux mettent plutôt l'accent sur des processus de co-évolution des objets technologiques et des publics concernés (tant par les conditions que par les conséquences de leur développement).

De plus, la notion d'acceptabilité sociale marque le fait de la primauté des intérêts et des capacités d'action des agents du marché et des politiques professionnels face à un social considéré comme résiduel. Elle constitue un révélateur d'une variété de légitimités et de capacités à transformer le monde commun et les règles qui le régissent. Elle rend compte d'une « dissymétrie structurelle du fonctionnement social »⁴, et désigne une ligne de front qui peut être saisie à travers le suivi des cadrages et débordements des projets opérés par les protagonistes concernés. D'une part, à travers les actions conduites par les

2. Rémi Barbier et Alain Nadaï, « Acceptabilité sociale : partager l'embarras », *Vertigo* - la revue électronique en sciences de l'environnement [Online], volume 15 n° 3, décembre 2015. URL : <http://journals.openedition.org/vertigo/16686> ; DOI : <https://doi.org/10.4000/vertigo.16686>

3. Francis Chateauraynaud et Didier Torny, « Les sombres précurseurs, une sociologie pragmatiste de l'alerte et du risque », 2002.

4. Barbier, Nadaï, op.cit.

promoteurs et leurs alliés pour rendre le projet acceptable et le faire accepter ; et d'autre part, et en prolongement, à travers les épreuves rencontrées par les populations concernées pour reprendre et poursuivre l'enquête sur les qualités des projets mis en cause, et aussi à travers les épreuves auxquelles elles soumettent le projet. En ce sens, l'acceptabilité sociale désigne moins une notion substantielle, qu'on pourrait mesurer précisément, qu'une dynamique de promotion/contestation ou qu'une activité de critique/régulation publique des innovations ou changements sociotechniques.

Cette dynamique peut prendre diverses formes et peut par exemple être appréciée à travers des enquêtes d'opinion, des manifestations publiques de soutien ou d'opposition, des publications savantes ou journalistiques, des actions en justice, etc. On tâchera de distinguer et d'articuler ce qui peut apparaître comme relevant d'une logique visant l'acceptation des projets, et ce qui peut apparaître comme relevant d'une critique sociale de l'acceptabilité des projets.

Résumé de la controverse par/pour un lecteur pressé

L'IA balance entre mythologie journalistique, littéraire, cinématographique, et projets de recherche et développements industriels. Le mythe de la singularité⁵, savamment entretenu par les auteurs, promoteurs et autres entrepreneurs, a largement dominé les représentations véhiculées pendant la décennie qui vient de s'écouler. Ce mythe de l'IA forte a – pour un certain nombre de sommités scientifiques en la matière – trop longtemps détourné l'attention des véritables problématiques et enjeux contemporains posés par le déploiement des IA faibles dans la vie courante, fussent-elles classiques ou issues du *machine learning* ou du *deep learning*.

Car, si elles sont invisibles, les IA faibles sont déjà partout ou presque. Elles ont conquis les espaces privés, domestiques, professionnels, publics, politiques, sociaux, culturels, civils et militaires de manière fulgurante. Le tableau ci-après, sans prétendre à l'exhaustivité, donne à voir comment les grandes fonctions identifiées des algorithmes se retrouvent dans différents secteurs.

5. Jean-Gabriel Ganascia, « Le Mythe de la Singularité, Faut-il craindre l'intelligence artificielle ? », éditions du Seuil, 2017.

Les grandes fonctions des algorithmes et de l'IA dans différents secteurs

	Education	Justice	Santé	Sécurité	Travail, RH	Culture	Autres
Générer de la connaissance	Mieux cerner les aptitudes d'apprentissage des élèves	Mettre en évidence les manières différenciées de rendre la justice selon les régions	Tirer profit de la quantité immense de publications scientifiques	Repérer des liens insoupçonnés pour la résolution d'enquêtes par les services de gendarmerie	Comprendre les phénomènes sociaux en entreprise	Créer des œuvres culturelles (peinture, musique)	Affiner le profil de risque de client d'un assureur
Faire du matching	Répartir les candidats au sein des formations d'enseignement supérieur (APB)		Répartir des patients pour participation à un essai clinique		Faire correspondre une liste de candidatures avec une offre d'emploi		Mettre en relation des profils « compatibles » sur des applications de rencontres, etc.
Prédire	Prédire des décrochages scolaires	Prédire la chance de succès d'un procès et le montant potentiel de dommages-intérêts	Prédire des épidémies Repérer des prédispositions à certaines pathologies afin d'en éviter le développement	Détecter les profils à risque dans la lutte anti-terroriste Prédire l'occurrence future de crimes et délits	Détecter les collaborateurs qui risquent de démissionner dans les prochains mois	Créer des œuvres ayant un maximum de chance de plaire aux spectateurs (Netflix)	
Recommander	Recommander des voies d'orientation personnalisées aux élèves	Recommander des solutions de médiation en fonction du profil des personnes et des cas similaires passés			Proposer des orientations de carrière adaptées aux profils des personnes	Recommander des livres (Amazon), des séries télévisées (Netflix), etc.	Individualiser des messages politiques sur les réseaux sociaux
Aider la décision		Suggérer au juge la solution jurisprudentielle la plus adéquate pour un cas donné	Suggérer au médecin des solutions thérapeutiques adaptées	Suggérer aux forces de police les zones prioritaires dans lesquelles patrouiller			Aider le conducteur à trouver le chemin le plus court d'un point à un autre (GPS)

Source : « Comment permettre à l'homme de garder la main ? Enjeux éthiques des algorithmes et de l'intelligence artificielle », CNIL, 2017.

Mais au-delà des promesses et opportunités auxquelles les IA sont liées – qu'il s'agisse de recommander, prédire, produire de la connaissance, apparier une offre et une demande, aider à la décision, à la délégation de fonctions à des machines « autonomes » –, les systèmes d'IA (c'est-à-dire des algorithmes évolutifs créant leurs propres modèles à partir des données qui leur sont fournies) soulèvent de nombreuses questions d'ordre éthique, juridique et politique. Ce sont aussi les problématiques qu'elles posent et les risques et les dommages qu'elles peuvent causer qui ont commencé à être publicisés, sur la base de situations concrètes et à travers des questionnements relatifs aux biais construits ou amplifiés par les algorithmes et les données avec lesquelles ceux-ci sont alimentés, aux discriminations et exclusions induites, aux risques d'enfermement ou de formatage, à l'opacité du fonctionnement des algorithmes, et à la dilution des responsabilités. Plus encore, ce sont les orientations, domaines et finalités mêmes des investissements dans l'IA, ainsi que les règles de droit encadrant le développement de technologies et de leurs usages pour garantir le respect des droits et libertés fondamentales, l'égalité des citoyens et la dignité humaine, qui se sont retrouvées sur la place publique.

Ce qui apparaissait encore il y a peu comme des extravagances asiatiques (le « crédit social ») ou américaines (la police prédictive et la décision de justice automatisée), tout droit sorties d'un scénario dystopique axé sur la surveillance par l'usage intempestif de capteurs et l'exploitation de données de toutes sortes au service d'une gestion

algorithmique du bien-être, se donnent à voir comme des réalités toujours plus proches de l'UE, dont il s'agit de se prémunir sans pour autant désertir le champ du développement de l'IA, au risque de perdre en souveraineté et en puissance normative.

À des réflexions d'abord centrées autour des impacts des IA sur des décisions critiques (au sens de très importantes), à travers par exemple les cas des armes « autonomes », de la voiture « autonome » ou encore du diagnostic médical, s'est adjointe une réflexion sur les impacts des effets d'échelle de décisions non critiques générées par des IA faibles notamment à travers l'exemple des applications dédiées à la circulation routière ou autres⁶. À travers ces différents axes de questionnement, ce sont en fait la fiabilité et la légitimité des IA qui se sont retrouvées mises à l'épreuve.

Durant la décennie écoulée – en Europe tout du moins et plus largement dans les pays occidentaux –, les questions éthiques ont émergé sur les scènes académiques, institutionnelles, militantes et médiatiques. À tel point qu'elles donnent sans doute une représentation trompeuse selon laquelle il n'y aurait plus de projet d'IA sans comité éthique... Des initiatives particulièrement intéressantes ont toutefois le mérite d'exister, telles que la déclaration de Montréal pour un développement responsable de l'IA.

L'émergence de l'éthique des IA ces dernières années peut être comprise comme un signe de l'avènement en la matière d'un principe d'autorégulation promu par les acteurs du marché, ou bien, ainsi que le soulignait la CNIL dans son rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle⁷, comme une préfiguration de ce que pourrait être une norme juridique en la matière pour interdire, autoriser et encadrer le développement de ces technologies, de leurs applications et de leurs usages.

Alors que le projet de règlement européen suscitait beaucoup d'espoir en termes de vision éthique des stratégies de développement des IA attachées à la protection des libertés et des droits humains fondamentaux vis-à-vis des nouveaux développements et applications technologiques, la version publiée en avril 2021 par la Commission européenne soulève doutes, inquiétudes et questionnements de la part de la société civile.

Quelques indicateurs clés

Considérant que l'opinion publique ou que l'action publique ne préexistent pas à leur fabrique, que les enjeux d'acceptabilité sociale peuvent être saisis à la croisée de logiques industrielles, consuméristes, culturelles, scientifiques et politiques, et devant l'absence de données consolidées sur le sujet (encore moins longitudinales), nous proposons néanmoins de piocher dans différentes sources pour combiner différents indicateurs susceptibles d'éclairer les tendances et les possibles bifurcations en matière d'acceptabilité sociale de l'IA.

Pour approcher la réalité industrielle de l'IA et apprécier l'ampleur du phénomène, nous mobiliserons différentes données produites par la Direction Générale des Entreprises (DGE), Cisco, et l'Office Mondial de la Protection Industrielle (OMPI).

6. CNIL, op.cit.

7. CNIL, op.cit.

Pour approcher la **réalité commerciale** et apprécier le degré d'acceptation des applications d'IA dans la vie quotidienne, nous mobiliserons des données du marché des smartphones et des enceintes connectées (en tant que produit grand public spécialement « marketé » IA).

Pour approcher la **réalité du tournant éthique** de la recherche en IA, nous mobiliserons des données issues de l'« AI index report 2021 » de l'Université de Standford.

Pour approcher la **réalité médiatique et institutionnelle** de l'IA, nous mobiliserons des données issues de l'« AI index report 2021 » de l'Université de Standford et du rapport 2017 de la CNIL, « Comment permettre à l'homme de garder la main ? Enjeux éthiques des algorithmes et de l'intelligence artificielle ».

Pour approcher la **réalité de la critique militante**, nous mobiliserons des données issues des publications de Human Rights Watch (HRW), de European Digital Rights (EDRi) et de la Quadrature du Net en ce qui concerne la France.

Rétrospective des tendances

Le développement industriel de l'IA

Pour commencer, il est utile de rappeler que les développements actuels autour de l'IA s'inscrivent dans une histoire relativement longue de la discipline et du numérique, et qu'ils ont comme principaux points d'appui le développement d'Internet et de ses usages, ainsi que les croissances exponentielles de la quantité de données générées et des capacités de calcul des ordinateurs. Aujourd'hui, de par le monde, il y aurait plus de 50 milliards d'objets connectés déployés. Ces derniers produiraient près de 10 % des données créées, selon la DGE⁸ et d'après les chiffres de Cisco.

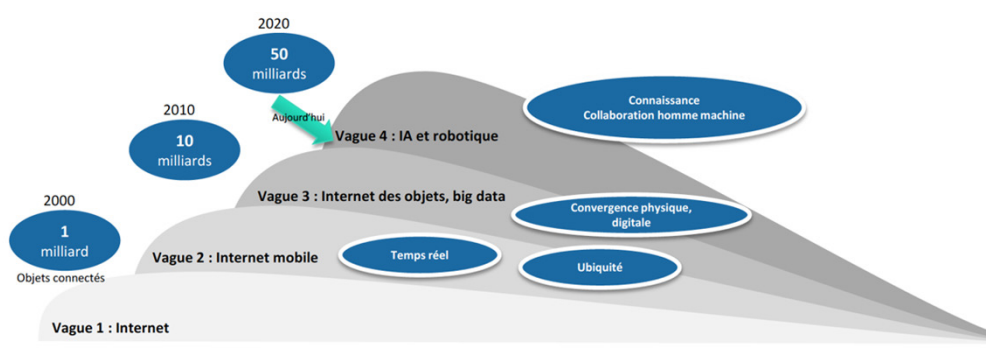


Figure 3 - Chronologie des évolutions numériques

S'agissant plus précisément de l'intelligence artificielle, le rapport daté de 2019 de l'Organisation Mondiale de la Protection Intellectuelle sur le sujet propose des éclairages sur les dynamiques à l'œuvre entre publications scientifiques, dépôts de brevets et

8. DGE, « Intelligence artificielle. État de l'art et perspectives pour la France », rapport final, 2017.

développement de technologies industrielles appliquées. Il indique en substance que les chercheurs ont publié plus de 1,6 million d'articles scientifiques en lien avec le thème de l'IA, et que près de 360 millions de demandes de brevets portant sur des inventions en lien avec des technologies d'IA ont été déposées. Plus de la moitié des brevets a été déposée à partir de 2013. Les premières publications scientifiques remontent à plusieurs décennies, mais leur nombre augmente à partir de 2001. En analysant l'évolution respective des publications scientifiques et des demandes de brevet, le rapport souligne également que nous sommes passés d'une recherche théorique à une utilisation industrielle et commerciale des technologies d'IA⁹.

Le rapport détaille¹⁰ :

« L'apprentissage automatique, qui est la principale technique d'intelligence artificielle divulguée dans les brevets, est inclus dans plus d'un tiers de l'ensemble des inventions recensées (134 777 documents de brevets). Le nombre de demandes de brevet en rapport avec l'apprentissage automatique a augmenté chaque année de 28 % en moyenne annuelle, 20 195 demandes de brevet ayant été déposées en 2016 (contre 9 567 en 2013). [...]

L'apprentissage profond et les réseaux neuronaux sont les techniques d'intelligence artificielle qui connaissent la croissance la plus rapide au regard du dépôt de demandes de brevet : l'apprentissage profond a connu un taux de croissance annuel moyen impressionnant de 175 % de 2013 à 2016, pour atteindre 2 399 dépôts de demandes de brevet en 2016 ; les réseaux neuronaux ont progressé à un taux de 46 % pendant la même période, avec 6 506 dépôts de demandes de brevet en 2016. [...]

La vision par ordinateur est mentionnée dans 49 % de l'ensemble des brevets en rapport avec l'intelligence artificielle (167 038 documents de brevet), soit une croissance annuelle moyenne de 24 % (21 011 demandes de brevet déposées en 2016). [...]

Les taux de croissance observés dans les données sur les brevets en rapport avec l'intelligence artificielle sont sensiblement plus élevés que le taux de croissance annuel moyen des brevets dans tous les domaines technologiques, qui s'élevait à 10 % entre 2013 et 2016. [...]

L'essor des technologies dans le domaine des transports devient plus évident lorsqu'on examine les tendances au cours de la période 2006-2016 : alors qu'elles représentaient seulement 20 % des demandes en 2006, elles constituaient en 2016 un tiers des demandes (plus de 8 700 dépôts de demandes). [...]

Les combinaisons les plus fréquentes dans les dépôts de demandes de brevet sont : l'apprentissage profond avec la vision par ordinateur, la vision par ordinateur avec le transport, les télécommunications et la sécurité, l'ingénierie ontologique avec le traitement du langage naturel, et l'apprentissage automatique avec les sciences de la vie et les sciences médicales. Ces combinaisons semblent indiquer les domaines où l'évolution rapide de l'intelligence artificielle dans un avenir proche est à suivre. [...]

Parmi les autres secteurs et sous-catégories au sein des secteurs ayant connu une croissance notable des dépôts de demandes de brevet, il convient de mentionner les villes intelligentes (croissance annuelle de 47 %), l'agriculture (32 %), l'informatique au niveau du gouvernement (30 %) et la banque et la finance (28 %). »

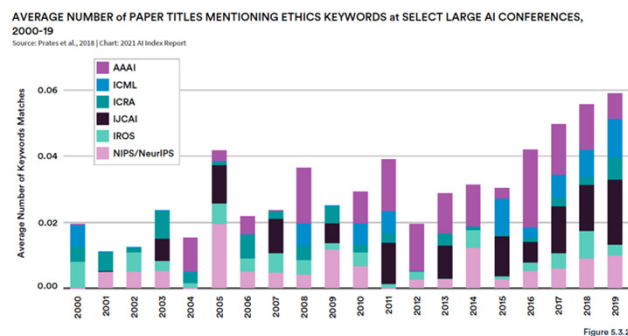
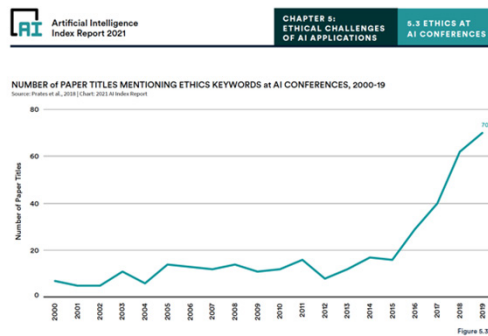
9. OMPI, Rapport sur les tendances technologiques, Intelligence artificielle, 2019.

L'émergence des questions éthiques dans la recherche sur l'IA

S'agissant du développement du thème de l'éthique dans les publications d'articles scientifiques qui traitent des IA, l'AI Index report 2021 publié par l'Université de Stanford apporte des éclairages intéressants. Ce thème connaît une croissance toute modérée des années 2000 à 2015, puis connaît depuis lors une croissance exponentielle. Mais si en dynamique le thème connaît une croissance importante, le rapport souligne qu'en volume le nombre d'articles publiés sur ce thème reste faible (au regard du nombre d'articles scientifiques publiés dans les principales conférences sur les IA).

Il y est spécifié que :

« Les chercheurs écrivent plus d'articles qui se concentrent directement sur l'éthique de l'IA, avec des soumissions dans ce domaine plus que doublées de 2015 à 2020. Pour mesurer le rôle de l'éthique dans la recherche en IA, des chercheurs de l'Université de Rio Grande do Sul, à Porto Alegre au Brésil, ont recherché des termes liés à l'éthique dans les titres d'articles de conférences sur l'IA, l'apprentissage automatique et la robotique. Comme le montre le premier graphique, il y a eu une augmentation du nombre d'articles portant sur le mot-clé « éthique » dans les titres soumis aux conférences sur l'IA depuis 2015. Une analyse plus approfondie du second graphique montre la moyenne de correspondances de mots-clés dans toutes les publications parmi les six grandes conférences sur l'IA. »



On peut s'interroger sur la continuité de cette dynamique, sur sa signification et ses éventuels effets sur les développements à venir des IA.

L'essor du marché des IA grand public

Les IA sont partout, sans pour autant que nous en soyons tout le temps bien conscients. Elles sont dans les smartphones, les plateformes, les applications, etc. Mais certains produits sont explicitement « marketés IA », par exemple les enceintes connectées commercialisées par les géants du numérique. À cet égard, les données commerciales qui les concernent nous permettent d'apprécier l'intérêt et les pratiques consuméristes délibérément tournées vers les IA. Elles nous indiquent que ce type d'application grand public d'IA trouve ses clients sans trop de difficultés. Elles nous révèlent un marché en très forte progression, avec toutefois des disparités très marquées entre USA, Asie et Europe.

Un article du *Siècle digital*, publié en février 2021, pointe que dans le monde les vendeurs chinois gagnent du terrain, et relate :

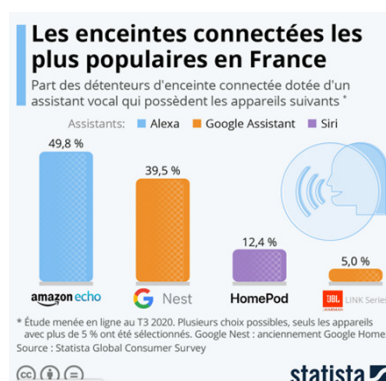
« En 2019, le marché des enceintes connectées a atteint un nouveau record : il représente désormais 146,9 millions d'euros. Entre 2018 et 2019, les observateurs de Strategy Analytics ont constaté une hausse de 70 %. »

Un article de *l'Usine nouvelle*, publié le 2 mars 2021, relate :

« Selon le cabinet Omdia, le nombre d'enceintes connectées écoulées dans le monde a bondi de 58 % en 2020 à 154 millions d'unités, portant le parc installé à 339 millions d'unités. Et ce chiffre pourrait passer à 512 millions à la fin de 2021. Les enceintes connectées constituent l'un des segments à la croissance la plus rapide du marché de la maison connectée. Selon le cabinet Omdia, le volume de ventes a augmenté de 58% en 2020 à 154 millions d'unités, représentant un chiffre d'affaires de 11,3 milliards de dollars. Les expéditions en Amérique du Nord ont bondi de 71 % à 72,9 millions d'appareils, la plus forte croissance du marché depuis 2017. Le marché asiatique a également connu un développement significatif augmentant de 48% pour atteindre 64 millions d'expéditions, tandis que l'Europe représentait près de 16 millions d'appareils. » (google translate)

Une enquête de Médiamétrie publiée en 2019 indique que la France compte 3,2 millions d'utilisateurs d'enceintes à commande vocale, alors qu'elle n'en comptait que 1,7 million en 2018.

S'agissant du marché français, ainsi que le montrent les données de la Statista Global Survey 2020, Amazon Echo est en position dominante, suivi par Google Nest.



La carrière médiatique de l'éthique des IA

L'IA Index Report 2021 indique que, pour identifier les défis éthiques posés par le développement et déploiement des IA tels que relatés dans la presse anglophone, les chercheurs de l'Université de Stanford ont exploité plus de 60 000 sources d'informations et 500 000 blogs en 2020. À travers les mots-clés « droits de l'homme », « valeurs », « responsabilité », « contrôle humain », « équité », « discrimination » ou « non-discrimination », « transparence », « explicabilité », « sûreté et sécurité », « responsabilité » et « vie privée », ils ont pu répertorier 3 046 articles liés aux technologies d'IA. Le graphique ci-dessous présente les 7 grands thèmes rattachés aux récits médiatiques analysés.

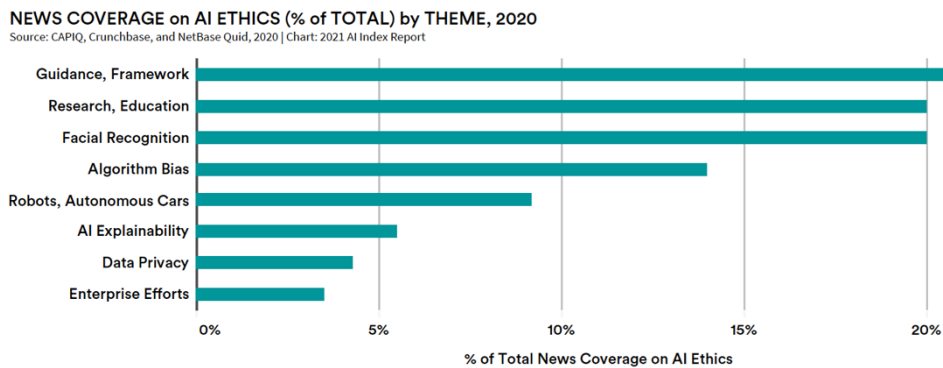


Figure 5.2.1

Les auteurs du rapport indiquent également que les 5 sujets d'actualité en lien avec l'IA qui ont retenu le plus l'attention en 2020 sont :

- La publication du livre blanc de la commission européenne sur l'IA
- Le licenciement par Google d'une chercheuse en éthique
- Le comité d'éthique formé par les Nations Unies
- Le plan éthique de l'IA du Vatican
- L'abandon par IBM des activités de reconnaissance faciale

L'institutionnalisation de l'éthique des IA

Des chercheurs du Boston AI Ethics Lab ont recensé 117 principes éthiques publiés par des organisations publiques ou privées entre 2015 et 2020. Ils mettent en évidence que des thèmes communs ressortent, tels que la confidentialité, la responsabilité, la transparence, ou l'explicabilité. Ainsi que le montre le tableau ci-dessous, les données exploitées indiquent que ce sont les entreprises privées qui ont émis à ce jour le plus grand nombre de publications.

NUMBER of NEW AI ETHICS PRINCIPLES by ORGANIZATION TYPE, 2015-20

Source: AI Ethics Lab, 2020 | Chart: 2021 AI Index Report

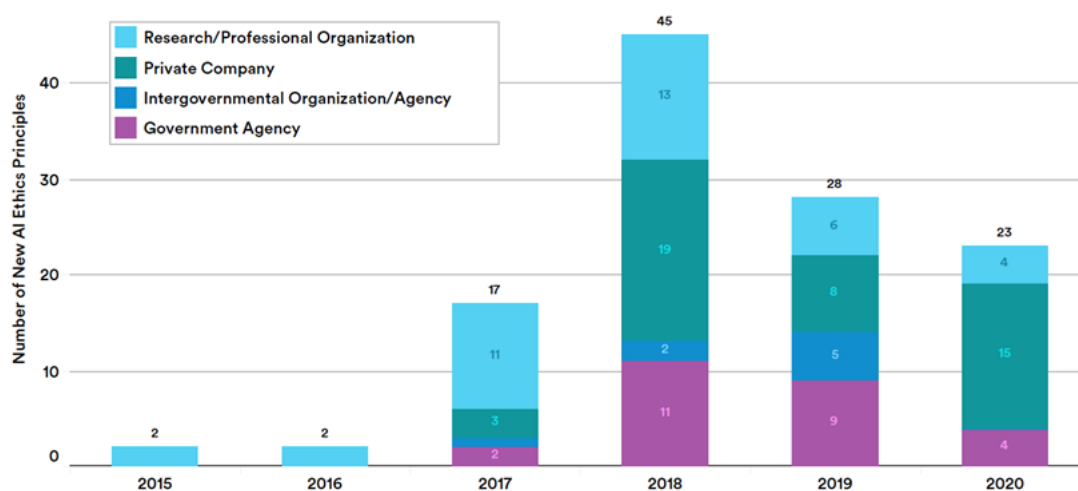


Figure 5.1.1

L'AI Index report 2021 souligne que l'année 2018 marque le point culminant du déploiement de principes éthiques pour les entreprises technologiques telles que IBM, Google ou Facebook, mais aussi pour diverses agences gouvernementales au Royaume-Uni, en UE et en Australie. Au-delà de ces chiffres, le rapport précise que « la prolifération des soi-disant principes éthiques ont été critiqués par les chercheurs en éthique et les praticiens des droits de l'homme qui s'opposent à l'utilisation imprécise des termes liés à l'éthique. Les critiques aussi soulignent qu'ils manquent de cadres institutionnels et sont sans engagement dans la plupart des cas. Le vague et l'abstrait, et la nature de ces principes n'offrent pas d'orientation sur la façon dont mettre en œuvre des directives éthiques liées à l'IA. »

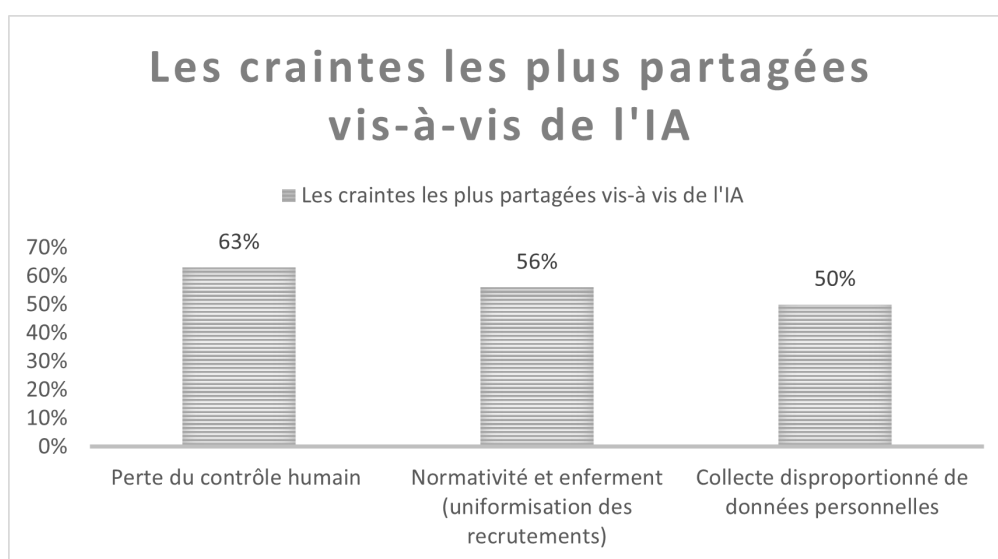
De plus, le Global Partners Digital and Stanford Global Digital Policy Incubator a publié en 2020 un rapport intitulé « Stratégies nationales d'intelligence artificielle et Droits de l'homme : un examen ». Ce rapport évalue dans quelle mesure les gouvernements ou les organisations régionales intègrent des considérations relatives aux droits humains dans leurs stratégies d'IA. Ce rapport indique que, parmi les stratégies des 30 États et des deux stratégies régionales (UE, pays nordiques et baltes) évaluées, un certain nombre font référence à l'impact des IA sur les droits de l'homme (le droit à la vie privée, l'égalité, la non-discrimination). Cependant, le rapport souligne que : « Très peu de documents de stratégie fournissent des informations approfondies, analyses ou évaluations concrètes de l'impact des applications de l'IA sur les droits de l'homme. Le détail de quand, comment et à quelle profondeur les droits de l'homme doivent être protégés dans le cadre de l'IA est largement absent, contrairement au niveau de spécificité sur d'autres questions, telles que l'économie, l'avantage concurrentiel et l'innovation. »

L'opinion publique sur les algorithmes et l'IA

D'après un sondage réalisé par l'IFOP pour la CNIL en janvier 2017, l'IA (et les algorithmes) est un sujet mal connu des Français. Si 83 % des sondés indiquaient qu'ils en avaient déjà entendu parler, plus de la moitié indiquait toutefois ne pas savoir précisément de quoi il s'agissait. Leur présence massive était appréhendée par 80 % des sondés, et 65 % d'entre eux estimaient que la dynamique allait se prolonger.

Si certains usages des algorithmes et de l'IA – comme la publicité – étaient bien repérés par 90 % des sondés d'une étude conduite en 2017 par Famille Rurales sur plus de 1 000 personnes dans le cadre d'un débat public organisé par la CNIL, seulement 46 % des sondés les identifiaient en ce qui concerne les applications de rencontre, et 33 % pour l'élaboration des diagnostics médicaux.

Selon une enquête réalisée en 2017 par l'IFOP pour la CFE-CGC, la perception publique des algorithmes et de l'IA est empreinte de méfiance. Ainsi que le montre le graphique ci-dessous, les trois craintes les plus partagées sont « la perte du contrôle humain », « la normativité et l'enfermement » et « la collecte disproportionnée de données personnelles ».



Source : Schéma du département Recherche et Prospective Stratégique de CNP Assurances adapté de l'enquête de la Confédération française de l'encadrement - Confédération générale des cadres CFE-CGC.

L'édition 2019 de l'Observatoire de la notoriété et de l'image de l'IA en France présente des résultats qui éclairent la situation sous d'autres aspects complémentaires. Elle indique notamment que 62 % des sondés sont confiants dans l'IA, même si cette confiance baisse de 5 points par rapport à l'édition 2018. Elle indique néanmoins que la vie privée devient le premier sujet de méfiance (38 %, + 10 points en un an) devant l'emploi (30 %, - 2 points en un an). L'étude indique également que 78 % des sondés considèrent que l'IA jouera un rôle important pour la santé dans le quotidien, et 72 % estiment qu'elle aura un rôle important dans l'éducation.

Les critiques militantes des IA

Si le projet de règlement européen de l'IA était porteur d'espoir pour nombre d'ONG afin de réguler la transition vers une gouvernance algorithmique du monde, force est de constater que celui-ci apparaît plutôt décevant. Il fait l'objet de différentes et vives critiques par les organisations qui militent pour la défense des libertés publiques et droits sociaux. Globalement, la critique générale consiste à montrer que ce projet de règlement est au service du marché et non des populations, qu'il protège davantage les promoteurs et développeurs d'IA que les populations, dans la mesure où il met en péril la protection

de droits fondamentaux alors qu'il s'annonce et se présente comme une manière de la renforcer tout en contribuant à l'augmentation générale du bien-être grâce au déploiement des applications d'IA.

Au-delà de la critique récurrente des applications d'IA concourant à la destruction des emplois, à une surveillance tous azimuts et au danger qu'elles représentent pour les libertés publiques, le document intitulé « How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net: Questions and Answers », et publié par l'ONG Human Rights Watch en réaction à la publication par la Commission européenne du projet de règlement de l'IA en avril 2021, pointe un autre danger : celui de la mise à mal du filet social par gestion algorithmique des droits sociaux.

Dans une première partie, le rapport pointe comment, à travers l'automatisation de la vérification des identités, de l'évaluation de l'éligibilité et du test de ressources, ou encore par le contrôle des fraudes et l'allocation de ressources, l'accès au droit et aux prestations, médié par des IA peut être entravé ou impacté et porter préjudice aux populations.

Dans une deuxième partie, le rapport souligne les lacunes et les limites de la proposition de réglementation européenne structurée autour d'une classification en 4 niveaux de risque (minime, limité, élevé, inacceptable), en ce qu'il ne serait pas assez protecteur des libertés et droits fondamentaux. Et ce, alors même que la Commission européenne le présente comme porteur d'une visée éthique et répondant aux demandes du Parlement européen et du Conseil européen, qui avaient exprimé la nécessité pour ce règlement de garantir que les droits des citoyens européens issus de la Charte des droits fondamentaux de l'UE soient pleinement respectés et appliqués. Le projet de règlement est ainsi passé au tamis d'une demi-douzaine de questions, telles que : « La réglementation empêche-t-elle la possibilité par les agences gouvernementales de noter ou de *profiler* les personnes qui demandent des prestations ou une aide à l'emploi ? » ou « La réglementation interdit-elle l'utilisation de la reconnaissance faciale pour filtrer les personnes qui demandent une aide à la sécurité sociale ? » À la première question, l'ONG répond que c'est flou, et à la seconde, elle répond non. Et, plus globalement, le jeu des questions/réponses s'attache à démontrer que « bien que le règlement de l'UE reconnaisse largement ces risques, il ne protège pas de manière significative les droits des personnes à la sécurité sociale et à un niveau de vie adéquat. En particulier, ses garanties étroites négligent la manière dont les inégalités existantes et les échecs à protéger adéquatement les droits – tels que la fracture numérique, les coupes dans la sécurité sociale et la discrimination sur le marché du travail – façonnent la conception des systèmes automatisés et s'y intègrent. »

Dans une troisième partie, l'ONG formule ses recommandations. Parmi celles-ci, on peut trouver par exemple celle visant à limiter de manière efficace les atténuations de règles qui, en première lecture, devaient permettre d'écarter des situations présentant des risques inacceptables, ou encore celle visant à garantir que les systèmes à haut risque respectent les droits humains à travers une exigence d'évaluation d'impacts (*ex ante*, en continu et *ex post*) tout au long des cycles de vie des systèmes : conception, approvisionnement, modification, post-déploiement.

De son côté, l'Européen Digital Rights (EDRi) publie le 3 août 2021 sa réponse au projet de réglementation porté par la Commission européenne. Dans le document intitulé « European Commission adoption consultation: Artificial Intelligence Act », l'ONG entend montrer « comment l'utilisation des systèmes d'IA peut systématiquement cibler, nuire et exclure les communautés marginalisées » et propose de s'interroger sur les

capacités du règlement à « remédier à ces dommages structurels, en raison de sa tendance à la dérèglementation de tous les usages « inacceptables », sauf les plus étroitement définis ; l'absence d'obligations pour les utilisateurs ; et l'absence de recours individuels et collectifs pour les personnes soumises à des systèmes d'IA ». Devant les inquiétudes pour les libertés et droits fondamentaux, l'EDRi recommande d'assurer une protection efficace contre les pratiques interdites et de traiter l'ensemble des risques inacceptables, d'adapter l'IA pour garantir un cadre holistique, démocratique et évolutif, de renforcer les obligations des utilisateurs de tous les systèmes d'IA, de mettre en œuvre une transparence publique significative pour les systèmes à haut risque, de faciliter la responsabilisation à travers des infrastructures de surveillance et d'application fonctionnant pour les personnes.

En France, les critiques formulées par l'association la Quadrature du Net sont encore plus vives. Pour l'association, le texte demande à être largement modifié car, dans son état actuel, « il risque de mettre en danger les maigres protections légales qu'offre le droit européen face au déploiement massif de techniques de surveillance dans l'espace public », dans la mesure où, « derrière le récit héroïque que propose la Commission européenne, se dissimule une attaque sournoise contre le droit européen à la protection des données personnelles, à travers une remise en question des principes acquis du RGPD et de la directive police-justice ». L'association estime que « les interdictions sont si limitées et si mal définies qu'on aurait presque l'impression que le but de la Commission européenne est d'autoriser le maximum de dispositifs plutôt que de véritablement en interdire certains ». Elle pointe également que des versions préparatoires du texte laissaient entrevoir un moratoire pour l'usage de la reconnaissance faciale et, qu'à travers cette version, la Commission a cédé au poids des lobbys sécuritaires. L'association dénonce le droit à l'analyse d'impacts promu par le texte, au détriment de mesures d'interdiction claires et fermes. Enfin, l'association souligne qu'aucune disposition du texte « n'offre de recours aux citoyen·nes vis-à-vis du déploiement de ces systèmes, la proposition se focalisant principalement sur la relation entre les entreprises fournissant ces systèmes et leurs client·es ».

Ces réactions contrastent avec le discours officiel qui supporte le projet de règlement européen de l'IA, qui entend « ouvrir la voie à des technologies éthiques dans le monde entier », selon les mots de la présidente de la Commission. Elles résonnent aussi avec les réactions des représentants des entreprises du numérique qui, elles, tendent plutôt à considérer que « ce projet de règlement européen sur l'IA encourage l'éthique du monde des affaires ».

Résumé, incertitude majeure et hypothèses prospectives

Résumé du champ problématique

Les investissements dans l'IA et les usages de l'IA n'ont jamais été aussi importants. Mais les inquiétudes restent vives. Les populations n'ont jamais autant eu recours aux IA et

acheté des enceintes connectées (subjuguées par la puissance des fonctionnalités), mais, dans le même temps, elles s'inquiètent pour la protection de leur vie privée. Au niveau des États, l'ambivalence demeure entre l'intérêt que représente l'IA pour la lutte contre le terrorisme ou contre la fraude fiscale, et la menace qu'elle représente pour la démocratie. Elle fait l'objet d'une compétition internationale où l'Union européenne n'est pas en position de force face à la Chine et aux USA. Elle entend toutefois créer des conditions de développement d'IA éthiques, dignes de confiance. Les applications de ces technologies recèlent un haut potentiel d'amélioration du bien-être humain. C'est pourquoi le projet de règlement européen affiche deux objectifs : promouvoir l'adoption de solutions basées sur l'IA, et faire face aux risques associés à certains usages de ces technologies. Construit autour d'une classification par niveau de risque (inacceptable, haut risque, modéré, minime), le règlement interdit certains usages jugés inacceptables (par exemple, l'utilisation de techniques subliminales) et encadre ceux qui sont porteurs de risques, y compris de haut niveau (par exemple, l'identification biométrique et la catégorisation des personnes).

Si cette proposition de règlement constitue une avancée juridique sans précédent en la matière, elle ne va pas sans poser de questions : apporte-elle réellement les garanties du respect des droits de l'homme, des libertés fondamentales et de la démocratie qu'elle prétend protéger ? Ne réduit-elle pas les voies de recours de la société civile et des populations face aux autorités administratives, aux puissances du marché et aux acteurs technologiques ?

Incertitude majeure

L'incertitude majeure qui pèse sur l'acceptabilité sociale des IA tient aux garanties offertes d'encadrement pour la garder sous contrôle humain, et la mettre au service de l'humain. Elle tient à la qualité des connaissances, informations et données produites et échangées, à la robustesse du cadre juridique, à la réalité de l'effectivité du droit – au sens des orientations stratégiques et politiques –, et aux dispositions pratiques et culturelles qui permettent aux populations d'exercer un travail démocratique d'évaluation sociale des technologies. Les limites de l'explicabilité des algorithmes, de la sécurisation offerte par le cadre juridique, des orientations stratégiques et politiques, de la culture des données et de leur gouvernance partagée constituent des freins à l'acceptabilité sociale des IA.

De plus, ici comme ailleurs, un certain climat d'instabilité politique semble s'être installé : entre capitalisme de surveillance débridé, technopopulisme autoritaire survolté et activisme démocratique renouvelé, nul ne sait encore de quel côté le vent peut tourner ! Que deviendrait-il de l'éthique de l'IA selon les circonstances ? Quelles orientations pourraient être données aux stratégies de développement des IA si d'aventure l'un ou l'autre de ses modèles venaient à s'affirmer plus clairement ? Et qu'en serait-il alors des conditions d'acceptabilité/d'inacceptabilité sociale des applications d'IA ?

Hypothèses

Hypothèse 1. Banalisation des IA et dévalorisation de ses critiques

La prolifération des IA et de leurs usages ne cesse de croître. Bien encadrés, les bénéfices l'emportent face aux risques. Sans pour autant les faire disparaître, les critiques sont amoindries ou démonétisées, car les catastrophes annoncées par les prophètes de malheur ne se sont pas réalisées. L'usage et le développement sont banalisés, même si les données sont largement collectées et confisquées par une poignée d'acteurs en position de force. L'acceptabilité sociale des IA est globalement renforcée dès lors qu'elles ne sont plus contestées, que certaines sont mêmes plébiscitées, et que le bilan s'avère positif dans le quotidien.

Hypothèse 2. Sécurisation des développements d'IA, enrayement des recours au contentieux et limitation de la portée des critiques

L'ambition normative et éthique de l'UE ne tient pas sa promesse, et s'étiole face à la concurrence internationale. Plutôt que de sécuriser le respect des droits humains fondamentaux, elle privilégie le développement d'expérimentations et d'initiatives publiques et privées pour rester dans la course, quitte à prendre quelques largesses avec les ambitions et orientations affichées initialement. Au lieu de construire un cadre de gouvernance partagée des données et des algorithmes, le droit est mis au service de l'innovation technologique et industrielle, et protège le marché de l'IA des velléités citoyennes. Le cadre juridique vient alors limiter les possibilités de contestation, de recours au contentieux. La critique n'est pas vraiment empêchée ni enrayée, mais sa portée s'en trouve limitée. De nouvelles applications d'IA sont imposées sans vraiment être débattues démocratiquement ni véritablement servir le bien-être des populations. L'acceptabilité sociale est entravée par l'écart entre une certaine évaluation sociale de ces technologies et la promotion institutionnelle qui en est faite. Sur le sujet, le champ médiatique – malgré sa mise sous tutelle par de grands groupes financiers – reste cependant controversé par des médias indépendants ou militants. Et des mouvements pour l'effectivité des droits humains à l'heure des IA se radicalisent et convergent dans les manifestations et mobilisations sociales qui ne cessent d'animer le pays depuis l'émergence des Gilets Jaunes.

Hypothèse 3. Expérimentations, institutionnalisation de la critique et renforcement démocratique des IA

Des expérimentations de nouvelles applications d'IA sont lancées dans différents secteurs ou domaines. La conduite de projet s'accompagne non seulement de comités d'éthique, mais associe de manière effective les parties prenantes à la gouvernance des projets d'expérimentation, de leur conception à leur évaluation, en passant par leur mise en œuvre et leur suivi. Les implications critiques ne sont pas repoussées, minimisées ni stigmatisées, mais incorporées au processus et sollicitées pour soutenir, enrichir, consolider et affiner des développements. La critique est institutionnalisée, sa valeur reconnue et son pouvoir accepté. L'acceptabilité sociale des IA est renforcée.

Les expérimentations contrôlées d'IA sont un levier de renouvellement de l'exercice démocratique dès lors qu'elles créent les conditions pour augmenter conjointement les capacités des IA, mais aussi des intelligences individuelles et collectives. Elles participent au développement des connaissances et des compétences, et s'accompagnent d'une nécessaire mise en culture partagée des données pour développer et faire fonctionner des IA de qualité.

Hypothèse 4. Tech-clash : entre ciblage et propagation

Considérant les menaces que représentent certains usages des données personnelles et applications d'IA de la *smart city* sur les libertés individuelles et collectives, une part grandissante des populations, mais aussi des gouvernements et des entreprises mêmes de « la tech » (à l'instar d'IBM ou du nouveau gouvernement allemand), rejettent certaines orientations et se refusent à procéder à certains développements ou usages qui seraient de nature à concourir à la surveillance généralisée des populations à partir de données biométriques. L'inacceptabilité sociale s'affirme et se concentre donc sur certaines orientations jugées particulièrement préoccupantes, notamment d'un point de vue démocratique et du respect de la dignité humaine. Si ces mesures de retrait sont circonscrites dans un premier temps, le front de la controverse gagne peu à peu, au gré des « leaks ». L'inacceptabilité s'étend à d'autres applications sur des secteurs jugés critiques (santé, énergie, etc.), malgré des opérations de séduction et transparence de la part des sociétés qui développent les applications. Certaines plus banales, comme celle du guidage routier, se voient elles aussi de plus en plus contestées et délaissées par une part grandissante de la population et des collectivités, tant sur les aspects de sécurisation des données et que sur leurs effets induits en matière de mobilité, d'urbanisme et de qualité de vie et ce, malgré des négociations et des programmes pour gagner la confiance des collectivités. Le déploiement des IA est entravé par des formes de mobilisations sociales et d'actions publiques exprimant leurs doutes sur les finalités et modalités du déploiement de ces technologies, leur méfiance à l'égard d'algorithmes incontrôlables aux effets démesurés, et leurs soupçons sur la captation et la réutilisation des données censées les alimenter.

Bibliographie

1. Barbier Rémi et Nadaï Alain, « Acceptabilité sociale : partager l'embarras », VertigO - la revue électronique en sciences de l'environnement [Online], Volume 15 Numéro 3, 2015.
2. Bradley Charles et Winfield Richard, « National artificial intelligence strategies and human rights: a review », Global Partners Digital and Stanford Global Digital Policy Incubator, 2020.
3. Chateauraynaud Francis et Torny Didier, « Les sombres précurseurs, une sociologie pragmatiste de l'alerte et du risque », Éditions de l'EHESS, 1999.
4. Cimino Valentin, « Les ventes des enceintes connectées sont en hausse : +70 % par rapport à 2018 », Le siècle digital, 2020.
5. CNIL, « Comment permettre à l'homme de garder la main ?, Enjeux éthiques des algorithmes et de l'intelligence artificielle », 2017.
6. Destribats Juliette, « 3,2 millions d'utilisateurs d'enceintes à commande vocale », Médiamétrie, 2019.
7. DGE, « Intelligence artificielle. État de l'art et perspectives pour la France », Rapport final, 2017.
8. EDRi, European Commission adoption consultation, Artificial Intelligence Act, 2021.
9. Ganascia Jean-Gabriel, « Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle ? », Seuil Éditions, 2017.
10. HRW, « How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net: Questions and Answers », 2021.
11. Human Centred Artificial Intelligence, Artificial Intelligence Index Report, Stanford University, 2021.
12. La Quadrature du Net, « Règlement IA : l'Union européenne ne doit pas céder aux lobbys sécuritaires », 2021.
13. La Quadrature du Net, « Règlement IA : la Commission européenne tend le piège de la reconnaissance faciale », 2021.
14. Loukil Rhida, « Les fabricants se frottent les mains devant le boom mondial des enceintes connectées », L'usine nouvelle, 2021.
15. OMPI, Rapport sur les tendances technologiques, Intelligence artificielle, 2019.
16. Siproudhis Jean-Baptiste, « Le projet de règlement européen sur l'intelligence artificielle encourage l'éthique des affaires », 2021.
17. UE, Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle, 2021.
18. Valenduc Gérard, « La technologie, un jeu de société : au-delà du déterminisme technologique et du constructivisme social », Éditions Académia, 2005.
19. Yuksel Tansu, « Les Français questionnent l'impact de l'IA en matière de confiance », Observatoire de la notoriété et de l'image de l'intelligence artificielle en France, 2019.

Règlementation sur l'IA et les données (monde, UE, France)

M. Bieri

Une tendance normative portée par l'Union européenne

Émergence des législations de protection des données personnelles

La création de la loi Informatique et Libertés, ainsi qu'une autorité administrative indépendante chargée de la protection des données et des libertés des citoyens (la CNIL), s'est déroulée dans un contexte particulier. Ses fondations ont été posées en 1974, en réaction au projet gouvernemental de regrouper trois fichiers d'État concernant les citoyens, afin d'en faire une base centralisée contenant une partie de l'ensemble des informations détenues sur les individus.

En effet, nous sommes au début de l'informatisation des fichiers d'État, et le projet SAFARI (« système automatisé pour les fichiers administratifs et le répertoire des individus ») est révélé par le journal *Le Monde* et provoque un certain tollé dans l'opinion publique, notamment chez certains intellectuels – la notion de fichage généralisé par l'État suscite de vifs débats, trente ans après la fin de la Seconde Guerre mondiale. Cette crise débouche sur l'abandon du projet : quatre ans plus tard, le 6 janvier 1978, est promulguée la loi Informatique et Libertés et est créée la CNIL, l'administration indépendante chargée de son application. L'idée est alors de créer un cadre protecteur des individus, en les dotant d'un certain nombre de droits techniques – cadre protecteur notamment vis-à-vis de l'État (à noter que d'autres pays ont eu des évolutions similaires, comme la Suède en 1973 et l'Allemagne en 1977).

Tandis que le ministère de l'intérieur développe la centralisation de ses renseignements Une division de l'informatique est créée à la chancellerie

En ordre dispersé, les départements ministériels tentent de développer à leur profit, à leur seul usage, l'informatique et son outil, l'ordinateur. Ce n'est pas tout à fait un hasard si, à l'époque où le Journal officiel va publier un arrêté créant une « division de l'informatique » au ministère de la Justice, celui de l'intérieur met la dernière main à la mise en route d'un ordinateur

puissant destiné à rassembler la masse énorme des renseignements graphiés sur tout le territoire; pas un hasard non plus si le projet SAFARI (système automatisé pour les fichiers administratifs et le répertoire des individus) destiné à détenir chaque Français par un « identifiant », qui ne définisse que lui, maintenant terminé, est l'objet de convoitises ardentes; le ministère de l'intérieur y souhaite

jouer le premier rôle. En effet, une telle banque de données, s'ouvrant à tous, donnera à quiconque la possibilité de connaître sans gêne

Ainsi se trouve d'évidence posé un problème fondamental, même s'il est rebattu : celui des rapports des libertés publiques et de l'informa-

tique. Son importance exigerait qu'il en soit, au Parlement, publiquement débattu. Tel ne paraît pas être, pourtant, la solution envisagée par le premier ministre dans les directives qu'il vient d'adresser au ministère de la Justice, intitulées au premier chef si l'on s'en rapporte à la Constitution qui dans son article 66 fait de l'autorité judiciaire le gardien des libertés individuelles.

« Safari » ou la chasse aux Français

Bref historique du concept contemporain de vie privée et de protection des données personnelles

Cette loi s'inscrit également dans une tendance plus globale autour du concept de vie privée (ou *privacy* chez les Anglo-saxons), qui s'est développé depuis le XIX^e siècle. Et ceci en contrepoint du développement des nouvelles technologies de l'époque, comme la photographie et l'imprimerie « de masse » qui ont permis la divulgation d'informations personnelles dans les journaux, rendant de plus en plus poreuse la frontière entre vie privée et informations publiques. On peut citer d'autres origines, comme le droit à la confidentialité des correspondances en France (identifié dès la Révolution) ou le droit à la préservation de son espace privé aux États-Unis (1791).

Ainsi, en 1890 apparaît la première formalisation d'un droit à la protection des données personnelles, attribuée à deux juristes américains, Samuel D. Warren et Louis D. Brandeis : ces derniers développent la notion d'un droit « *to be left alone* », littéralement d'« être laissé tranquille ».

Au cours du XX^e siècle, ce droit s'affine dans un paradigme libéral et s'inscrit comme droit fondamental au même titre que d'autres libertés fondamentales (liberté d'opinion, de circulation, politique, etc.). Au cours du XX^e siècle va émerger l'idée de la nécessité du contrôle par l'individu sur la circulation des informations qui le concernent : c'est la notion d'autodétermination informationnelle, qui va devenir le « socle des législations de protection des données personnelles élaborées à partir des années 1960 dans les pays occidentaux »¹.

Élargissement européen et obstacles

Cette loi nationale se développe pourtant dans un contexte de construction européenne. Cette dernière va faire émerger des problématiques² concernant les relations entre ses membres : les droits nationaux manquent parfois d'harmonisation pour correctement

1. CNIL, « Cahier IPB : Scènes de la vie numérique », avril 2021, <https://linc.cnil.fr/fr/cahier-ip8-scenes-de-la-vie-numerique-une-exploration-du-rapport-quotidien-la-protection-des-donnees>

2. OCDE, « En définissant les principes essentiels applicables dans ce domaine, les lignes directrices jouent un rôle important en aidant les gouvernements, les entreprises et les représentants des consommateurs à protéger la vie privée et les données personnelles, tout en évitant des restrictions inutiles aux flux transfrontières de données, en ligne ou hors ligne », 1980.

s'accorder. Le refus en 1989 par la CNIL de transférer des données de l'entreprise Fiat France à sa maison-mère italienne³ pour des raisons de protection des données des citoyens français, considérée comme « équivalente », est l'un des moteurs de l'adoption de la Directive 95/46/CE sur la protection des données personnelles de 1995⁴. La question de la protection des données se heurte ainsi à des considérations économiques pour les pays membres.

Ainsi, la directive de 1995 reprend les grandes lignes de la loi de 1978, mais aussi des principes d'autres pays, notamment en demandant l'institution d'une autorité de protection des données au niveau national, sur le modèle de la CNIL. Elle entraîne la création du Groupe de travail de l'article 29 (puisque c'est cet article qui instaure l'existence du groupe) sur la protection des données – appelé G29 –, qui est alors un organe indépendant et consultatif de l'UE sur la question de la protection des données.

Dès 2012, des négociations débutent autour d'une actualisation du cadre commun sur la protection des données, cette fois sous la forme d'un règlement, et non d'une directive⁵. On y retrouve les grands principes déjà présents dans la version de 1995, mais avec quelques nouveautés : l'intégration du critère de ciblage (le fait de traiter les données d'un résident dans l'UE oblige le responsable de traitement à appliquer la réglementation), le changement de fonctionnement des sanctions (20 millions d'euros et jusqu'à 4 % du CA mondial), ou encore la mise en place du principe de responsabilité (*accountability*). Le texte, intitulé « Règlement européen sur la protection des données personnelles » (ou RGPD) entre en vigueur dès 2016, et en application dans les pays de l'Union européenne le 25 mai 2018⁶.

Il est également intéressant de noter que, pendant le processus de révision de cette directive, a éclaté l'affaire Edward Snowden et ses révélations concernant l'usage des données à des fins de surveillance par l'État américain, mais également les révélations de Wikileaks concernant l'espionnage pratiqué par ce dernier envers ses alliés européens.

Concernant l'intelligence artificielle, le RGPD ne l'ignore pas (la régulation vise les « traitements » et non les données en elles-mêmes) Deux notions importantes émergent du texte : la question de la décision entièrement automatisée, et celle de l'explicabilité. En effet, le texte européen a pris la mesure de l'évolution des technologies : « Le RGPD prévoit des règles plus restrictives dans ces cas, pour éviter que l'homme ne subisse ces décisions émanant uniquement de machines. L'article 22 du RGPD encadre ainsi les processus de prise de décision entièrement automatisés, lorsqu'elles produisent des effets juridiques ou des effets significatifs »⁷. Les effets sont jugés significatifs lorsqu'ils concernent les droits et libertés de la personne, mais également s'ils provoquent des conséquences discriminatoires. Le texte prévoit aussi le droit à une intervention humaine pour réexaminer son cas, mais également pour obtenir une explication de la décision prise automatiquement.

3. Directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

4. <https://linc.cnil.fr/1970-2021-la-protection-des-donnees-essaime-le-monde> et https://linc.cnil.fr/sites/default/files/atoms/files/20171116_rapport_annuel_cnil_-_rapport_dactivite_1989_vd.pdf
Extrait : « Se pose donc un problème de flux transfrontière et la notion de protection équivalente entre en jeu. Or, l'Italie ne dispose d'aucune législation générale relative à la protection des données et n'a pas non plus ratifié la Convention du Conseil de l'Europe. »

5. « À la différence d'un règlement communautaire qui s'applique totalement et directement, une directive donne des objectifs à atteindre par les pays membres, avec un délai. » [Wikipédia]

6. Comme pour la directive de 1995, sont séparées les données personnelles utilisées dans le cadre des activités de police et de justice (qui donnera donc la directive Police Justice en 2018, avec le règlement européen à la protection des données).

7. CNIL, « Profilage et décision entièrement automatisée », 29 mai 2018, <https://www.cnil.fr/fr/profilage-et-decision-entierement-automatisee>

Ces questions « éthiques » dans la relation entre utilisation des données par des algorithmes ne sont pas nouvelles, et la CNIL s'en est saisie à plusieurs reprises, notamment en publiant en 2017 son rapport éthique, intitulé « Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle »⁸.

À noter également : les traitements de police-justice sont encadrés de manière spécifique par un texte européen différent (la directive police-justice⁹).

L'Intelligence artificielle dans le viseur

Vers un projet de régulation de l'IA

La démarche est la même concernant la question de l'intelligence artificielle. Depuis 2017, Commission et Parlement européens ont multiplié les réflexions sur le sujet, en commençant par la constitution du AI HLEG en 2018. Ce « High-Level Expert Group on Artificial Intelligence » est composé de 52 experts chargés de produire plusieurs livrables. A ainsi été défini le concept de « *trustworthy AI* » ou « IA de confiance », ses 7 exigences-clés pour l'encadrer (agence de contrôle, sécurité et robustesse technique, confidentialité et gouvernance des données, transparence, diversité, non-discrimination et équité, bien-être sociétal et environnemental, responsabilité), et son outil associé (ALTAI : *Assessment List for Trustworthy AI*), testé par plus de 350 organisations. Le mandat du HLEG AI s'est terminé à la fin de l'été 2020.

L'arrivée à la présidence de la Commission européenne d'Ursula von der Leyen en 2019 a par ailleurs accéléré le mouvement, avec le plan de lancement « 100 jours pour définir une approche européenne des implications humaines et éthiques de l'IA » en novembre de la même année.

Ainsi, en février 2020 a été publié le livre blanc de la Commission européenne sur l'IA, qui bénéficie de plus de 1 200 contributions externes de la part d'entreprises, de chercheurs, d'acteurs de la société civile, d'associations, d'autorités publiques, de particuliers, etc.

Enfin, en avril 2021, la Commission européenne a publié un plan de coordination sur l'IA incluant un projet de régulation. Les principales propositions peuvent être résumées ainsi :

- La proposition d'un cadre juridique sur l'IA aux règles harmonisées concernant la mise sur le marché, la mise en service et l'utilisation de systèmes d'IA dans l'UE.
- Une définition de l'IA large applicable aux acteurs à l'intérieur et à l'extérieur de l'UE, pour autant que le système d'IA soit mis sur le marché de l'UE ou que son utilisation affecte des personnes situées dans l'UE.

8. CNIL, « Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle », 15 décembre 2017, <https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de>

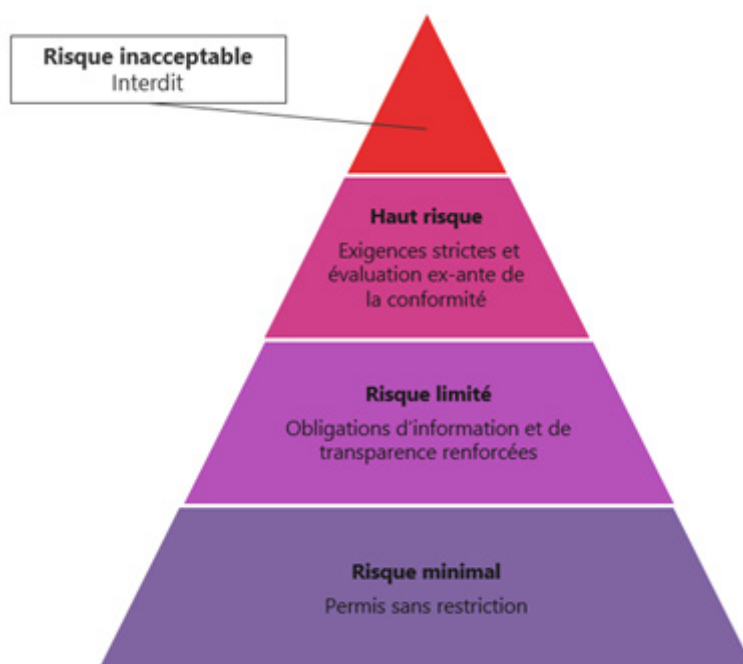
9. <https://www.cnil.fr/fr/directive-police-justice-de-quoi-parle-t>

- Des règles relatives à la surveillance du marché s'appliquant aux fournisseurs et aux utilisateurs de systèmes d'IA dans les secteurs public et privé.
- Une approche basée sur la classification des systèmes en fonction de leur niveau de risque.

La particularité de cette approche est qu'elle appréhende la technologie non pas de manière technique, mais à travers ses usages et ses applications. Le « risque » est celui qui pèse sur les populations, et ce sont des valeurs « éthiques » qui guident l'autorisation ou non d'un système d'IA.

- Des obligations strictes pour les systèmes d'IA à haut risque : systèmes d'évaluation et d'atténuation des risques, qualité élevée des données alimentant le système, traçabilité des résultats, documentation détaillée, informations claires à l'intention de l'utilisateur, contrôle humain, niveau élevé de robustesse, de sécurité et d'exactitude.
- Des mécanismes de certification (« marquage CE ») par des organismes notifiés.
- La création d'un comité européen de l'IA (CEIA) rassemblant la Commission européenne, l'EDPS et les autorités nationales compétentes.
- Des sanctions pécuniaires allant jusqu'à 30 M€ ou 6 % du CA, en plus des possibilités offertes par la surveillance du marché : actions correctrices, restriction, retrait, rappel, etc.

Des mesures visant à favoriser l'innovation (bacs à sable réglementaires¹⁰ notamment).



Pyramide des systèmes d'IA selon les risques liés à ses usages

10. Un bac à sable réglementaire est un cadre permettant à des technologies innovantes de tester leurs applications sans devoir respecter dans son intégralité la réglementation s'appliquant habituellement.

Ici, qu'il s'agisse du critère de ciblage, du modèle des sanctions, du principe de responsabilité (*accountability*) ou encore du marquage CE, il s'agit bien de reprendre des mécanismes éprouvés par le passé (notamment par le RGPD pour les deux premiers) et qui ont permis la diffusion d'un modèle normatif sur le marché mondialisé.

Les algorithmes, la régulation et les travailleurs des plateformes numériques

En France, l'émergence et la spécificité du cadre des travailleurs des plateformes numériques (ce que l'on nomme alors la *gig economy*¹¹) a entraîné une réaction juridique, à travers la création d'une nouvelle autorité administrative : l'Autorité des relations sociales des plateformes d'emploi (ARPE), en avril 2021. Et c'est notamment le recours aux algorithmes dans l'organisation du travail (management algorithmique) qui a justifié la mise en place de cette administration. Cette dernière sera chargée de la régulation des relations sociales (représentation, dialogue social), d'un observatoire sur l'activité des plateformes numériques, ou encore de l'indemnisation de ces travailleurs (compensation)¹². « [Ce dialogue social] pourra ainsi s'instaurer au niveau de deux secteurs d'activité : celui des activités de conduite d'une voiture de transport avec chauffeur (VTC), et celui des activités de livraison de marchandises à vélo, scooter ou tricycle, qui représentent près de 100 000 travailleurs indépendants. »

Dans le cadre du RGPD, par exemple, l'implémentation d'un nouveau droit – le droit à la portabilité des données (c'est-à-dire de pouvoir transmettre les données fournies à un service à son concurrent ou à un autre service similaire, afin de renforcer le pouvoir des individus dans leurs choix et de lutter contre des phénomènes d'enfermement) – a permis de faire émerger les revendications des travailleurs de ces plateformes numériques. Le fait de pouvoir récupérer ses données permet alors d'objectiver une « lutte » contre un management algorithmique opaque. Et encore plus dans un phénomène de précarisation : « L'OCDE estime qu'aujourd'hui, au sein de ses pays membres, les emplois « atypiques » (indépendants, temps partiel, intérimaires) représentent plus d'un emploi sur trois et sont, sur tous les plans sociaux, moins bien couverts, à commencer par les perspectives de retraites, un des enjeux à l'origine de la grève interprofessionnelle ayant lieu depuis décembre 2019 en France. » On pourrait également rajouter l'exemple d'un ancien chauffeur de ces plateformes qui a porté plainte pour licenciement abusif après la désactivation de son compte et ce, sans explication. Se pose donc la question du recours, en plus de celle de l'opacité générale, que ce soit dans l'encadrement des conditions de travail ou dans le fonctionnement des algorithmes. L'Italie a par ailleurs sanctionné le service de livraison Foodinho (Glovo) d'une amende de 2,6 millions d'euros pour ne pas avoir « expliqué ses algorithmes de gestion automatique des commandes, ne pas s'être assuré que l'évaluation automatique des performances des livreurs était correcte et ne pas avoir proposé de solution de contestation des mesures prises par les algorithmes à l'encontre des livreurs »¹³.

11. H. El Aoufir, « Données : poison et remède pour les travailleurs des plateformes ? », 17 février 2020, <https://linc.cnil.fr/fr/donnees-poisson-et-remede-pour-les-travailleurs-des-plateformes>

12. « Les travailleurs indépendants des plateformes de mobilité pourront désigner leurs représentants en 2022 », <https://travail-emploi.gouv.fr/actualites/presse/communiqués-de-presse/article/les-travailleurs-independants-des-plateformes-de-mobilite-pourront-designer>

13. « L'Italie condamne le service de livraison Foodinho (Glovo) à une amende de 2,6 millions d'euros », L'usine digitale, 06 juillet 2021, <https://www.usine-digitale.fr/article/l-italie-condamne-le-service-de-livraison-foodinho-glovo-a-une-amende-de-2-6-millions-d-euros.N1119024>

Autres travaux européens en cours

- **Le Digital Governance Act (DGA)** : développement du marché européen de la donnée en soutenant la réutilisation et le partage de certaines catégories de données : adaptation du cadre juridique européen le rendant plus incitatif et facilitateur pour la réutilisation des données concernées, y compris les données à caractère personnel. Ces dernières sont considérées comme source d'innovation. À noter également, la notion d'« altruisme des données » : mise à disposition de données pour le bien commun par des entreprises ou des particuliers.
- **Le règlement sur les services numériques (DSA)** : mise à jour du cadre juridique horizontal concernant les responsabilités et obligations des fournisseurs de services numériques dans l'Union européenne, et en particulier les plateformes numériques (transparence, système de réception et traitement de notifications, recours et suppression des contenus, audit et évaluation des risques systémiques liés à la modération, recommandation, gestion et publicité en ligne, etc.).
- Et d'autres, comme le **Data Act**.

Le pouvoir normatif de l'UE

« L'Europe [...], depuis 1957, s'est construite sur la norme et le projet politique qu'elle porte au plan mondial peut être qualifié de « projet normatif ». [...] L'Europe est porteuse d'une narration du monde, d'un *storytelling* planétaire qui privilégie l'interdépendance et la norme au détriment de la *realpolitik* et de la force »¹⁴.

Le rayonnement de la réglementation européenne sur la protection des données

Le RGPD a provoqué une forte médiatisation lors de son entrée en application, notamment concernant ses modalités et son application concrète par les organismes, qu'ils soient privés ou publics. Le critère de ciblage a permis d'étirer les standards européens aux entreprises étrangères (et qui n'avaient pas de présence en Europe, tout en traitant des données européennes), qui échappaient dans certains cas à cette législation sur les données.

Les projets d'adéquation se multiplient depuis l'entrée en vigueur du RGPD : le Japon a obtenu son adéquation en 2019, la Corée du Sud en 2021¹⁵, l'Angleterre post-Brexit, etc. D'autres lois sont en gestation dans d'autres pays, cherchant à reprendre les standards du RGPD. On peut citer le LGPD au Brésil, la loi californienne autour de la *privacy*, ou encore celle de l'Inde.

14. Zaki Laïdi, « La norme sans la force », 2009.

15. « RGPD : vers une décision d'adéquation de la Corée du Sud », NextINpact, 17 juin 2021, <https://www.nextinpact.com/lebrief/47465/rgpd-vers-decision-dadequation-coree-sud>

De la même manière, la Chine s'est dotée d'un règlement similaire au RGPD, avec cependant quelques différences assez notables. Tout d'abord, les traitements de données personnelles par les acteurs publics ne sont pas concernés par ce règlement. Le second point, qui découle du premier, est l'absence d'autorité indépendante chargée de veiller à son application : le contrôle sera fait par un service rattaché à un ministère, le State Internet Information Department.

Les précédents tendent à montrer que les standards européens, souvent élevés, sont suivis pour faciliter les transferts et continuer de gagner des parts de marché. L'exemple de Fiat France et la maison-mère italienne en témoigne pour la protection des données. L'adéquation d'autres grands pays dans d'autres parties du monde également (Japon et Corée du Sud). De la même manière, le fait que la Californie prennent un modèle similaire pourrait avoir un effet domino : « L'adoption en 2018 du California Consumer Privacy Act a signé le début d'un morcellement du paysage législatif étatsunien en matière de protection des données. Depuis, il a été suivi du Colorado Privacy Act et du Consumer Data Protection Act de l'État de Virginie, et d'autres textes devraient probablement suivre. » En effet, l'économie ne raffole pas de la coexistence au sein d'un même pays de standards trop différents.

« Avec le développement de l'économie de la donnée, ces considérations économiques revêtent une importance croissante et favorisent des changements de position des entreprises quant aux projets législatifs »¹⁶.

C'est par ailleurs un objectif assumé par la Commission : « Faciliter le développement d'un marché unique pour des applications d'IA légales, sûres et dignes de confiance, et empêcher la fragmentation du marché. »¹⁷

Les limites de cette stratégie normative

Le rayonnement de l'Europe est à relativiser : pour nombre d'États émergents, les standards du RGPD sont « disproportionnés » par rapport à leur intégration dans le pays, car les obligations sont quasiment impossibles à mettre en place, notamment par manque de moyens, voire d'infrastructures suffisamment interconnectées¹⁸.

D'autres hypothèses pourraient être le déclin ou l'attractivité moins forte du marché européen, qui fait sa puissance normative : la montée en puissance des marchés asiatiques et le recentrement sur le commerce intra-Asie, par exemple, pourrait favoriser le développement de nouvelles approches en termes d'encadrement des technologies, qui seraient bien différentes de celles existantes en Europe, et plus largement en Occident¹⁹.

Et, de manière générale, cette tendance semble se confirmer par le fait que certains pays (dont des pays asiatiques) choisissent d'avoir une adoption « partielle » des obligations de protection des données. En effet, dès 2012, le chercheur Graham Greenleaf montrait que

16. J. Saliou, « 1970-2021 : la protection des données essaime le monde », 18 octobre 2021, <https://linc.cnil.fr/1970-2021-la-protection-des-donnees-essaime-le-monde>

17. Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'union, avril 2021, <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52021PC0206>

18. Clarisse Girod, entretien avec le LINC, « La mise en œuvre peut s'avérer très complexe pour des pays qui intègrent en quelques années des décennies de construction normative européenne. Pour ces États, le RGPD est un standard « disproportionné » imposant de nombreuses obligations administratives : études d'impact, portabilité, certification, codes de conduite... que ni les entreprises ni les autorités n'ont les moyens de mettre en œuvre », 19 mai 2021, <https://linc.cnil.fr/le-modele-europeen-point-de-convergence-source-de-divergences>

19. Ibidem.

la Malaisie, le Viêtnam, l'Inde, le Qatar ou encore Dubaï, rejoints récemment par la Chine, avaient décidé d'exempter complètement le secteur public de cette réglementation²⁰.

De la même manière, la possibilité d'un RGPD un peu remanié, plus souple quant à la question des cookies et des échanges transfrontaliers, est à l'étude au Royaume-Uni, afin de fluidifier l'économie.

Parallèle avec la réglementation IA

Si le projet de réglementation a été plutôt salué (même si des critiques ont émergées sur les questions d'application, de manque de contrôle (notion de responsabilité), notamment pour les niveaux « acceptables », plusieurs critiques émanant d'acteurs importants dans le développement de ces technologies montrent qu'il y a de fortes dissensions. Ces dernières s'appuient notamment sur un contexte géopolitique qui pourrait faire échouer un consensus sur le sujet.

Les grandes plateformes et les grandes entreprises du numérique (Big Tech et autres ex-GAFAM) tendent à avoir des positions plus ambivalentes sur le sujet. Si elles estiment qu'il y a bien des utilisations de l'IA qui mériteraient d'être interdites²¹, certaines n'hésitent pas à critiquer ouvertement cette réglementation, donnant le ton de leurs propres orientations. L'ancien CEO de Google, Eric Schmidt, n'a pas mâché ses mots contre le projet de régulation européen, le qualifiant même de « désastre »²². Sa vision des choses est autre : l'IA n'étant pas assez mature pour être réglementée, il faut d'abord la nourrir et l'« inventer » avant de lui mettre des barrières. Son de cloche presque similaire du côté de Yann Le Cun, tête de file bien connue de la recherche en intelligence artificielle chez Facebook. Pour lui, il s'agit non pas de réguler une technologie, mais plutôt ses usages : le risque serait pour l'Europe de retarder la recherche sur l'IA, ce qui lui ferait prendre du retard dans la course à la technologie. Les chercheurs partiraient alors en Grande-Bretagne ou aux États-Unis, selon Eric Schmidt.

Ce dernier est par ailleurs le codirecteur d'une commission indépendante (National Security Commission on Artificial Intelligence), qui a publié un rapport en mars 2021²³ préconisant d'abord :

- D'augmenter considérablement les dépenses fédérales (hors défense) pour la recherche et le développement de l'IA en les doublant chaque année, de 2 milliards de dollars en 2022 à 32 milliards de dollars en 2026.
- La création d'un corps numérique (Digital Corps) afin d'intégrer des travailleurs qualifiés en technologie au sein du gouvernement.
- La création d'une Digital Service Academy : une université accréditée offrant une éducation subventionnée en échange d'un engagement à travailler un certain temps dans le gouvernement.

20. Ibid., « Dès 2012, Graham Greenleaf identifiait un groupe d'États asiatiques exemptant le secteur public des obligations de protection. Au rang de ce groupe, il comptait la Malaisie, le Viêtnam, l'Inde, le Qatar, ainsi que Dubaï. La Chine a également confirmé sa place en son sein avec l'adoption en août 2021 de la Personal Information Protection Law, les activités étatiques étant exclues du champ d'application. »

21. Paresch Dave and Jeffrey Dastin, « Money, mimicry and mind control: Big Tech slams ethics brakes on AI », Reuters, <https://www.reuters.com/technology/money-mimicry-mind-control-big-tech-slams-ethics-brakes-ai-2021-09-08/>

22. Eric Schmidt, « L'IA est trop récente pour la réglementer, il faut d'abord l'inventer », CBnews, <https://www.cbnews.fr/digital/image-eric-schmidt-ia-est-trop-recente-reglementer-il-faut-abord-inventer-62151>

23. <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>

- D'inclure des rapports sur les droits civils et les libertés civiles pour les nouveaux systèmes d'IA ou les mises à jour majeures des systèmes existants.
- D'étendre les allocations de cartes vertes basées sur l'emploi et les donner à tout titulaire d'un doctorat en IA d'une université américaine accréditée.
- De réformer le système de gestion des acquisitions du ministère de la Défense afin d'accélérer et de faciliter l'introduction de nouvelles technologies.

Cette commission indépendante oriente notamment la nécessité d'accélérer le mouvement pour l'IA pour une raison liée à la défense et à la sécurité nationale : « Nous devons gagner la compétition de l'IA qui intensifie la concurrence stratégique avec la Chine ». Son rapport cite l'Europe 81 fois, et la Chine 605 fois (sources comprises).

Enfin, le rôle de frein que peuvent jouer ces grands acteurs est aussi à prendre en compte, notamment à travers le lobbying. Les récentes révélations de documents internes lors d'une procédure judiciaire contre Google, au Texas, ont mis en lumière certains éléments de stratégie, et notamment des e-mails se félicitant d'avoir réussi à « ralentir et retarder » le règlement européen ePrivacy.

Des dissensions internes

Si le projet de régulation a été bien accepté par plusieurs commentateurs, certains points ont soulevé interrogations et critiques. En particulier, les aspects inhérents à la question du travail et la protection des travailleurs. L'European Trade Union Institute (ETUI) a publié sur le sujet un article soulignant l'inadéquation de ce projet au monde du travail, qui est un cadre particulier. Si cette question du travail est appréhendée par le projet, l'ETUI note un manque de prise en compte du travailleur, notamment en ce qui concerne les travailleurs des plateformes (*gig economy*). Les questions du recours et de la réparation ne sont pas évoquées, notamment dans le cas d'une atteinte (discrimination ou autre) sur l'individu et sur le travailleur.

De la même manière, une approche par les risques aussi large leur semble laisser trop de libertés aux systèmes d'IA considérés comme à « faible » risque ou « modéré » : il manque ainsi un contrôle s'exerçant sur ces IA (dont l'implémentation au travail devrait se faire via un dialogue social). Par ailleurs, l'ETUI critique le manque de diversité dans les acteurs ayant contribué à la consultation sur ce sujet, et notamment le manque de visibilité des partenaires sociaux sur cette question. Ainsi, ils plaident pour une déclinaison d'un projet de régulation de l'IA dans le monde du travail pour prendre en compte les responsabilités de l'employeur, pour gérer la question de la protection des données et de la vie privée au travail²⁴, l'explicabilité pour les travailleurs, pour s'assurer qu'il y ait un contrôle humain, et pour gérer la question de la surveillance au travail.

La Chine comme contre-exemple

Le modèle chinois d'usage des technologies numériques est souvent décrit, à l'instar du rapport de la commission américaine, comme un épouvantail, notamment par son développement extrêmement rapide, qu'il s'agisse de l'Internet chinois ou de ses grands

24. Par exemple, la question du consentement (base légale possible dans le RGPD) dans le cadre professionnel est problématique : il ne peut y avoir un consentement libre dans une relation de subordination entre employeur et employé.

groupes industriels et plateformes que sont les fameux BATX. Et, dans le domaine de l'intelligence artificielle, la Chine se positionne à la pointe des nouveaux usages, notamment dans le domaine de la surveillance (reconnaissance faciale, surveillance « intelligente » de l'espace public, etc.).

L'utilisation de ces technologies contre les Ouïghours montre une face plutôt sombre de l'utilisation de l'IA dans l'espace public et à des fins de surveillance. Encore plus quand les principales start-ups et programmes de recherche sont financés par l'État lui-même (à l'instar de Megvii, SenseTime, Cloudwalk et Yitu, les quatre licornes du marché chinois de la reconnaissance faciale²⁵).

Cette force de frappe chinoise lui permet de peser dans la « détermination de standards et de réglementations qui soient les plus favorables au développement de leur activité économique, dans le domaine de la sécurité ou de l'usage industriel de la reconnaissance faciale ». Et ce choix se fait dans une optique de concurrence directe avec les standards occidentaux dans les autres pays et organisations internationales : « Elles sont parmi les acteurs les plus actifs à proposer des standards dans les instances internationales, telles que l'Union Internationale des Télécommunications (ITU), qui font référence dans les pays d'Afrique, du Moyen-Orient et d'Asie (l'Europe et l'Amérique du Nord ont leurs propres organismes de standardisation, tels que l'IETF, l'IEEE ou le 3GPP, où les industriels occidentaux dominent les débats). »²⁶

Dans son ouvrage *Red Mirror*, l'auteure et journaliste Simone Perrianni explique qu'il n'y aurait pas tant de différences entre les États-Unis et la Chine, si ce n'est que la surveillance en Chine est bien plus présente dans les mains de l'État qu'elle ne l'est outre-Atlantique : « Ce que l'on appelle le capitalisme de surveillance rapproche de façon frappante les deux grandes puissances mondiales. La Chine et les États-Unis montrent la voie qui sera suivie par les autres pays. La différence entre les deux modèles est la suivante : dans notre monde, les données sont gérées par des entreprises qui les utilisent à des fins privées, alors qu'en Chine, c'est l'État qui détient les informations des citoyens. »²⁷ De la même manière, « brandir la menace chinoise contribue à orienter le débat sur l'angle de la concurrence et des politiques industrielles nationales et à éviter l'imposition de réglementations restrictives qui nuiraient aux entreprises nationales sur les marchés internationaux ». ²⁸

Tendances & Hypothèses

- La volonté de l'Europe en termes de rayonnement normatif, comptant continuer dans ce cadre (avec l'exemple du RGPD). Sa force réside notamment dans l'intérêt économique que représente le marché européen : si sa puissance s'effrite, notamment en termes d'attractivité (dissensions internes, montée d'autres marchés régionaux, etc.), quels effets pour ses réglementations ?

25. A. Courmont, « La « menace » des start-ups chinoises : une figure bienvenue dans les débats sur la reconnaissance faciale ? », 2 janvier 2020, <https://linc.cnil.fr/fr/la-menace-des-start-ups-chinoises-une-figure-bienvenue-dans-les-debats-sur-la-reconnaissance-faciale>

26. Ibid.

27. Simone Perrianni, « Red Mirror, L'avenir s'écrit en Chine », p. 37, C&F Editions, 2021.

28. A. Courmont, « La « menace » des start-ups chinoises : une figure bienvenue dans les débats sur la reconnaissance faciale ? », 2020.

- La concurrence entre la Chine et les États-Unis : l'éloignement de plus en plus important entre la Chine et les Occidentaux pourrait renforcer encore la guerre économique, notamment à travers des champions nationaux (ex-GAFAM contre BATX et Huawei). Quel impact sur les volontés de normalisation internationale ?

Hypothèse 1. Uniformisation des réglementations sur le modèle européen : un modèle qui fonctionne

La réglementation européenne, même si elle n'est pas reprise dans son intégralité, arrive à poser des standards élevés sur l'ensemble des technologies (à l'instar d'un marquage CE), et permet d'avoir une certaine uniformisation autour d'une IA « digne de confiance ». Il existe des divergences, notamment autour de l'application aux services publics. Concernant le monde du travail, dans cette hypothèse, les risques de l'IA en termes de surveillance et de respect des droits restent relativement bien encadrés et l'IA se développe sans entraver les droits des travailleurs.

Hypothèse 2. Vers une IA de confiance : le modèle COP (Conference of Parties) ?

Si une norme « en dur » a échoué, les pays se sont rassemblés autour d'une charte et d'une série d'engagements pour accompagner le développement de l'IA. Les engagements ne sont pas coercitifs, mais rassemblent l'ensemble des parties prenantes. Ils permettent de construire une feuille de route au global.

Hypothèse 3. Coquille vide : les tentatives de réglementation globale échouent, et les grandes solutions d'IA sont distribuées par les « champions » numériques

Dans la compétition géopolitique, les États-Unis ont largement préféré le soutien aux champions nationaux (GAFAM, ou AMAMA désormais) et ont accepté un cadre plus lâche de la réglementation : c'est l'autorégulation qui prime, avec un droit de regard de la part de l'État dans les programmes qui sont financés par lui. Les grandes solutions d'IA dans le monde du travail sont trustées par de grandes plateformes, dont la régulation se fait plutôt *a posteriori*. Ce sont donc leurs visions du travail et de son encadrement qui priment, avec des scandales à prévoir (sur le modèle de la régulation des conditions des travailleurs de la *gig economy*).

Hypothèse 4. Un modèle à part : les fournisseurs de solutions IA sont européens et pour les Européens (balkanisation)

Morcellement des réglementations sur le sujet : si les principes occidentaux ont réussi à infuser, les autres pays ont opté pour une sélection de ce qui est proposé dans les cadres réglementaires fournis par l'UE. D'autres pôles, notamment asiatiques et moyen-orientaux, émergent avec des marchés régionaux de plus en plus importants, où les standards européens ne priment pas. Au sein de l'Union, se développent des technologies spécifiquement pour le marché unique et correspondant à des critères spécifiques.

Acceptabilité de l'IA dans le monde du travail : employeurs/salariés

S. Halluin

Cette fiche vise à traiter des questions que l'introduction d'IA dans l'entreprise, et dans la gestion de la S&ST en particulier, peuvent poser aux acteurs, en séparant le traitement du point de vue des employeurs de celui des salariés.

Dans les deux cas, il s'agira d'explorer les réactions possibles face à des questions telles que la compréhension de l'algorithme (comportement de l'IA), le monitoring des lieux ou des situations de travail, l'asservissement, la dépendance...

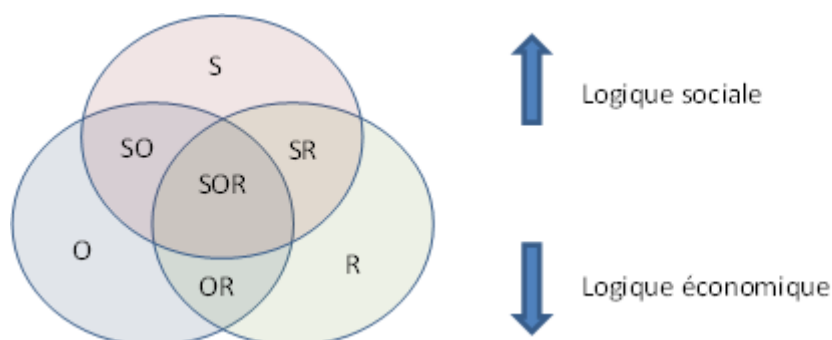
Contexte

D'après de nombreux experts, l'intelligence artificielle (IA) est susceptible de modifier profondément le travail dans les années à venir. Cette capacité transformative est supposée s'accroître au cours du temps, au fur et à mesure que l'IA pourrait passer du statut de technique permettant d'améliorer certaines phases du travail dans un cadre préalablement défini sans son concours, à celui d'élément central autour duquel l'ensemble de la production s'organiserait.

L'utilisation à des niveaux significatifs de l'IA en milieu de travail est encore trop récente pour disposer d'un recul suffisant permettant d'évaluer au fil des années les évolutions de l'opinion des employeurs et des salariés à son égard. On dispose uniquement de quelques études ponctuelles. On notera aussi que la plupart d'entre elles accordent un intérêt particulier à l'utilisation de l'IA dans le domaine des ressources humaines (RH), ce qui peut s'expliquer par l'avance relative de ce secteur dans l'utilisation des algorithmes, mais aussi par le fait que, dans d'autres activités (la production industrielle par exemple, à travers les applications robotisées), il est plus difficile d'« identifier » ce qui ressort de l'IA et ce qui correspond à d'autres techniques (mécanique, automatique, électronique...).

Objectifs de l'IA dans le cadre du travail

- **Soulagement** : diminuer l'exposition aux risques physiques et mentaux des salariés
- **Optimisation** : améliorer la productivité et augmenter les capacités des utilisateurs
- **Remplacement** : diminuer le coût du travail en substituant un salarié par de l'IA



La logique sociale tend vers le soulagement, tandis que la logique économique tend vers l'optimisation et le remplacement.

Exemples :

- Remplacement : *chatbot*, maintenance en milieu fermé, contrôle qualité, tâches sous environnement dangereux
- Optimisation : calculs de rendement (trajets, production...), planification de l'activité, traducteurs, gestion des stocks, expert, suivi médical
- Soulagement : EPI connectés
- Optimisation + Remplacement : démarchage, publicité ciblée, aide à la codification Cramif, maintenance prédictive, exosquelettes
- Soulagement + Optimisation : analyseur de mails, réassortiment
- Soulagement + Remplacement : robots de maintenance
- Soulagement + Optimisation + Remplacement : entrepôt automatique

Critères d'acceptabilité de l'IA aujourd'hui

I. Oracle et Future Workplace ont réalisé une enquête sur ce sujet dans dix pays (dont la France) auprès de 8 370 personnes (1 350 Français), âgées entre 18 et 74 ans. Les répondants appartenaient au monde de l'entreprise : environ un quart au titre de managers, autant au titre d'employés des ressources humaines, et une moitié en tant qu'employés.

Comme on peut le voir sur la figure ci-dessous, globalement, l'IA bénéficie d'un préjugé favorable auprès de l'ensemble de la population. C'est sa capacité à libérer du temps pour les répondants qui est d'abord mise en avant, suivie par la possibilité d'acquérir de nouvelles compétences.

What opportunities do you think will be created for you through using AI?

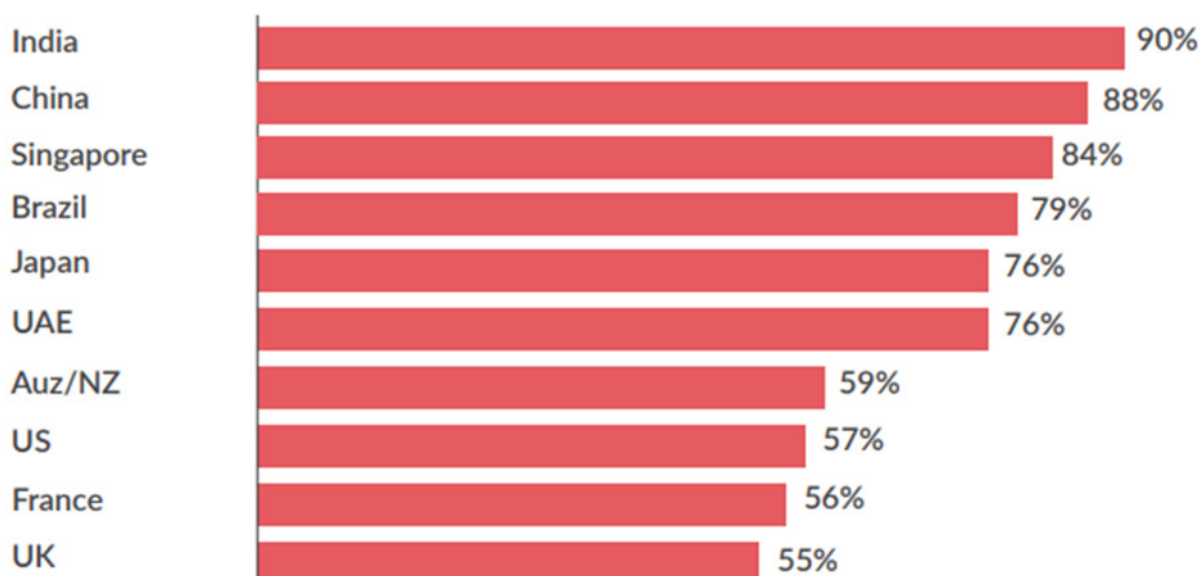


Il convient néanmoins de nuancer ce propos en fonction du statut des répondants. Les responsables de RH ont une vision clairement optimiste du déploiement de l'IA pour 38 % d'entre eux, alors qu'il n'est que de 31 % pour les autres encadrants, et seulement de 19 % pour les employés.

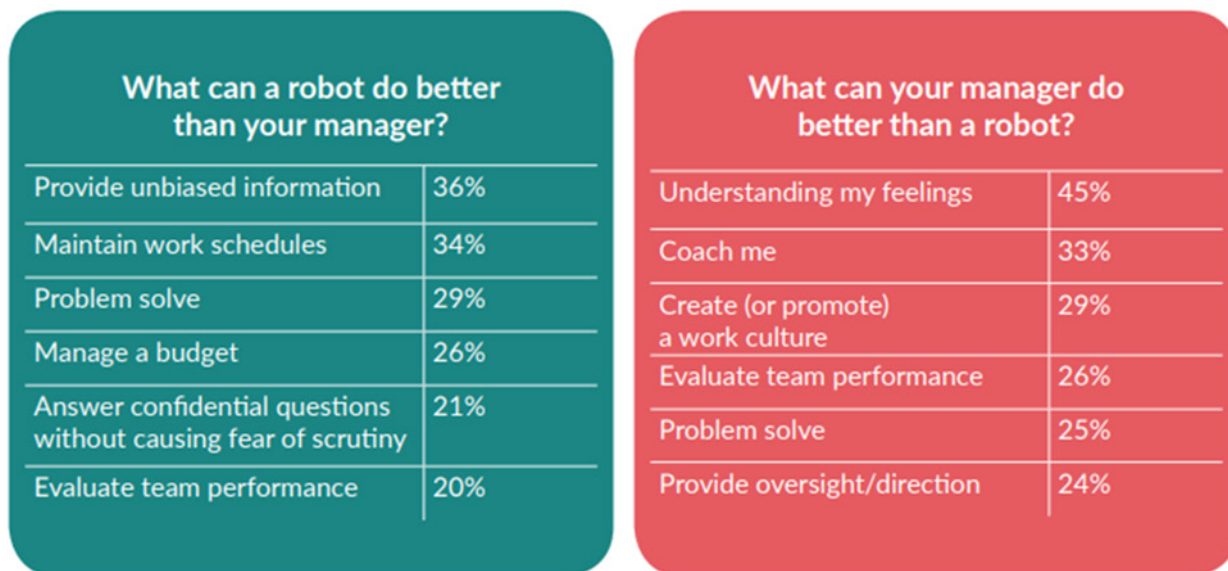
La stratification par pays est encore plus significative : la marge est grande entre les 60 et 56 % de personnes « enthousiastes » respectivement en Inde et en Chine, et les 8 et 20 % enregistrés pour la France et le Royaume-Uni. De même pour l'âge : les millénials (génération Y) sont « enthousiastes » à 31 %, suivis des générations Z et X avec respectivement des scores de 24 et 22 %. Ce sont les *baby boomers* qui ferment la marche avec seulement 14 %.

Même s'il existe, comme dans le cas précédent, des différences importantes entre les pays, le niveau de confiance envers les robots (assimilés à l'IA) se situe à un niveau élevé, comme la figure suivante le met en évidence.

People now trust robots more than their managers.



Pour autant, il y a encore des tâches pour lesquelles un manager est jugé supérieur à un robot : sans surprise, celles qui sont principalement fondées sur l'échange interpersonnel. En revanche, pour la résolution de problèmes techniques ou des évaluations, les performances des deux sont jugées plus ou moins équivalentes. Les sondés en déduisent que le rôle de l'encadrement va devoir évoluer fortement. Des indicateurs de ces tendances apparaissent.



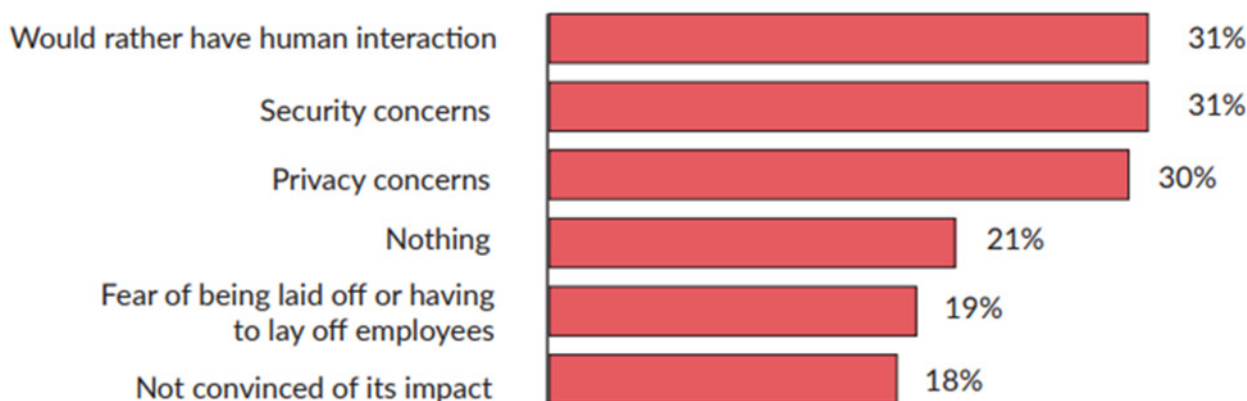
Les auteurs de l'étude analysent favorablement ces données sur la base du constat que la moitié du temps de l'encadrement est consacrée à des tâches routinières de production ou de compilation de données, qui pourraient être avantageusement effectuées par des techniques issues de l'IA, libérant du temps pour les encadrants pour la réalisation de tâches plus conceptuelles ou les relations humaines.

L'hypothèse, faite en introduction, que ce sont d'abord les aspects algorithmiques de l'IA plutôt que sa contribution au développement d'automates ou de robots industriels qui sont identifiés par les répondants, se vérifie à travers les réponses faites à une question consacrée aux principaux usages des techniques d'IA. En effet, les six premières réponses à la question des applications les plus courantes sont les suivantes :

- Recueil de données concernant les employés et les clients
- Développement de logiciels pour la formation
- Gestion de la relation client
- Fonctionnement des *chatbots* et autres assistances en ligne
- Recrutements mettant en œuvre des algorithmes informatiques
- Gestion prévisionnelle de l'emploi

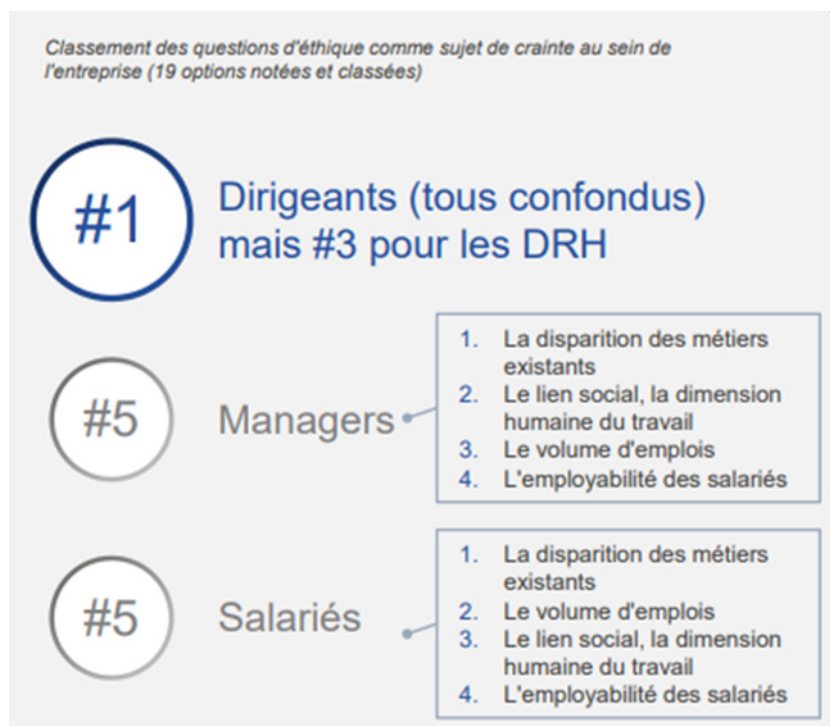
Parmi les paramètres qui constituent un frein à l'utilisation de l'IA, les questions de sécurité et de confidentialité des données se situent au premier rang et s'accompagnent d'une certaine dose de technophobie, ainsi qu'on peut le voir sur l'infographie ci-contre.

What is preventing you from using AI at work?



Une deuxième étude, intitulée « Intelligence Artificielle et capital humain en entreprise », a été effectuée à la demande de Malakoff Humanis par le Boston Consulting Group. Ce travail est assez fortement focalisé sur l'utilisation de l'IA dans les fonctions RH, plutôt que dans l'activité générale des entreprises. Cela est en grande partie dû au fait que, parmi les 321 dirigeants d'entreprises interviewés, c'est cette fonction qui était majoritaire. Ont également été sollicités 551 managers et 951 salariés.

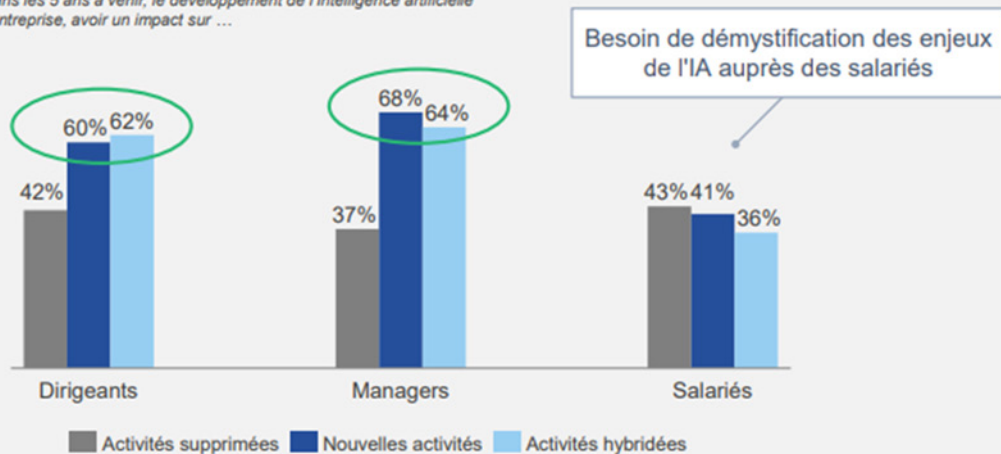
Si les questions d'éthique se situent au premier rang des préoccupations des dirigeants, ce sont d'abord les questions d'emplois (disparition des métiers existants, volume d'emplois, employabilité des salariés), ainsi que le lien social et la dimension humaine du travail, qui retiennent l'attention de l'encadrement et des salariés. Ces résultats apparaissent sur la figure ci-dessous.



Un accord général se fait pour considérer que le développement de l'IA va se traduire à la fois par une hybridation des activités existant actuellement, mais aussi par l'apparition de nouveaux métiers. Cependant, ces aspects sont beaucoup plus mis en avant par les directions et l'encadrement que par les salariés. En revanche, les préoccupations en termes de destructions possibles d'emplois se situent quasiment au même niveau dans les trois catégories, comme on peut le voir ci-dessous.

L'hybridation et la création de nouvelles activités sont les impacts les plus anticipés dans les 5 ans à venir

De votre point de vue, dans les 5 ans à venir, le développement de l'Intelligence artificielle va-t-il, au sein de votre entreprise, avoir un impact sur ...



Si la légitimité de l'utilisation de l'IA en matière de ressources humaines est évidente pour la quasi-totalité des dirigeants et de l'encadrement, l'adhésion des salariés est beaucoup moins massive. Ce phénomène apparaît clairement dans les réponses à différentes questions rassemblées dans l'infographie ci-après. Les enjeux évoqués dans le titre sont du domaine des RH et sont explicités dans les réponses aux cinq questionnements.

Dirigeants et managers considèrent les DRH légitimes sur ces enjeux

Sélection de questions illustratives

Face aux enjeux liés à l'IA, diriez-vous que c'est le rôle des Ressources humaines de conduire les missions suivantes ?



En matière de santé et sécurité au travail, on note une progression faible, mais significative, sur les trois catégories de l'idée que le développement de l'IA peut être bénéfique. Cependant, si cette conviction est fortement répandue chez les dirigeants et managers, respectivement à 70 % (+ 4 % par rapport à l'enquête précédente) et 73 % (+ 3 %), elle n'est cependant partagée que par 38 % (+ 3 %) des salariés. D'après les personnes interrogées, ce sont la prévention et les actions de santé individuelle et collective, notamment grâce à l'émergence des objets connectés d'une part, et la sécurité et les risques d'accidents d'autre part, qui devraient le plus bénéficier du développement de l'IA.

L'UGICT-CGT a mobilisé ses adhérents (cadres et techniciens) pour mener une réflexion sur l'utilisation de l'IA dans les entreprises. Quoique de façon moins délibérée que dans les sondages précédents, cette réflexion réserve aussi une large place aux questions relatives aux RH, ce qui tend à confirmer l'hypothèse émise précédemment que c'est dans ce secteur que les solutions IA sont le mieux identifiées.

L'IA est identifiée comme pouvant favoriser la libération de l'innovation et de la créativité des professionnels. À ce titre, elle est considérée comme un outil important dans une logique d'émancipation au travail, en particulier si elle s'inscrit dans une perspective de progrès social et environnemental.

C'est à la société de décider ce qu'elle attend de l'utilisation de l'IA.

Il est donc important qu'un contrôle puisse être effectué sur ses utilisations, et il apparaît nécessaire d'encadrer les pratiques professionnelles de tous les corps de métiers intervenant sur les systèmes IA : « L'auditabilité des algorithmes ne doit pas trouver sur sa route « le secret des affaires », qui empêcherait au nom de la propriété intellectuelle et industrielle la transmission des informations pertinentes pour expliciter les choix qui prévalent dans la restitution des résultats. »

On est donc confronté à une responsabilité du concepteur et de l'utilisateur, et cela implique des obligations de certification, de qualification et d'explication (au sens humain pour cette dernière).

« Les outils de gouvernance – regard sur la conception, contrôle *a posteriori* de l'intégrité des algorithmes – doivent permettre d'assurer l'éthique des usages de l'IA. Un tiers de confiance habilité, sous la forme d'une autorité indépendante de conformité des algorithmes, doit pouvoir veiller au respect de ces recommandations en assurant la transparence via l'*open-source* des algorithmes. »

« La création d'un Comité d'audit des technologies numériques et de l'IA doit inclure la participation des représentants des salariés. »

Plus loin, l'IA est considérée comme un outil au service de la justice sociale et environnementale : « Pour cela, les nouvelles technologies, dont l'IA, doivent être mobilisées pour assurer la traçabilité des conditions sociales et environnementales de fabrication des services et des produits manufacturés. Le traçage et le suivi numérique du début du processus de fabrication (matière première) à la livraison au client ou à l'utilisateur sont possibles notamment en utilisant la technologie *blockchain* et les étiquettes RFID. »

En ce qui concerne l'utilisation de l'IA dans le domaine des RH, les rédacteurs insistent sur le risque d'émergence de dispositifs standardisés, équivalents des progiciels de gestion RH, qui présentent le risque d'éluder les vraies questions devant être traitées au cours des différents contacts qu'un salarié a avec sa hiérarchie et les services de ressources

humaines tout au long de sa carrière. L'IA ne doit pas être un outil de formatage, mais répondre aux besoins réels des entreprises, dans toutes leurs composantes humaines.

Les partenaires sociaux européens (patronat et syndicats) sont parvenus à un accord sur la digitalisation, qui ne se réduit évidemment pas à l'IA, mais qui partage beaucoup de promesses et contraintes d'usage avec elle.

Une partie est plus spécifiquement consacrée à l'IA. Elle pose un certain nombre de principes :

- Affirmation de la nécessité de garantir un contrôle humain sur les machines comme sur l'IA, ainsi que leur conformité en termes de santé et sécurité.
- Affirmation de trois principes pour une IA de confiance :
 - Elle doit être équitable, transparente, « *safe and secure* », respectueuse des lois en vigueur et des droits fondamentaux ;
 - Elle doit être éthique et égalitaire, notamment respectueuse des droits fondamentaux tels que définis dans l'UE ;
 - Elle doit être robuste et soutenable, aussi bien sur le plan technique que social, notamment éviter les dommages non-intentionnels.
- Affirmation que les partenaires sociaux doivent s'impliquer sur les usages possibles de l'IA, y compris au niveau de l'entreprise, et notamment dans une optique d'amélioration des conditions de travail. Pour ce faire, il leur faudra notamment s'intéresser aux tensions potentielles entre autonomie, protection de la santé et explicabilité des décisions.
- Affirmation que le développement des systèmes d'IA doit :
 - Respecter le principe du contrôle humain ;
 - Être sûr, ce qui implique une évaluation des risques aussi bien physiques que psychosociaux (en tenant compte des biais possibles et de la charge cognitive) ;
 - Respecter des principes d'équité ;
 - Garantir la transparence et l'explicabilité de manière proportionnée à la gravité des conséquences possibles, avec des procédures de vérification.

Tendances

- L'acceptation de l'IA dans l'entreprise est plus grande chez les responsables des entreprises et des ressources humaines que chez les salariés. La disparition des métiers existants est la première crainte chez les salariés.
- Les nouvelles générations prenant de l'ampleur sur le marché du travail pour les quinze prochaines années sont moins enthousiastes vis-à-vis de l'IA que leurs prédécesseurs.
- L'acceptabilité de l'IA chez les salariés est bien plus grande en Asie et dans les pays émergents qu'en Occident, où les questions d'éthique restent aujourd'hui prédominantes.

- L'explicabilité de l'IA contribue à son acceptabilité lorsqu'il s'agit de diriger certaines tâches opérationnelles.
- La connaissance fine du fonctionnement de l'IA est pour l'instant réservée à un petit nombre, ce qui rend difficile l'appropriation des enjeux.

Incertitudes majeures

- La création d'un encadrement global de l'IA par exemple à travers des comités d'audit indépendants des technologies numériques et de l'IA, dans les entreprises et/ou des institutions intermédiaires.
- La capacité des acteurs (salariés comme dirigeants et RH) et de leurs représentants à comprendre les enjeux, expertiser les algorithmes et négocier les conditions d'utilisation de l'IA.
- La gestion et la traçabilité des données, capacité d'identification et d'opposition, etc.
- L'emprise des technologies étrangères dans les entreprises, notamment celles issues de pays où les exigences éthiques sont moins prégnantes.
- Évolution de l'explicabilité de l'IA pour faciliter son acceptation.
- Focalisation des partenaires sociaux sur les sujets sensibles (comme les biais de recrutement, données biométriques, ou encore la supposée perte d'emplois) au détriment d'autres questions, comme les usages de l'IA dans les processus de fabrication.
- Paramètres extérieurs à l'acceptabilité dans les entreprises et pour les individus :
 - Coût du travail : si le coût du travail reste élevé, il accélérera le déploiement de l'IA. Au contraire, si le poids des charges sociales diminue, l'humain restera compétitif par rapport à l'IA.
 - Mise à disposition de la main-d'œuvre : les emplois peu qualifiés et peu concernés par l'IA sont ou non pourvus par des chômeurs ou via l'immigration.
 - Progression des capacités de l'IA : la technologie et les algorithmes progressent significativement ou, au contraire, l'hiver de l'IA s'installe.
 - Protection sociale : la France confirme sa spécificité sociale (protection des salariés élevée, prédominance des CDI, droits au chômage...).

Hypothèses

Hypothèse 1. L'IA acceptée de tous

Plus explicable grâce à des évolutions technologiques, contrôlée et basée sur des données plus traçables, l'IA devient acceptable pour la grande majorité des salariés. Un grand travail de pédagogie et de montée en compétence a été nécessaire, tout comme la mise en place d'instances de contrôle efficaces, dont la saisine par les salariés et les arbitrages judiciaires ont permis d'installer la confiance.

Hypothèse 2. L'IA au cœur des débats dans les entreprises

L'intérêt de l'IA progresse aux yeux des dirigeants, managers et RH, mais elle reste mal acceptée par les salariés. Elle se déploie progressivement pour différents usages malgré cette incompréhension et malgré le manque d'instances de contrôle. Cette situation crée des tensions difficiles à résoudre, où parfois l'IA est prise comme un tout et est assimilée à d'autres problèmes sociaux sources de conflits dans les entreprises.

Hypothèse 3. L'IA rejetée par les forces vives

La crainte que les jeunes générations expriment à l'égard de l'IA a amplifié les doutes préexistants chez les salariés. Avec la montée en puissance de ces nouvelles forces vives sur le marché du travail, suite à divers scandales et avec un imaginaire culturel de plus en plus critique à l'égard de l'implantation de l'IA, les entreprises ont progressivement abandonné le recours trop avancé à l'IA, mal compris et non souhaité. Cela provoque des difficultés concurrentielles face aux entreprises d'autres régions du monde, où l'IA est nettement mieux acceptée et par conséquent implantée.

Hypothèse 4. L'IA s'installe par nécessité malgré une forte défiance

La défiance à l'égard de l'IA chez les salariés se renforce. Les technologies issues d'autres régions du monde deviennent toutefois prépondérantes et s'installent malgré tout dans les entreprises par nécessité. Les instances de contrôle étant inefficaces face à ces outils pilotés depuis l'étranger, la lassitude s'installe et tous les acteurs se résignent à adopter l'IA malgré leurs réticences.

Les questions éthiques sur l'acceptabilité en matière de prévention de l'usage de l'IA dans le monde du travail

A. Olympio

Cette fiche vise à interroger les limites éthiques aux usages possibles de l'IA en matière de santé et sécurité au travail (S&ST). Quelles sont les freins possibles ? Quel encadrement éthique de l'intelligence artificielle ? Qu'est-ce qui pourrait être accepté ou non ? À quelle condition ?

Préambule

Les usages de l'intelligence artificielle ont transformé, transforment et continueront de transformer le monde du travail. Ces innovations technologiques, principalement basées sur l'exploitation de données massives – structurées ou non structurées – par des algorithmes évolués touchent plusieurs maillons de la chaîne de création de valeurs. Les géants de l'économie digitale ont fondé leur modèle d'affaires sur l'usage de l'intelligence artificielle afin d'assurer leurs succès. En effet, elles soutiennent les activités de développement des entreprises en apportant des solutions automatiques pour améliorer le ciblage des clients et prospects, proposer des produits et services innovants, faciliter les interactions entre l'organisation et ses parties prenantes. Elles permettent également aux entreprises d'optimiser leurs processus opérationnels notamment en réduisant certaines tâches pénibles et répétitives et l'exposition à certains risques pour la santé au travail, en améliorant la sécurité et la protection des salariés dans l'exercice de leurs activités au quotidien. Leurs usages possibles pour soutenir la prévention et en matière de santé et de sécurité dans le monde du travail sont de plus en plus nombreux. Malgré cela, même si l'usage de l'IA se démocratise progressivement dans les entreprises et ce, dans tous les secteurs d'activité de l'économie, il existe de réelles préoccupations d'ordre éthique et d'acceptabilité de ces nouvelles technologies, telles que les biais algorithmiques (de discrimination et/ou d'exclusion), la protection des données personnelles, le risque d'usage abusifs de ces technologies (surveillance, cyberrisque...), le risque du remplacement des humains par les automates et de perte d'emploi, facteurs d'inquiétude et de stress au travail. Au vu de la place qu'occupe et occupera la technologie dans nos vies, il est plus que crucial de s'interroger sur l'approche éthique permettant d'encadrer son usage au

regard des considérations de la société sur la notion du bien et du mal, du juste et de l'injuste, de l'équitable et de l'inéquitable. Ainsi, ces dernières années, si plusieurs acteurs (publics comme privés) réfléchissent sur l'encadrement éthique de l'IA de manière générale, encore très peu de travaux et de propositions sont produits sur les limites éthiques que posent les usages possibles de l'IA en matière de santé et sécurité au travail.

Définitions et concepts

Les problématiques soulevées par l'intelligence artificielle sont imposantes, puisque l'intelligence artificielle est de plus en plus utilisée : dans les téléphones portables, les salles de marchés, les robots, les objets intelligents, la ville intelligente, la médecine, la justice, l'agriculture, la police et l'armée, les services, et progressivement dans l'industrie...

Dans ce contexte global, concernant les limites éthiques de l'usage de l'IA, quelles sont les notions structurantes permettant d'appréhender cette variable ?

Les principes éthiques de l'IA

Dans la philosophie antique, la notion d'éthique est un concept lié à la morale, avec l'idée de « ce que l'on doit faire ». Autrement dit, les normes. Contrairement aux normes juridiques, les normes éthiques sont de nature implicite.

En 1983, la création du Comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE) fait de l'éthique une norme « éclairceuse du droit ». Ce comité a pour mission d'« éclairer les progrès de la science, soulever des enjeux de société nouveaux et poser un regard éthique sur ces évolutions »¹. La définition des normes éthiques par le CCNE constitue une préfiguration des normes juridiques, selon la Commission nationale de l'informatique et des libertés (CNIL)². En France, ce sont ces deux organismes qui ont été principalement sollicités afin d'explorer les enjeux éthiques des avancées technologiques et leurs incertitudes.

En 2019, le gouvernement français a chargé le CCNE de constituer un comité pilote d'éthique du numérique. Ce comité a pour mission de « remettre de premières contributions sur l'éthique du numérique et de l'intelligence artificielle et de déterminer les équilibres pertinents pour l'organisation du débat sur l'éthique des sciences et technologies du numérique et de l'intelligence artificielle »³. En 2021, le comité a rendu un avis sur les problématiques éthiques posées par les agents conversationnels (*chatbots*). Dans cet avis⁴, le comité identifie les trois principaux enjeux éthiques suivants : le brouillage des frontières entre une machine et un être humain ; l'imitation du langage et des émotions par les *chatbots* ; la capacité des agents conversationnels à manipuler les interlocuteurs humains. Treize préconisations spécifiques aux agents conversationnels sont formulées à l'attention des autorités publiques, afin que celles-ci puissent émettre des principes éthiques veillant à la protection des individus et de leurs biens contre toutes les formes de distorsion de biais et discriminations pouvant altérer les décisions

1. [CCNE](#)

2. [CNIL](#)

3. <https://www.ccne-ethique.fr/node/351?taxo=0>

4. <https://www.ccne-ethique.fr/node/464?taxo=0>

prises par l'usage des algorithmes d'IA.

Les principes éthiques devraient permettre de légitimer l'usage des IA dans une certaine mesure, en garantissant sa conformité éthique et en instaurant un cadre de confiance qui contribue à une meilleure compréhension des décisions prises.

Les biais posant des problèmes éthiques à l'usage de l'IA

Les algorithmes permettent une automatisation des process et peuvent proposer des prédictions pour l'aide à la décision. Ces avantages conduisent également à certains défis, comme la présence de biais dans les classifications et les prédictions issues de ces algorithmes.

Il existe deux types de biais : le biais d'erreur d'estimation d'un modèle d'IA et le biais introduit dans le modèle pour discriminer de manière volontaire ou non. Le biais comme erreur d'estimation est l'une des deux erreurs utilisées pour définir la qualité d'un algorithme d'apprentissage. On le définit aussi parfois comme la « distance » entre le meilleur modèle pouvant être appris par l'algorithme et le vrai modèle. De son côté, le biais de discrimination correspond à un préjugé qui avantage ou désavantage une personne ou un groupe, ou à quelque chose considéré comme inéquitable. Les biais algorithmiques peuvent donc provenir d'un biais dans la sélection des données ou d'un biais dans le traitement des données. Dans les deux cas, c'est l'usage du modèle qui entraînera des questions sur la fiabilité des décisions.

Les humains peuvent également amplifier involontairement et de manière inconsciente (ou biais implicite) les biais des modèles d'apprentissage automatique. Le problème des biais de données et des biais de modèles d'apprentissage est tel, que la plupart des leaders en IA proposent des outils de détection. Pour ce faire, il faut commencer par analyser les données d'apprentissage.

En matière d'IA, voici une liste des principaux biais et discriminations pouvant poser des questions éthiques.

Le biais d'échantillonnage

Un jeu de données peut ne pas représenter l'intégralité du périmètre du problème (exemple de l'apprentissage d'IA d'un véhicule autonome avec uniquement des données de jour).

Les préjugés

Un jeu de données peut également incorporer des données dont l'utilisation pourrait donner lieu à des discriminations (par exemple, la race ou le genre d'une personne).

Le biais d'exclusion

Étant donné que les données sont généralement nettoyées avant d'être utilisées pour l'apprentissage ou le test d'un modèle d'IA, ce type de biais peut se produire par exemple lorsqu'on supprime des colonnes que l'on juge non pertinentes.

Le biais de mesure

Un biais de mesure a lieu lorsque les données collectées pour l'apprentissage diffèrent de celles utilisées pendant la production. Cela peut se produire par exemple lorsqu'un jeu de données de type photo/vidéo est collecté avec un type de caméra spécifique, alors que les données de production proviennent d'une caméra présentant des caractéristiques différentes.

Le biais algorithmique

Ce dernier ne provient pas des données à partir desquelles un modèle a été construit, mais du modèle d'apprentissage automatique lui-même. Cela inclut la façon dont le modèle a été développé ou entraîné, qui produit des résultats biaisés.

Principes éthiques et valeurs de l'IA

Concernant l'intelligence artificielle, les enjeux éthiques sont identifiés par la commission mondiale d'éthique des connaissances scientifiques et des technologies de l'UNESCO. Nous pouvons citer entre autres :

- Le manque de transparence des outils d'IA
- La vie privée et la surveillance par la collecte de données personnelles
- La propriété des données
- La partialité et la discrimination
- La manipulation de l'information et la confiance
- L'impact environnemental lié à sa consommation d'énergie

Pour tenter de répondre à cela, les principaux acteurs de l'IA ont adopté, en 2017, le premier guide de référence pour un développement éthique de l'intelligence artificielle, connu sous le nom des « 23 principes d'Asilomar »⁵. Parmi les principes définis, certains sont dédiés à l'éthique et aux valeurs de l'IA. Les 13 principes concernés sont les suivants :

- **Sécurité** : les systèmes d'IA doivent être sûrs et sécurisés tout au long de leur durée de vie opérationnelle et ce, de manière vérifiable lorsque cela est possible et faisable.
- **Transparence des défaillances** : si un système d'IA cause un préjudice, il devrait être possible d'en déterminer la raison.
- **Transparence judiciaire** : toute implication d'un système autonome dans la prise de décision judiciaire doit fournir une explication satisfaisante et vérifiable par une autorité humaine compétente.
- **Responsabilité** : les concepteurs et les constructeurs de systèmes d'IA avancés sont parties prenantes des implications morales de leur utilisation, de leur mauvaise utilisation et de leurs actions, avec la responsabilité et la possibilité de façonner ces implications.

5. <https://futureoflife.org/ai-principles/>

- **Alignement des valeurs** : les systèmes d'IA hautement autonomes doivent être conçus de manière à ce que leurs objectifs et leurs comportements puissent être assurés de s'aligner sur les valeurs humaines tout au long de leur fonctionnement.
- **Valeurs humaines** : les systèmes d'IA devraient être conçus et exploités de manière à être compatibles avec les idéaux de dignité humaine, de droits, de libertés et de diversité culturelle.
- **Vie privée** : les personnes devraient avoir le droit d'accéder, de gérer et de contrôler les données qu'elles génèrent, étant donné le pouvoir qu'ont les systèmes d'IA d'analyser et d'utiliser ces données.
- **Liberté et vie privée** : l'application de l'IA aux données personnelles ne doit pas restreindre de manière déraisonnable la liberté réelle ou perçue des personnes.
- **Bénéfices partagés** : les technologies de l'IA doivent bénéficier au plus grand nombre de personnes possible et les rendre autonomes.
- **Prospérité partagée** : la prospérité économique créée par l'IA doit être largement partagée, au bénéfice de l'ensemble de l'humanité.
- **Contrôle humain** : les humains devraient choisir si et comment déléguer des décisions aux systèmes d'IA, afin d'atteindre des objectifs choisis par les humains.
- **Non-subversion** : le pouvoir conféré par le contrôle de systèmes d'IA hautement avancés devrait respecter et améliorer (plutôt que subvertir) les processus sociaux et civiques dont dépend la santé de la société.
- **Course à l'armement de l'IA** : une course à l'armement en matière d'armes autonomes létales devrait être évitée.

En réalité, il existe d'autres déclarations de ce type, par exemple la Déclaration de Montréal pour l'IA responsable⁶, mais aussi plusieurs travaux de corps divers. Nous pouvons citer par exemple, les analyses réalisées par AI Index⁷ ou par le Global Partnership on Artificial Intelligence (GPAI)⁸. Les objectifs poursuivis par ces différents acteurs sont d'encourager et de contribuer au développement, à l'utilisation et à la gouvernance responsables des systèmes d'IA centrés sur l'homme, en accord avec les objectifs de développement durable des Nations Unies.

Indicateurs clés

Nous proposons dans cette partie des indicateurs clés permettant d'apprécier les évolutions de l'usage des algorithmes de l'IA, notamment pour la sécurité et la santé au travail, ainsi que les questions éthiques que cela pose.

- Incidences liées aux biais algorithmiques dans l'environnement de travail
- Travaux récents en matière d'usage éthique de l'IA
- Freins et avancées vers l'usage éthique de l'IA dans les entreprises

6. <https://www.declarationmontreal-iaresponsable.com/>

7. <https://hai.stanford.edu/research/ai-index-2021>

8. <https://gpai.ai/projects/responsible-ai/>

Analyse rétrospective des indicateurs clés

Incidences liées aux biais algorithmiques dans l'environnement de travail

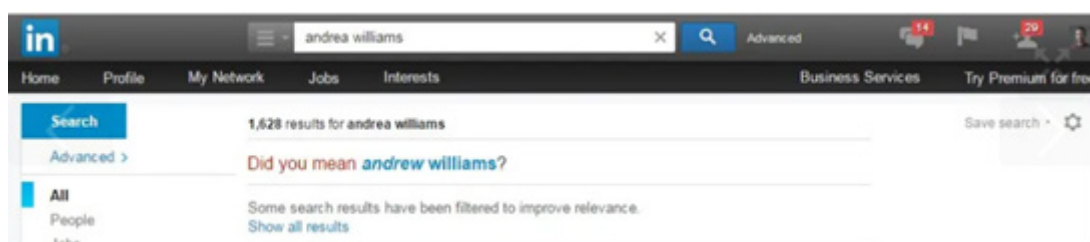
Le management algorithmique dans la gestion du personnel

L'avancée et la culture technologiques, l'accumulation de données historiques, mais aussi la croissance de la puissance numérique des entreprises, ont permis le développement des technologies d'IA au sein des organisations. Les outils d'apprentissage ont été une révolution.

En matière de management algorithmique, on peut citer l'usage de l'IA en ressources humaines pour le recrutement, la formation opérationnelle, l'amélioration de l'expérience salariée ou d'autres activités tactiques, comme l'élaboration d'indicateurs afin d'évaluer le capital humain sur la performance de l'entreprise. Mais il faut noter que le management algorithmique intervient dans le contexte plus large de la mise en place d'outils technologiques et de systèmes de supervision numériques visant à gérer le personnel⁹. Plus largement, l'usage de l'intelligence artificielle peut intervenir dans l'ensemble des processus de décision.

Néanmoins, bien que ces technologies permettent de limiter les biais cognitifs (à ce jour, il existerait 180 biais cognitifs¹⁰), de nombreux exemples ont mis en valeur les biais discriminatoires de ces algorithmes.

L'un des principaux champs de discrimination algorithmique est le recrutement. Dans une enquête menée sur LinkedIn¹¹, des chercheurs ont mis en évidence le fait que l'algorithme de LinkedIn suggérerait aux recruteurs un nom d'homme lorsqu'un nom de femme était inscrit dans la barre de recherche. Dans ce cas de biais de genre, le risque d'une perte de profil intéressant représente un manque à gagner pour l'entreprise.



Enquête « Comment l'algorithme de recherche LinkedIn reflète le biais de genre »

Source : Seattle Times

En termes d'éthique réglementaire, un filtrage algorithmique des *curriculum vitae* peut être contraire au RGPD dans l'UE. Toute décision finale concernant la recevabilité des candidats devrait toujours être prise par un être humain¹².

9. <https://www.etui.org/sites/default/files/2021-06/Management%20algorithmique%20et%20n%C3%A9gociation%20collective-web-2021.pdf>

10. <https://home.kpmg/fr/fr/blogs/home/posts/2021/1/comment-maitriser-utilisation-ia.html>

11. Seattle Times, « [How LinkedIn's search engine may reflect a gender bias](#) ».

12. https://www.uniglobalunion.org/sites/default/files/imce/uni_pm_algorithmic_management_guide_fr.pdf

L'équité algorithmique contre la discrimination

En 2014, Amazon a commencé à développer un système pour automatiser la sélection des candidats à interviewer, en fonction du texte figurant sur leur *curriculum vitae*. Plus tard, l'entreprise a pris conscience que l'algorithme semblait favoriser les hommes au détriment des femmes occupant des postes d'ingénieur. En conséquence, Amazon a tenté d'ajuster l'algorithme pour plus d'équité, mais, faute de résultats satisfaisants, l'a abandonné¹³.

Au regard des biais inhérents au déploiement de l'intelligence artificielle, le concept d'équité algorithmique s'est développé. Dans une enquête sur les biais en *machine learning*, Mehrabi et al. définissent l'équité algorithmique comme étant : « L'absence de tout favoritisme ou discrimination à l'égard d'un individu ou d'un groupe formé par des caractéristiques innées ou acquises »¹⁴. L'équité est un objectif difficile à atteindre : de plus, il n'y a pas de consensus sur sa définition mathématique.

L'équité algorithmique : quelles données sont concernées ?

En France, il est interdit de collecter des informations qui, directement ou indirectement, font apparaître les origines sociales ou ethniques, les opinions politiques ou religieuses, les appartenances syndicales, et les informations relatives à la santé et à la vie sexuelle.

Ce concept est un enjeu de taille pour les entreprises qui doivent concilier entre l'efficacité de leur process et les contraintes d'exploitation des données pour des raisons éthiques. Le choix des métriques de performance servant dans la production des modèles d'IA doit donc prendre en compte les biais possibles afin de les corriger et/ou les éviter.

Travaux récents en matière d'usage éthique de l'IA

Les travaux de l'université de Stanford - AI100

Éric Horvitz¹⁵, directeur de recherche chez Microsoft, diplômé de médecine à l'université de Stanford, a lancé et financé avec sa femme fin 2014 une étude qui va durer un siècle. L'objectif est d'étudier et anticiper les impacts de ces technologies sur tous les aspects de la société et de l'économie au travers des époques.

Le premier rapport de AI100¹⁶ est paru en septembre 2016. Ce premier rapport a été largement couvert par la presse populaire et est connu pour avoir influencé les discussions au sein de conseils consultatifs gouvernementaux et d'ateliers dans de multiples pays. Il a également été utilisé dans une variété de programmes d'études en intelligence artificielle.

L'étude récente est sortie en septembre 2021¹⁷. Ce rapport est structuré comme un recueil de réponses du groupe d'étude 2021 à un ensemble de douze questions

13. [Les Échos](#).

14. <https://arxiv.org/abs/1908.09635>

15. <https://qz.com/973005/microsofts-new-head-of-research-has-spent-his-career-building-powerful-ai-and-making-sure-its-safe/>

16. Peter Stone et al. « Intelligence artificielle et vie en 2030 », étude de cent ans sur l'intelligence artificielle, rapport du panel d'étude 2015-2016, Stanford University, septembre 2016. <http://ai100.stanford.edu/2016-report>.

17. <https://ai100.stanford.edu/2021-report/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence>

permanentes (QS) et de deux questions d'atelier (QA), posées par le Comité permanent AI100. L'objectif des ateliers était de s'appuyer sur l'expertise d'informaticiens et d'ingénieurs, de chercheurs en sciences sociales et humaines (notamment des anthropologues, des économistes, des historiens, des spécialistes des médias, des philosophes, des psychologues et des sociologues), d'experts en droit et en politique publique, et de représentants de la gestion des entreprises, ainsi que des secteurs privé et public.

Un comité permanent élargi, avec plus d'expertise en matière d'éthique et de sciences sociales, a formulé un appel et encouragé activement les propositions de la communauté internationale des chercheurs et des praticiens de l'IA, avec une large représentation des domaines pertinents pour l'impact de l'IA dans le monde. En réunissant des chercheurs issus de l'ensemble des disciplines explorant rigoureusement les impacts éthiques et sociétaux des technologies, les ateliers d'étude visaient à élargir et approfondir les discussions sur la manière dont l'IA façonne les espoirs, les préoccupations et les réalités de la vie des gens et, par conséquent, sur les défis éthiques et sociétaux que l'IA soulève. Après avoir diffusé un appel à propositions et examiné plus de 100 soumissions du monde entier, deux ateliers ont été sélectionnés pour être financés.

Le premier, intitulé « Prediction in Practice », a étudié l'utilisation de prédictions du comportement humain basées sur l'IA (telles que la probabilité qu'un emprunteur finisse par rembourser un prêt) dans des contextes où les données et la modélisation cognitive ne tiennent pas compte des dimensions sociales qui façonnent la prise de décision des personnes. L'autre étude, intitulée « Coding Caring », portait sur les défis et les possibilités liés à l'intégration des technologies d'intelligence artificielle dans le processus de prise en charge des personnes et sur le rôle que jouent le genre et les relations de travail pour répondre au besoin urgent d'innovation dans le domaine des soins de santé. S'appuyant sur les conclusions de ces ateliers d'étude, ainsi que sur le rapport annuel AI Index – un projet dérivé d'AI100 –, le Comité permanent a défini un cahier des charges¹⁸ pour le groupe d'étude à l'été 2019¹⁹ et a recruté Michael Littman, professeur d'informatique à l'université Brown, pour présider le groupe. Le groupe d'étude de 17 membres, composé d'un ensemble diversifié d'experts en IA issus du monde universitaire et des laboratoires de recherche de l'industrie et représentant l'informatique, l'ingénierie, le droit, les sciences politiques, la politique, la sociologie et l'économie, a été lancé au milieu de l'automne 2020. En plus de représenter un éventail de spécialités savantes, le panel avait une représentation diversifiée en termes de régions géographiques d'origine, de sexes et de stades de carrière. Les prochaines études sortiront tous les 5 ans.

Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle

Les 193 États membres de l'UNESCO ont adopté un nouveau texte le 25 novembre 2021 qui vise à donner « une orientation éthique » aux technologies d'intelligence artificielle²⁰. Ce premier instrument juridique mondial sur l'IA n'est pas contraignant pour les signataires, mais les pousse à adopter leur propre cadre afin de lutter contre certaines dérives, comme les biais ou la violation de la vie privée.

18. https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100_charge_2020_final.pdf

19. « One Hundred Year Study on Artificial Intelligence (AI100) », Stanford University, accessed August 1, 2016, <https://ai100.stanford.edu>.

20. <https://www.usine-digitale.fr/article/l-unesco-adopte-une-recommandation-sur-l-ethique-de-l-intelligence-artificielle.N1163202>

Donner une orientation responsable

L'objectif du texte est de « donner une orientation responsable aux technologies de l'IA ». Il cite « les biais susceptibles de comporter ou d'accentuer des inégalités et exclusions » et le risque de menacer « la diversité culturelle, sociale et écologique et entraîner des clivages sociaux ou économiques ». L'UNESCO rappelle aussi que l'IA possède de nombreux bénéfices, à condition d'en assurer « un accès équitable » tout en tenant compte « des dommages et des impacts ».

La recommandation ne fournit pas de définition « unique » de l'IA, « celle-ci étant appelée à évoluer en fonction des progrès technologiques ». Elle vise en général les « systèmes technologiques capables de traiter l'information par un processus s'apparentant à un comportement intelligent, et comportant généralement des fonctions de raisonnement, d'apprentissage, de perception, d'anticipation, de planification ou de contrôle ».

Rappelant que les systèmes doivent respecter la vie privée, être robustes et ne pas provoquer de discriminations, l'UNESCO recommande aux États de mettre en place « des études d'impact » pour identifier les avantages et les risques. Ils doivent mettre en évidence « les répercussions sur les droits de l'Homme, le droit du travail, l'environnement et les écosystèmes ». Des mécanismes de « diligence » et de « supervision » devraient également être mis en place par les États et les entreprises.

Pas de personnalité juridique

La recommandation se penche également sur la responsabilité en cas de dommages causés par un système d'apprentissage automatique. Elle conseille aux États de veiller à ce que les préjudices causés fassent l'objet « d'enquêtes et de réparation » en mettant place « des mécanismes d'exécution solides et des mesures correctives ». Précision importante : elle estime que les systèmes ne doivent pas disposer d'une personnalité juridique, mais que la responsabilité doit toujours incomber à une personne physique ou morale.

Étant intimement liées à l'IA, les données doivent faire l'objet de politique, rapporte l'UNESCO. Il convient d'évaluer « la qualité des données d'entraînement », notamment « l'adéquation des processus de collecte et de sélection des données ». Il est également indispensable de trouver un équilibre entre la collecte des données et la protection de la vie privée.

En revanche, l'UNESCO n'aborde pas directement le sujet de la reconnaissance faciale. Un thème pourtant central, dont l'Union européenne s'est par exemple emparée. La Commission européenne propose une utilisation strictement encadrée des systèmes d'identification biométrique à distance dans les lieux publics.

Synthèse (non exhaustive) des lignes directrices, principes et déclarations sur l'IA

Le tableau ci-après présente les principes et les lignes directrices sur l'éthique de l'IA formulés par différentes organisations depuis 2016.

Nom de l'organisation	Principes et lignes directrices sur l'éthique de l'IA
ACM	<p>2017, « 2018 ACM Code of Ethics and Professional Conduct: Draft 3 », Association for Computing Machinery Committee on Professional Ethics, https://ethics.acm.org/2018-code-draft-3/</p> <p>2017, « Statement on Algorithmic Transparency and Accountability », Association for Computing Machinery US Public Policy Council, www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf</p>
Asilomar	2017, « Asilomar AI Principles », Future of Life Institute, https://futureoflife.org/ai-principles/
COMEST	2017, « Rapport de la COMEST sur l'éthique de la robotique », Commission mondiale d'éthique des connaissances scientifiques et des technologies, https://unesdoc.unesco.org/ark:/48223/pf0000253952_fre
FATML	2016, « Principles for Accountable Algorithms and a Social Impact Statement for Algorithms », Fairness, Accountability and Transparency in Machine Learning, www.fatml.org/resources/principles-for-accountable-algorithms
FPF	2018, « Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models », The Future of Privacy Forum, https://fpf.org/wp-content/uploads/2018/06/Beyond-Explainability.pdf
GEE	2018, « Déclaration sur l'intelligence artificielle, la robotique et les systèmes "autonomes" », Groupe européen d'éthique des sciences et des nouvelles technologies, https://publications.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-fr/format-PDF
HLEG AI (CE)	2019, Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle de la CE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60427
IEEE	2017, « Ethically Aligned Design: Version 2 », Global Initiative on Ethics of Autonomous and Intelligent Systems, Institute of Electrical and Electronics Engineers, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf
Montréal	2017, « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, https://www.declarationmontreal-iaresponsable.com/

Source : OCDE

Freins et avancées vers l'usage éthique de l'IA dans les entreprises

L'IA est aujourd'hui loin de remplir toutes les conditions nécessaires à son acceptation sociale par les travailleurs. Cela s'explique notamment par son opacité et son caractère « éthéré ». Concevoir des IA socialement approuvées suppose de les rendre interactives, universellement accessibles, transparentes et d'assurer un contrôle permanent sur leur devenir qui ne soit pas simplement technique, mais aussi éthique.

Dans ce contexte, certains facteurs aggravants peuvent freiner l'acceptabilité sociale des IA et leurs usages au sein des entreprises :

- **La méconnaissance de l'IA impacte son degré d'acceptabilité par les travailleurs** : en 2017, une étude de l'IFOP affirmait que 80 % des Français jugent la présence des algorithmes déjà massive dans leur vie, mais un peu plus de la moitié avoue ne pas savoir précisément ce que sont les algorithmes²¹. La connaissance des outils d'intelligence artificielle permet de démystifier son usage. Le cabinet Keley Consulting et l'institut YouGov France ont réalisé une étude sur la perception des Français sur l'intelligence artificielle qui met en évidence les différents degrés de connaissance des outils et leurs impacts sur l'optimisme ou le pessimisme ressenti par rapport à l'IA²². Selon les conclusions de l'étude : « Les inactifs, les sans emploi et les moins diplômés voient beaucoup moins les opportunités que peut offrir l'IA. Aussi, les plus diplômés sont 66 % à considérer l'IA comme une opportunité pour la France, contre 56 % des détenteurs du baccalauréat et 52 % de ceux qui ne l'ont pas ».
- **L'usage de l'IA dans les entreprises est perçu comme une menace pour l'emploi** : l'usage de l'IA et son développement est perçu par les Français comme une menace pour l'emploi, contrairement à son développement dans différents autres domaines. Cette crainte crée une incertitude sur l'acceptabilité sociale de l'IA.
- **Apparition de nouveaux risques de santé et sécurité au travail liés à l'usage de l'IA** : dans un rapport traitant des avantages et risques de l'utilisation d'outils d'IA dans un environnement de travail²³, l'agence européenne de la santé et sécurité au travail (EU-OSHA) met en évidence que certains risques psychosociaux suivants sont particulièrement représentés dans les environnements de travail digitalisés. Ce sont le stress, la discrimination, la précarité accrue, les troubles musculo-squelettiques, l'intensification du travail et les pertes d'emploi potentielles. Selon l'agence, ils sont exacerbés lorsque l'IA sert à augmenter des outils technologiques existants ou est introduite pour la première fois sur le lieu de travail à des fins de gestion et de conception.

Malgré ces freins, il faut noter toutefois certains progrès pouvant améliorer la perception des personnes et favoriser l'acceptabilité sociale d'un usage éthique de l'IA dans le cadre du travail :

- **Les évolutions technologiques ont contribué à améliorer le bien-être au travail** : d'après l'étude mondiale d'Oracle & Workplace Intelligence sur l'IA au travail²⁴, l'IA a contribué à l'amélioration de la santé mentale des travailleurs en leur

21. <https://www.ifop.com/publication/notoriete-et-attentes-vis-a-vis-des-algorithmes/>

22. <https://www.actuia.com/actualite/une-etude-sinteresse-a-la-perception-de-lintelligence-artificielle-par-les-francais/>

23. <https://osha.europa.eu/en/publications/osh-and-future-work-benefits-and-risks-artificial-intelligence-tools-workplaces/view>

24. <https://www.oracle.com/fr/news/announcement/artificial-intelligence-supports-mental-health-2020-10-07.html>

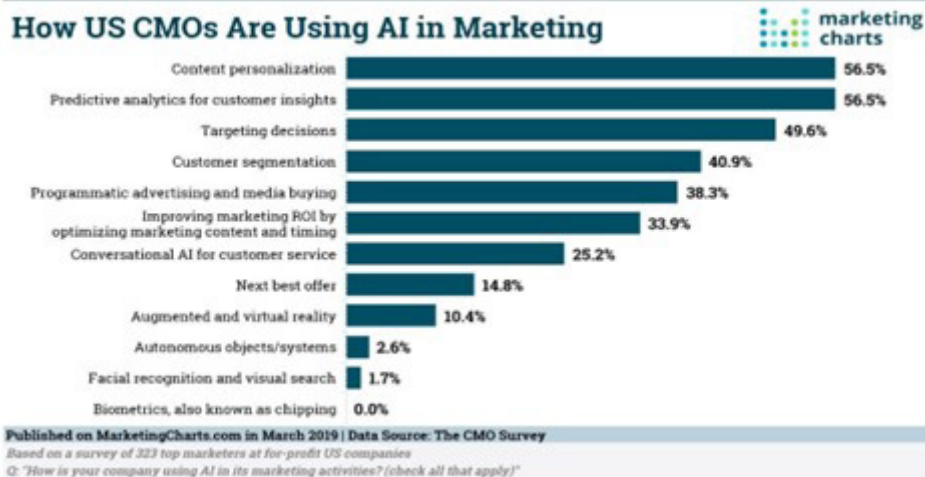
fournissant les informations nécessaires à leur travail (31 % des répondants), en automatisant les tâches (27 % des répondants), et en réduisant le stress (27 % des répondants). Toujours dans cette même étude, il est constaté que les salariés attendent plus de la technologie que la simple mise à disposition d'outils collaboratifs : ils veulent pouvoir compter sur la technologie pour soulager leur santé mentale. L'intelligence artificielle aurait déjà permis à la moitié des salariés de raccourcir la semaine de travail et de prendre de plus longues vacances (51 % au global – 35 % en France). Plus de la moitié des répondants indique que l'intelligence artificielle permet d'accroître la productivité des salariés (63 % au global – 45 % en France), contribue à l'amélioration de la satisfaction au travail (54 % au global – 37 % en France) et améliore le bien-être général (52 % au global – 37 % en France).

- **Formations en IA ouverte à tous** : la Fondation Abeona s'associe avec l'Institut Montaigne et Open Classrooms afin de proposer une formation en intelligence artificielle ouverte à tous. Cette formation a pour but de sensibiliser le grand public aux problèmes d'équité et de justice en IA et de promouvoir de meilleures pratiques. Elle consacre plusieurs chapitres de son cours à restituer les enjeux éthiques, non seulement en démystifiant certaines idées reçues, mais également en précisant les points de vigilance et en proposant des solutions concrètes aux enjeux d'utilisation des données, d'information, de décisions algorithmiques et de biais²⁵.
- **Arsenal éthique dans le cadre de l'environnement de travail** : actuellement, l'intégration d'un cadre éthique dans l'IA vise davantage l'architecture des systèmes d'information, en complément des normes juridiques. Cette approche est appelée « *by design* ». Elle permet d'adresser directement les questions de biais ainsi que les enjeux de l'anonymisation des données personnelles, tout en optimisant les processus opérationnels. En parallèle, on assiste au développement d'un cadre juridique au sein des entreprises, notamment avec la création de comités éthiques de l'IA. Ces comités sont les garants de l'acceptabilité de l'IA au sein des entreprises par les clients, ainsi que par les collaborateurs et les partenaires. En avril 2021, la Commission européenne a proposé le premier cadre juridique sur l'intelligence artificielle au sein de l'Union européenne. Aujourd'hui, la plupart des institutions internationales ont un organisme dédié pour les questions liées à l'intelligence artificielle afin de traiter de ses enjeux éthiques.

De manière parallèle aux efforts des différents acteurs à tendre vers un usage plus éthique de l'IA, ces outils sont de plus en plus adoptés par les entreprises et les salariés. L'IA permet d'assurer des solutions plus optimales et efficaces. En effet, au vu du volume de données traitées, les prévisions faites par les outils de l'IA peuvent être plus précises, fiables et immédiates. De plus, son utilisation permet de réduire les coûts opérationnels. Dans une étude réalisée par la plateforme américaine de données et d'analyses marketing Marketing Charts, l'utilisation de l'IA dans la prise de décision est son troisième type d'usage. La prise de décision vient après les analyses de prévision des clients et la personnalisation des offres.

25. <https://www.institutmontaigne.org/blog/vivre-au-temps-des-algorithmes-quels-sont-les-enjeux>

Les questions éthiques sur l'acceptabilité en matière de prévention de l'usage de l'IA dans le monde du travail



L'IA peut être aussi utilisée comme un service. Par exemple, la mise en place des *chatbots* dans la relation client connaît un fort développement depuis ces dernières années, en raison notamment de la rapidité des réponses obtenues. En 2017, le cabinet Garner envisageait que, en 2022, 85 % des échanges entre les clients et les entreprises se feront via des *chatbots*²⁶. Actuellement, une entreprise sur trois met en œuvre des cas d'utilisation limitée d'IA selon une étude de 2021 du cabinet PwC²⁷.



Certaines technologies actuellement mises en place dans le monde du travail visent à déterminer l'état de santé physique et mentale des travailleurs. Au Québec, la start-up EmoScienS²⁸ a créé un logiciel de reconnaissance faciale qui permet d'analyser l'état émotionnel des utilisateurs. Le logiciel prend à intervalles réguliers des photos du visage à partir de la webcam de l'utilisateur. Les expressions faciales de son visage sont analysées pour dessiner son profil émotionnel. Il propose également des outils pour mieux gérer ses émotions. Ces données ne sont visibles que par ce dernier, à moins qu'il ne souhaite les partager, auquel cas elles seront dépersonnalisées. Malgré ces précautions, ces formes de neuro-surveillance soulèvent d'énormes problèmes éthiques²⁹ dans le cadre de son usage en entreprise.

26. <https://www.alternance-professionnelle.fr/statistiques-bots-conversationnel-relation-client/>

27. <https://www.pwc.fr/fr/publications/data/predictions-ia-2021.html>

28. <https://www.emosciens.com/>

29. V. Stefano, « Masters and servers: collective labour rights and private government in the contemporary world of work », *International Journal of Comparative Labour Law and Industrial Relations*, 2020.

Tendances, signaux faibles et incertitudes majeures

L'analyse rétrospective précédente des principaux indicateurs retenus comme facteurs d'influence des questions et limites éthiques de l'usage de l'IA dans la prévention pour la sécurité et la santé au travail (S&ST) nous conduit à relever les points structurants suivants.

Les tendances lourdes

- Les incidents liés aux biais et discriminations associés à l'usage de l'IA ont des conséquences majeures sur l'environnement de travail et les travailleurs.
- L'usage de l'IA dans les entreprises est perçu comme une menace pour l'emploi.
- L'acceptabilité sociale de l'usage de l'IA est plutôt négative à cause des questions éthiques que cela soulève, ainsi que les menaces potentielles pour l'emploi.

Les tendances émergentes

- Les biais inhérents au déploiement de l'intelligence artificielle ont conduit au développement du concept d'équité algorithmique.
- La perception de l'usage de l'IA est positive pour l'amélioration de la qualité de vie, y compris au travail.
- La multiplication des réglementations autour de l'éthique de l'IA est favorisée par les nombreux principes et lignes directrices proposés par différentes organisations.

Les signaux faibles

- L'apparition de nouveaux risques de santé et sécurité au travail liés à l'usage de l'IA.
- Le développement des outils de surveillance de la santé physique et mentale des travailleurs.

Dans ce contexte, différentes incertitudes majeures peuvent être relevées :

Qui seront les futurs acteurs concernés par l'usage éthique de l'IA?

Ces dix dernières années, nous constatons qu'un grand nombre d'acteurs (nationaux et internationaux) proposent des principes et des lignes directrices pour encadrer l'usage éthique de l'IA.

- En matière de responsabilisation de l'usage de l'IA pour la prévention et la S&ST au sein des entreprises, quels seront les acteurs qui auront un rôle décisif à jouer ? Le secteur public, le secteur privé, la société civile ?
- Quels seront les rôles des travailleurs pour garantir l'équilibre des pouvoirs et l'usage éthique de l'IA au sein des entreprises ?

Dans quelle mesure le développement des normes éthiques sur l'usage de l'IA engendrera des contraintes pour les entreprises ?

Le rapport rendu par le CCNE et la CERNA sur « Numérique et santé » a alerté sur les risques d'une législation française trop rigide qui freinerait la recherche nationale en intelligence artificielle, entraînant un développement de l'usage de solutions étrangères³⁰.

- La réglementation nationale/européenne sur l'usage éthique de l'IA sera-t-elle en faveur de la protection des salariés, au détriment des innovations technologiques ?
- Quel sera le sens des arbitrages sur les contraintes imposées par les nouvelles normes pour les entreprises face à la concurrence internationale (Chine, USA...) ? Limitation des opportunités et accès aux marchés internationaux ?

Comment les travailleurs arbitreront entre l'utilité de l'IA en matière de santé et de sécurité au travail d'une part, et les questions de confiance et d'éthique dans les outils de l'IA d'autre part ?

Dans l'édition 2020 de l'Observatoire de la Notoriété et l'Image de l'Intelligence Artificielle en France, qui associe l'IFOP et le *do tank* IA Impact, un répondant sur deux pense que l'IA améliore les performances, le bien-être au travail et aide à l'évaluation des compétences³¹.

- Les travailleurs, surtout les jeunes générations, opteront-ils massivement pour l'innovation et l'usage de l'IA au détriment de normes éthiques jugées trop restrictives ?
- Est-ce que le développement de l'usage éthique de l'IA permettra une plus grande confiance des travailleurs envers ces outils ? En particulier dans la prévention et la S&ST ?

Hypothèses

Hypothèse I. Initiatives du secteur privé

Le développement éthique par les acteurs privés

Les acteurs privés affichent leurs normes IA. On assiste à une multiplication des comités d'éthique dans les organisations, mais sans grand résultat en termes de réglementation.

L'émergence de labels éthiques reconnus permet d'apaiser les inquiétudes et de faciliter l'acceptabilité sociale des usages de l'IA pour la prévention et la S&ST.

30. <https://www.ccne-ethique.fr/fr/publications/numerique-sante-quels-enjeux-ethiques-pour-quelles-regulations>

31. <https://www.impact-ai.fr/fr/2021/01/22/observatoire2020/>

Hypothèse 2. Initiatives des États

L'éthique devient structurante dans l'usage de l'IA

La plupart des États développent une réflexion structurée donnant lieu à des réglementations encadrant les principes éthiques. En revanche, il y a une diversité de normes au niveau mondial. La concurrence entre les normes éthiques chinoises, américaines, UE, etc. est forte.

Les travailleurs des entreprises internationales ne bénéficient pas des mêmes protections en termes d'usage éthique de l'IA.

Hypothèse 3. Consensus entre tous les acteurs

Consensus éthique pour le développement de l'IA

Les acteurs privés et publics, nationaux et internationaux, parviennent à tomber d'accord sur des règles minimales en termes d'éthique, et un cadre normatif global est adopté.

À partir des recommandations de l'UNESCO, des principes éthiques sont adoptés, favorables à la protection de l'individu, mais sans freiner le développement de la recherche et des innovations. En matière de prévention et de S&ST, les travailleurs bénéficient d'un cadre clair et transparent.

Hypothèse 4. Divergences majeures entre les acteurs

Plusieurs mondes d'éthique assez incompatibles

Émergence de trois approches divergentes cristallisées autour de visions concurrentes de l'éthique de l'usage de l'IA (l'axe Russie, Chine et leurs alliés ; l'axe États-Unis, Canada et leurs alliés ; et l'axe UE et ses alliés).

Ainsi, il y a plusieurs mondes d'éthique d'IA assez incompatibles, car s'excluant, en concurrence, et qui s'affrontent sans cadre global.

Les entreprises et les salariés multiplient les recours en justice, tout en s'adaptant à une concurrence internationale plus complexe.

Croissance économique et géographie de la production (industrielle et services)

M. Héry

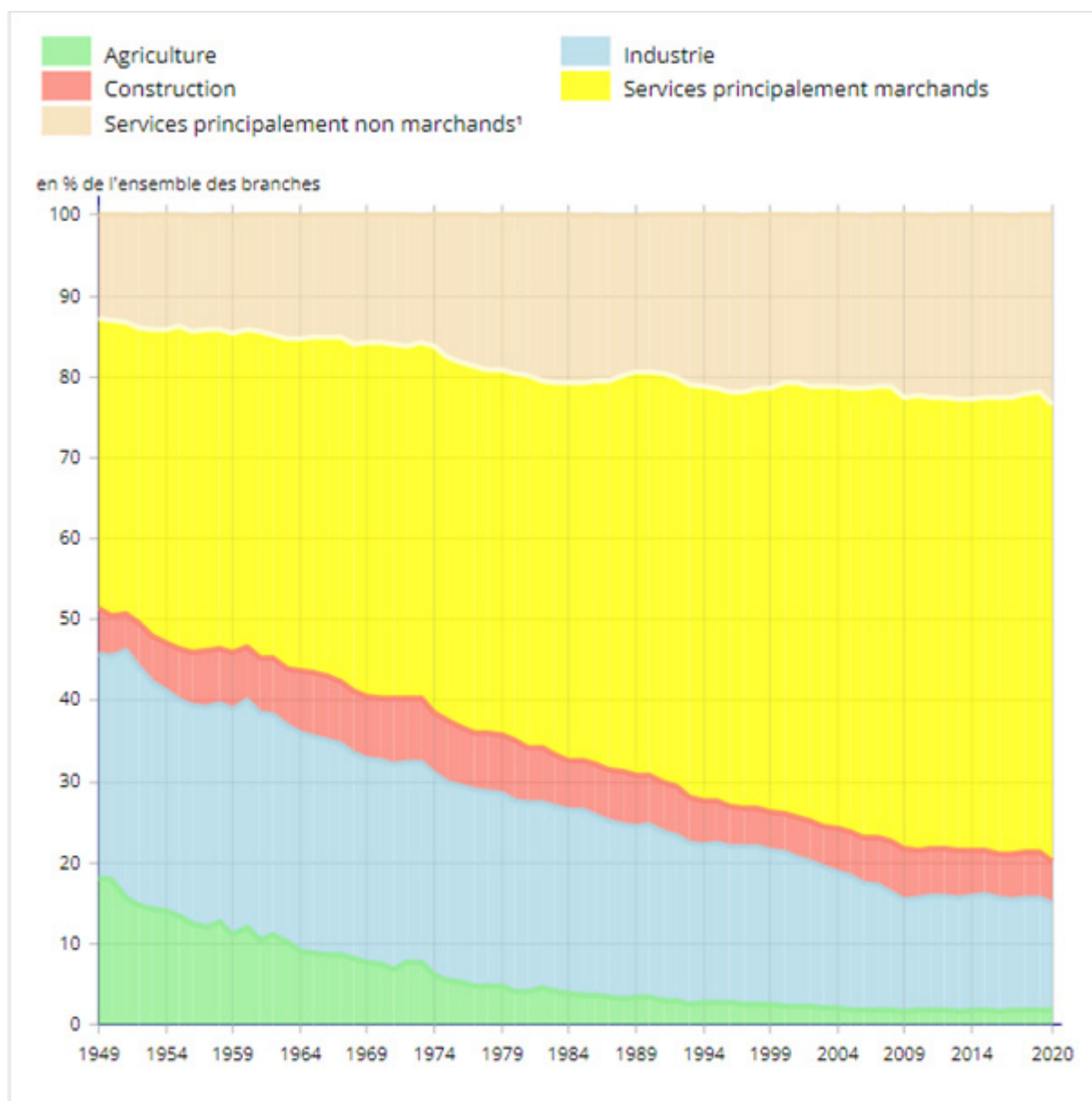
Dans cette fiche variable, nous allons nous intéresser à ce qu'on produira en France dans la dizaine d'années à venir. Nous nous efforcerons en particulier d'apporter des éléments permettant d'identifier quels secteurs et métiers connaîtront des transformations significatives en termes de volume ou de contenu.

Rétrospective

Depuis plusieurs décennies, les parts respectives des secteurs primaire (agriculture et pêche) et secondaire (industrie et construction) n'ont cessé de baisser dans le produit intérieur brut (PIB) de la France au profit du secteur tertiaire (services marchands et non marchands). Depuis une dizaine d'années, on assiste à une relative stabilisation de la production industrielle, mais celle-ci s'effectue à un niveau relativement faible. La part de la construction est restée relativement stable sur les dernières décennies. Le secteur primaire (agriculture, pêche) a également connu une stabilisation relative à un niveau très faible au cours de la dernière décennie. Le secteur tertiaire a vu sa valeur absolue et sa part dans l'économie française augmenter, tant dans le secteur marchand que dans le secteur non marchand.

Les proportions relatives des différentes branches d'activité au cours du temps sont représentées sur la ci-après (en prenant comme référence l'évolution des valeurs ajoutées respectives).

Les évolutions en termes de répartition du PIB entre les différents secteurs et de nombre de travailleurs employés suivent les mêmes tendances que celles de la valeur ajoutée représentées sur le graphique.



1. Regroupement « Administration publique, Enseignement, Santé humaine et action sociale ».

Lecture : en 2020, les services principalement marchands génèrent 56,4 % de la valeur ajoutée de l'ensemble des branches d'activité.

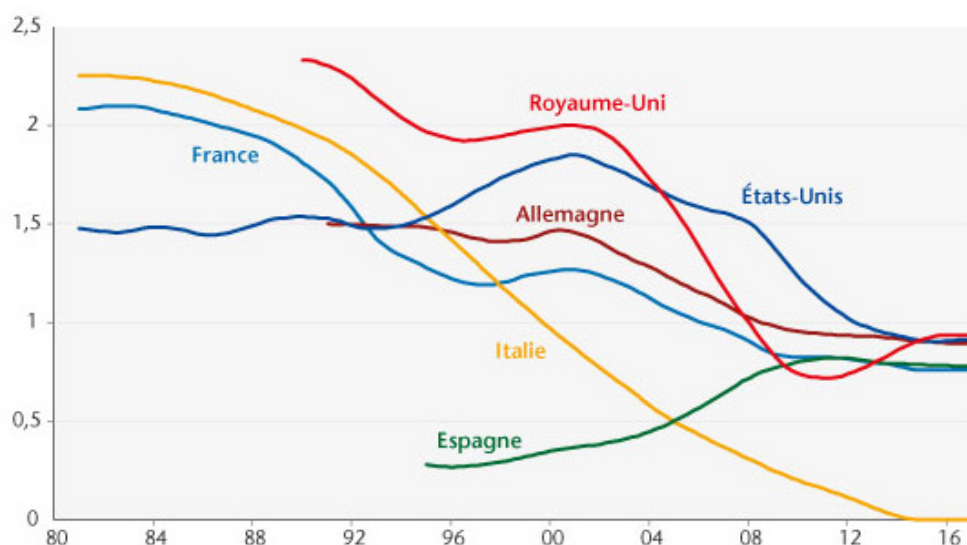
Champ : France.

Source : Insee, comptes nationaux - base 2014.

Source : <https://www.insee.fr/fr/statistiques/2830197>

Depuis des décennies, on assiste à une diminution tendancielle des gains de productivité du travail en France, comme dans les autres pays développés. À l'exception des États-Unis¹ pendant une brève période à la fin des années 1990 (au plus fort du développement de l'utilisation des technologies de l'information et de la communication dans les activités productives), cette tendance n'a pratiquement pas connu d'évolution durable à la hausse depuis une quarantaine d'années.

1. À l'exception aussi de l'Espagne dont l'industrie a bénéficié d'un effort important de modernisation au tournant du siècle.



Note : gains tendanciels de productivité horaire en rythme annuel.

Source : <https://www.ofce.sciences-po.fr/blog/nouveau-sentier-de-croissance-de-productivite-travail/>

Comme on peut le constater dans le tableau ci-dessous, le partage de la valeur ajoutée entre la rémunération (immédiate ou différée) des salariés, l'excédent brut d'exploitation, et les impôts a assez peu évolué au cours des dernières décennies.

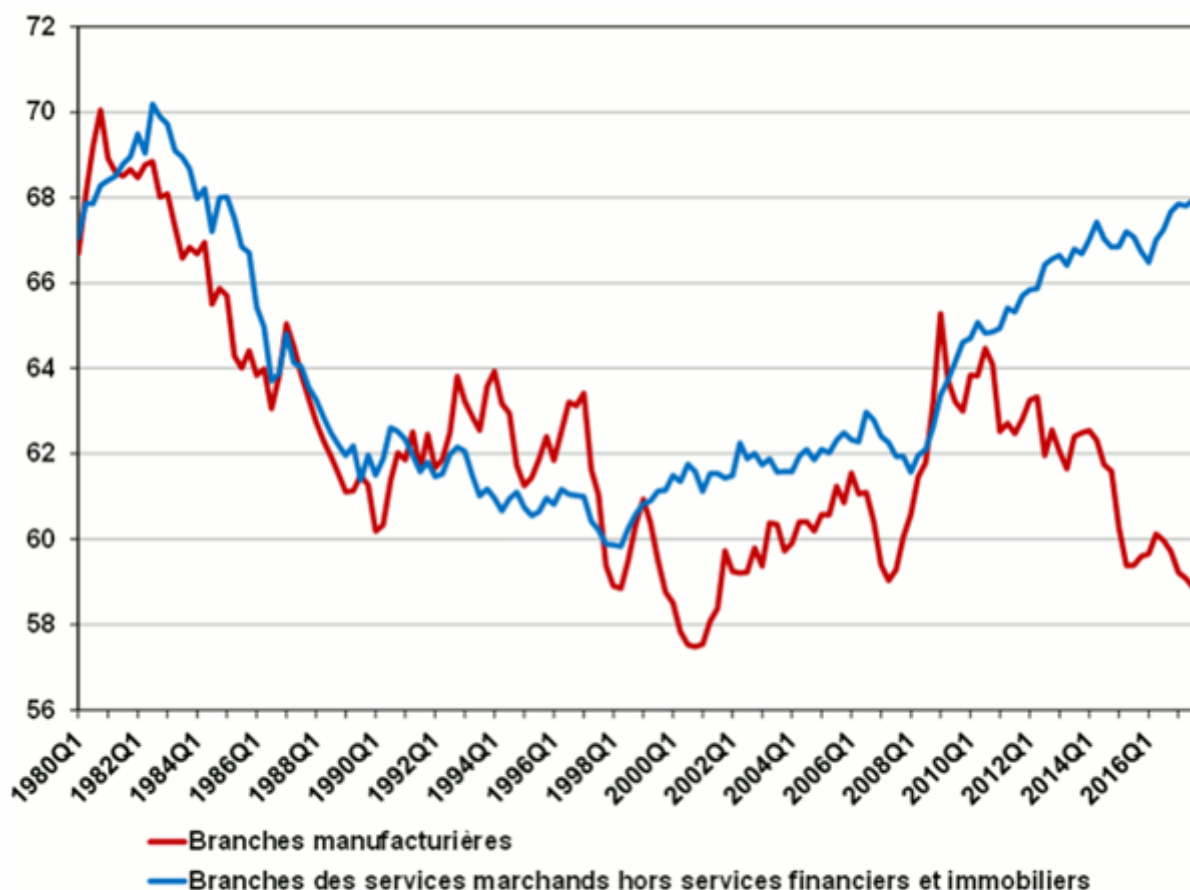
Figure 4 - Partage de la valeur ajoutée à prix courants en 2018

	1990	2000	2008	2018
	(en %)			
Rémunération des salariés	56,6	56,7	56,3	58,9
Salaires et traitements bruts	41,1	41,6	41,6	43,2
Cotisations sociales à la charge des employeurs	15,5	15,1	14,7	15,7
Excédent brut d'exploitation	31,6	31,9	33,2	32,4
Impôts sur la production	3,7	4,5	4,6	5,2
Subventions d'exploitation	-1	-1	-1,4	-2,3
Revenu mixte brut des entrepreneurs individuels	9,1	7,9	7,3	5,8
Ensemble	100,0	100,0	100,0	100,0
	(en milliards d'euros)			
Valeur ajoutée brute	943,9	943,9	1 792,8	2 090,9
+ impôts sur les produits ¹	124,1	169,3	212,5	283,9
- subventions sur les produits	-14,4	-17,0	-13,0	-21,8
Produit intérieur brut (PIB)	1 053,5	1 096,1	1 992,4	2 353,1

1. TVA, impôts sur les importations et autres impôts sur les produits (TIPP, taxes sur les tabacs, sur les alcools, etc.).

Champ : France.

En revanche, la part dévolue au travail dans la valeur ajoutée des branches ayant fluctué dans des proportions relativement importantes dans les dernières décennies se singularise par un double phénomène très net au cours des dix dernières années : d'une part, une diminution significative dans les branches manufacturières et, d'autre part, une augmentation constante dans les services marchands (hors services financiers et immobiliers). Il est tentant d'y voir, dans le premier cas, le résultat d'une forte augmentation de la robotisation dans l'industrie et, dans le second, le développement des différentes activités de services à la personne, qui demandent peu d'investissement matériel, mais impliquent souvent beaucoup de travail humain puisqu'elles sont très peu automatisables.



Graphique 2 : Part du travail dans la valeur ajoutée brute des branches (en %) Source : Insee, calculs des auteurs

Source : <https://blocnotesdeleco.banque-france.fr/billet-de-blog/part-du-travail-dans-la-valeur-ajoutee-un-diagnostic-difficile>

Les dynamiques de changement

Depuis une bonne vingtaine d'années, on a vu se succéder une série de plans de réindustrialisation de la France. Les différents résultats présentés dans la partie précédente amènent à les considérer comme autant d'échecs, sauf qu'ils ont permis de stabiliser la

situation et d'éviter un déclin plus accentué. Industriellement, la France conserve des points forts dans les secteurs suivants : aéronautique, luxe, agroalimentaire, chimie fine et pharmacie (même si la fabrication des molécules les plus simples a souvent été délocalisée dans des pays à faible coût de main-d'œuvre et aux réglementations du travail et de l'environnement plus laxistes qu'en France). Dans beaucoup d'autres secteurs (fabrication de voitures, textile, confection, sidérurgie, métallurgie, etc.), les délocalisations ont été nombreuses. Certaines activités précédemment délocalisées ont été relocalisées (pour des raisons de qualité ou de réactivité de la production, de préservation du secret industriel, pour mieux répondre aux attentes du consommateur en termes de durée de livraison, etc.), mais, en règle générale, ce retour s'est traduit par une automatisation de la production plutôt que par des créations d'emplois.

La France est en revanche nette exportatrice d'activités de services.

Deux phénomènes pourraient venir remettre en cause la logique d'organisation de la production qui a façonné l'économie française depuis quelques décennies.

Le premier concerne les questions environnementales, et notamment le réchauffement climatique. Afin d'en limiter les effets, la consommation d'énergies fossiles devrait diminuer, au profit d'énergies renouvelables (ou nucléaires). Mais cette substitution ne pouvant s'effectuer que progressivement (pour des raisons technologiques et de coût, notamment dans les pays émergents ou pauvres), il sera probablement nécessaire de revoir les modèles de production et de consommation, pour plus de durabilité des biens et pour une plus faible consommation des ressources naturelles. On pense à l'économie circulaire, mais d'autres modèles sont probablement possibles. Le modèle productif mondial (et bien sûr celui de la France) pourrait donc connaître des changements significatifs, dont le bon sens (mais pas forcément les décisions politiques) voudrait qu'ils soient mis en œuvre rapidement.

Le second renvoie à la pandémie de Covid-19. Celle-ci a fait prendre conscience de la fragilité des systèmes d'échanges mondiaux, notamment des longueurs et ramifications excessives des chaînes d'approvisionnement, en particulier pour des biens de première nécessité tels que certains médicaments ou composants (surtout électroniques) produits par un nombre limité d'entreprises et de pays. Des sécheresses limitant la quantité d'eau disponible ou la production d'énergie, des restructurations de la production ou des problèmes de qualité s'ajoutant à la relative désorganisation des voies de transport (en particulier maritimes) suffisent à ralentir significativement la production mondiale de certains biens de consommation, freinant la reprise d'activité. La question d'une pénurie (matières premières, mais aussi de transports) se pose et s'est déjà traduite par une explosion de certains prix. Reste à connaître la durée de ce phénomène.

Une restructuration du système de production industrielle français, comme le rapatriement déjà évoqué de certaines productions ou des reconversions des ateliers de certaines entreprises vers d'autres productions (de l'aéronautique vers les transports collectifs terrestres, par exemple) pourrait donc intervenir.

La question des services, pour lesquels La France est globalement exportatrice, pourrait présenter un tout autre cas de figure.

Si les services à la personne sont par nature non délocalisables, la crise sanitaire de la Covid-19 a fait tomber beaucoup de tabous sur les questions de télétravail. Compte tenu des résultats des premières négociations dans les entreprises, un retour au *statu quo* apparaît peu probable, même si la situation n'est pas figée et que des clauses de revoyure ou des négociations ultérieures pourront faire bouger ces accords. Les entreprises vont donc apprendre à travailler sur un nouveau mode.

Inévitablement, la question des contrats de travail va se poser dans certaines entreprises. Quand un cadre n'aura plus que des contacts en présentiel épisodiques avec son entreprise, qu'il effectuera son travail en presque totale autonomie, il n'est pas impossible que certains employeurs réfléchissent à transformer le contrat de travail en contrat commercial : cette tendance existait déjà avant la crise sanitaire dans certains métiers, le plus souvent assez peu qualifiés (par exemple, ceux générés par la plateforme Amazon Mechanical Turk). La période post-crise sanitaire risque donc de voir se renforcer et se multiplier les statuts de travailleurs indépendants, notamment pour des travailleurs qualifiés. Cela ne pourra qu'être facilité par les progrès technologiques, qui permettent la collaboration entre des personnes à distance. On pense à l'outil Marcel, développé par Publicis et Microsoft, qui a pour objectif de faire collaborer, grâce à l'utilisation d'intelligences artificielles faibles, les 80 000 employés de l'entreprise présents dans 130 pays. On pense aussi à l'algorithme Foundry, développé par l'université Stanford, qui permet le recrutement d'indépendants dans le monde entier pour des contrats de quelques heures, sans intervention d'un service de ressources humaines, et qui organise la remontée des travaux vers le donneur d'ordres, tout en réservant un fort intérêt au retour de suggestions de la part des indépendants qui ont effectué le travail.

Certaines entreprises de services pourraient alors envisager de procéder à une « délocalisation » virtuelle de tout ou partie de leur activité, en confiant à des travailleurs vivant à des milliers de kilomètres dans des pays à faible coût de vie (et à faibles salaires) des tâches effectuées auparavant par des travailleurs de haut niveau en France. La formule n'est pas nouvelle, mais jusqu'à présent elle concerne surtout des travaux souvent répétitifs à faible valeur ajoutée.

Depuis quelques années, l'utilisation de la RPA (automatisation robotisée des processus), qui apprend les tâches à automatiser en observant le comportement de travailleurs, va croissante. Elle peut utiliser des techniques d'intelligence artificielle faible pour des tâches souvent répétitives (comme la génération de factures ou les réponses à des mails, les relances de primes d'assurance, etc.). La mise en place des techniques de RPA se traduit généralement par des suppressions de postes. Ces techniques sont relativement bon marché et assurent un retour très rapide sur investissement, notamment parce qu'elles s'insèrent dans les processus déjà en place et ne remettent pas en cause l'organisation générale de la production.

Hypothèses

Hypothèse 1.

Sous l'influence des contraintes environnementales, économiques et sociales, on assiste à une revitalisation de l'industrie française, correspondant notamment au rapatriement de certaines activités, à la poursuite de l'automatisation et au redéploiement de certains secteurs vers des productions socialement responsables, et d'un point de vue environnemental (transports en commun, économies d'énergie, usages partagés).

Ce renouvellement de l'industrie française s'accompagne d'un développement d'activités de services directement associées, y compris à l'exportation. Dans le même temps, les activités de services à la personne – un des moteurs du développement du secteur tertiaire en France – connaissent également un nouvel essor lié à la fourniture par l'industrie de biens conçus pour faciliter le travail des soignants et des aidants.

Hypothèse 2.

L'industrie française connaît un renouveau grâce au choix d'une forte libéralisation des conditions d'emploi et de travail. Ces décisions se traduisent d'une part par des délocalisations vers des pays à bas coût de main-d'œuvre des activités dont la rentabilité n'est pas assurée en France, même dans un contexte d'automatisation. Et d'autre part, par la création de nouvelles activités en France qui requièrent de la dextérité plutôt que des compétences techniques de haut niveau, ce qui, compte tenu de la baisse du coût de la main-d'œuvre, ne favorise pas l'automatisation. Certaines entreprises étrangères font d'ailleurs le choix d'une implantation en France de nouvelles unités ou du rachat de plus anciennes pour profiter de ces conditions favorables et de la localisation de la France à proximité de lieux de consommation importants.

Un schéma identique s'applique pour les services, tant à l'export que pour le marché intérieur.

Hypothèse 3.

Faute d'investissements et de volonté politique, l'industrie française ne peut suivre le fort développement technologique caractéristique des années 2025-2035. Elle peine à s'imposer dans les nouvelles technologies et les mutations énergétiques et reste cantonnée à ses domaines d'excellence (luxe, industrie agro-alimentaire, etc.).

L'accent est mis sur le développement des services, aux entreprises comme aux particuliers (tourisme, services à la personne). Dans ce domaine, de grands groupes français spécialisés dans l'environnement, l'industrie hôtelière, les services bancaires ou assurantiels voient leurs affaires se développer avec succès à l'export. Le marché intérieur reste relativement atone, à l'exception des prestations pour les clients étrangers (tourisme).

Hypothèse 4.

L'industrie française est de plus en plus intégrée à quelques grands opérateurs européens, les seuls capables de dégager les moyens financiers nécessaires au développement

toujours plus coûteux de biens mettant en œuvre les nouvelles technologies. Parallèlement, un marché de régulation de la pollution, très actif financièrement, s'est mis en place à travers la négociation de quotas d'émissions.

Dans le domaine des services, l'activité ne connaît pas d'évolutions notoires par rapport à la situation d'avant la crise sanitaire des années 2020-2021.

Démographie du travail (vieillesse, qualification, formation...)

M. Héry

Contexte

À travers cette fiche variable, il s'agit de faire un point global des évolutions que va connaître la population au travail : du point de vue démographique, de celui du niveau de la formation initiale et du nombre de travailleurs formés, mais également de celui de la formation continue et des moyens qui lui sont accordés. Les données présentées ici sont diverses : rétrospectives centrées sur les dernières années et, dans certains cas, prospectives. Dans la mesure du possible, nous avons cherché des chiffres se référant à la période de l'exercice à horizon 2036 : compte tenu de la ressource documentaire identifiée, de nombreux accommodements par rapport à cette règle ont été nécessaires.

Ces travaux doivent être replacés dans le contexte décrit par un rapport du Massachusetts Institute of Technology (MIT) consacré à l'avenir du travail¹. De façon relativement contre-intuitive, ce ne sont pas les capacités d'innovation², ni celles des entreprises à les mettre en œuvre, qui constituent selon le MIT un éventuel frein au développement de l'activité économique dans les dix à vingt ans à venir, mais le fait que la population au travail pourrait ne pas avoir un niveau de formation suffisant. En conséquence, il préconise de nombreuses mesures visant à faciliter l'accès à la formation des minorités (les femmes, les travailleurs pauvres, etc.) à travers des mesures financières d'incitation, mais aussi en baissant le prix des formations (développement par exemple des *community colleges*, universités axées sur les formations du supérieur en deux ou trois ans, ouvrant un large accès aux travailleurs dans le cadre de la formation continue).

Part de la population diplômée

France Stratégie a produit des séries longues et des projections de la population active par niveau de diplôme³ dans le cadre de la réalisation de « Prospective des métiers et des

1. D. Autor, D. Mindell, E. Reynolds, « The Work of the Future: Building Better Jobs in an Age of Intelligent Machines », Massachusetts Institute of Technology, 2 novembre 2020. <https://workofthefuture.mit.edu/>

2. Même si le MIT insiste sur le fait que les applications novatrices actuelles dans le domaine de l'industrie et des services sont le fruit de politiques massives d'investissement dans la recherche publique entre les années 1950 et 1980 et que la forte diminution des budgets qui y sont consacrés depuis pourrait bien constituer un handicap majeur dans les décennies à venir.

3. J. Flamand, « Séries longues et projections de population active par niveau de diplôme », France Stratégie, février 2020. <https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/fs-dt-projections-population-active-fevrier-2020.pdf>

qualifications 2030 » (paru en mars 2022).

Ainsi, pour les femmes, la part des diplômées du supérieur dans la population active est représentée sur la figure 1. Cette projection est faite jusqu'en 2033.

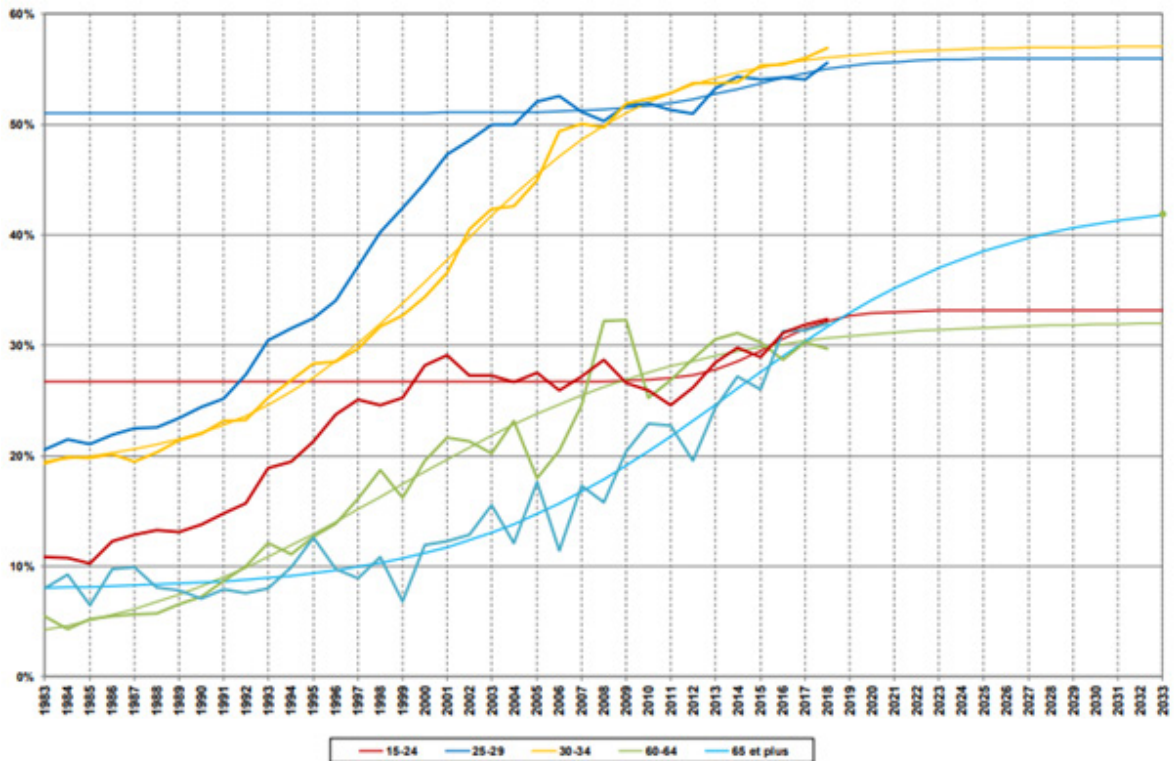


Figure 1. Projections des parts de diplômées du supérieur dans la population active par sexe et tranche d'âge quinquennal pour le sexe féminin. Catégories 15-24, 25-29, 30-34, 60-64 et au-delà de 65 ans.

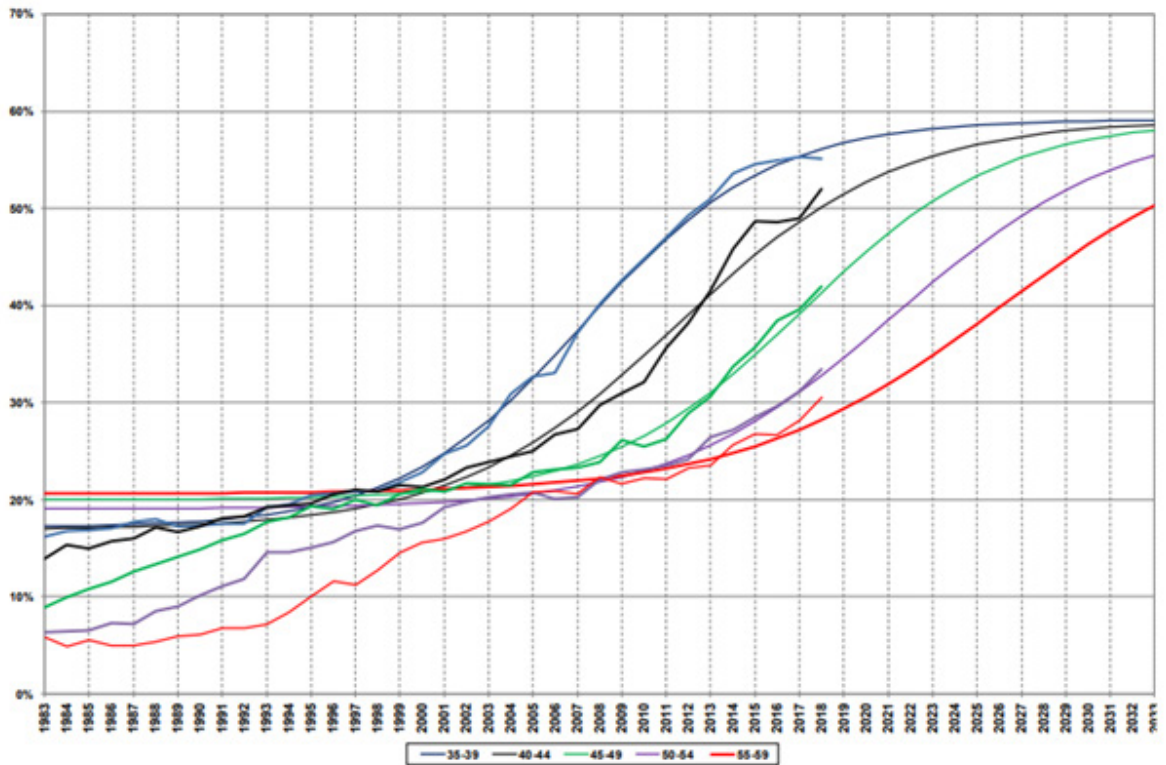


Figure 2. Projections des parts de diplômées du supérieur dans la population active, par sexe et tranche d'âge quinquennal pour le sexe masculin.

tranche d'âge quinquennal pour le sexe féminin. Catégories 35-39, 40-44, 45-49, 50-54 et 55-59. Des schémas analogues pour les hommes sont représentés sur les figures 3 et 4.

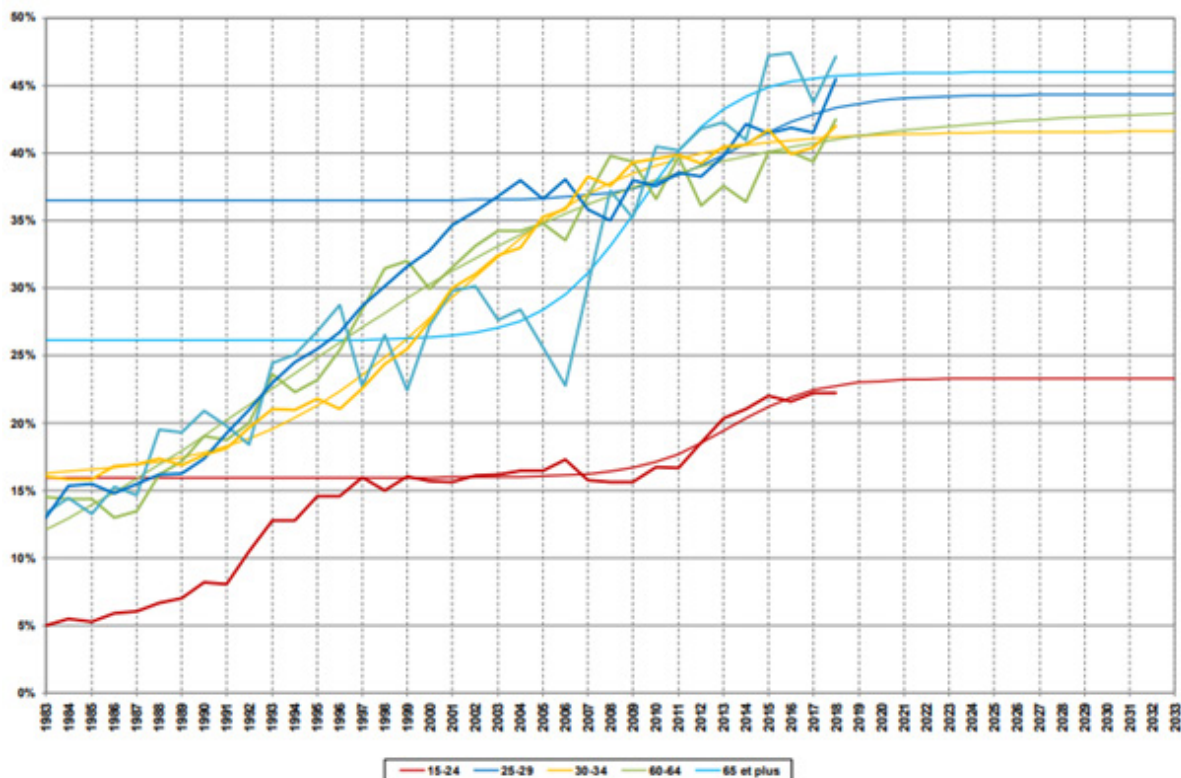


Figure 3. Projections des parts de diplômés du supérieur dans la population active par sexe et tranche d'âge quinquennal pour le sexe masculin. Catégories 15-24, 25-29, 30-34, 60-64 et au-delà de 65 ans.

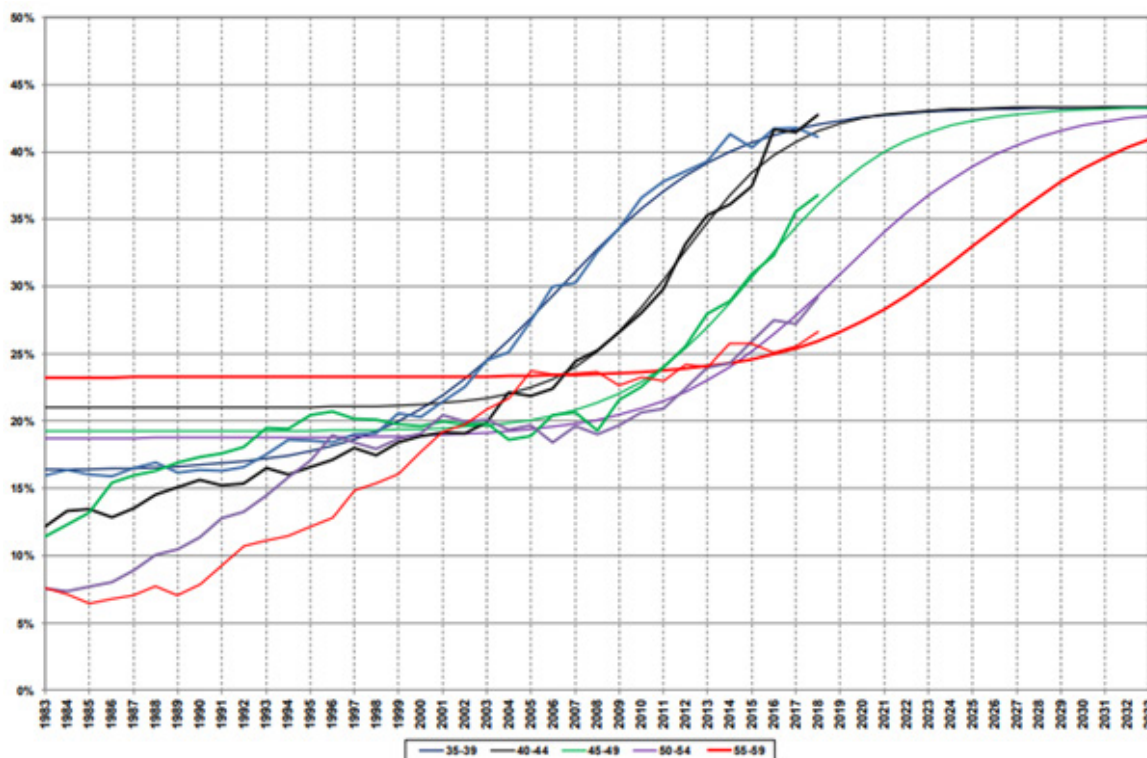


Figure 4. Projections des parts de diplômés du supérieur dans la population active par sexe et tranche d'âge quinquennal pour le sexe masculin. Catégories 35-39, 40-44, 45-49, 50-54 et 55-59.

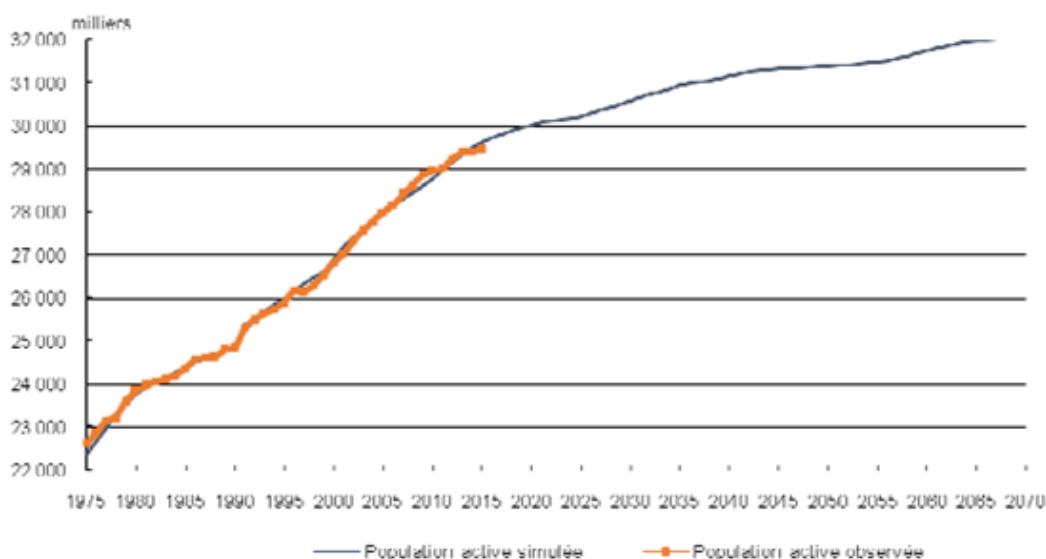
Part de la population en emploi

L'Insee a publié en 2017 les projections de la population active à l'horizon 2070 en France⁴. Ces projections effectuées à intervalles réguliers (cinq à six ans) s'appuient sur les estimations de population les plus récentes, auxquelles sont appliquées des projections de taux d'activité par sexe et tranche d'âge. Elles donnent lieu à l'émission d'un certain nombre d'hypothèses et de scénarios, avec un accent particulier apporté aux tranches d'âge entre 55 et 69 ans. Pour ces dernières, un modèle particulier de microsimulation – dont l'objectif principal est la projection à long terme des retraites – a été utilisé.

La population active comprend les personnes en emploi et celles au chômage, les deux étant entendues au sens du Bureau International du Travail (BIT). Les personnes en emploi sont des personnes âgées de 15 ans ou plus ayant travaillé au moins une heure rémunérée pendant la semaine de référence, celles ayant une activité en tant qu'aide familiale ou conjoint collaborateur, et celles ayant un emploi rémunéré, mais n'ayant pas travaillé durant la semaine de référence pour un certain nombre de raisons (congrés payés, arrêt maladie, congé maternité, congé parental, formation pour une entreprise, chômage partiel... avec dans certains cas des critères de durée d'absence). Les chômeurs sont des personnes n'étant pas en emploi au sens du BIT, disponibles pour travailler dans les quinze jours, et ayant effectué des démarches de recherche active d'emploi dans les quatre semaines finissant la semaine de référence, ou bien ayant trouvé un emploi.

Entre 2005 et 2015, le nombre d'actifs s'est accru en moyenne de 152 000 personnes par an. Selon le scénario central, entre 2015 et 2040, l'application des chiffres tendanciels donnerait une augmentation de la population active d'environ 1,5 million de personnes, atteignant 31,1 millions, soit une croissance annuelle moyenne presque divisée par deux à 62 000 personnes.

La population active progresserait ensuite encore plus modérément jusqu'en 2055, sur un rythme annuel moyen de 22 000 personnes. Sur la figure ci-dessous, on peut comparer les projections obtenues en appliquant le même modèle au réel observé pour la période de 1975 à 2015.



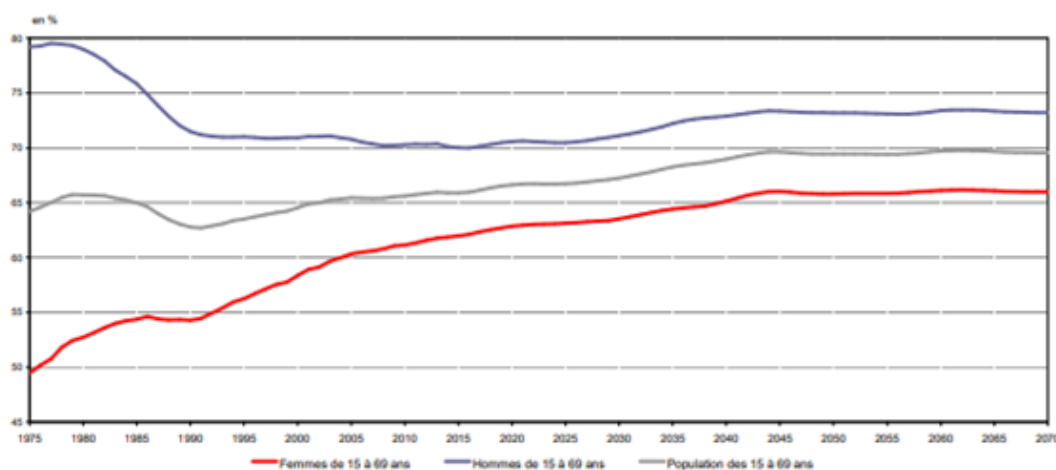
Champ : personnes de 15 ans ou plus en ménages ordinaires; France métropolitaine jusqu'en 1990, France hors Mayotte de 1991 à 2013, France à partir de 2014.

Source : projections de population active 2016-2070, Insee.

4. M. Koubi, A. Marrakchi, « Projection de la population active à l'horizon 2070 », Insee, F1702, 2017. <https://www.insee.fr/fr/statistiques/2845558>

Cette représentation ne rend pas compte des disparités significatives en termes de sexe et d'âge.

La figure 6 montre que, si on s'attend à ce que le taux d'activité des hommes et des femmes de la population augmente d'ici 2036, la différence entre les sexes ne se résorbe que très légèrement.

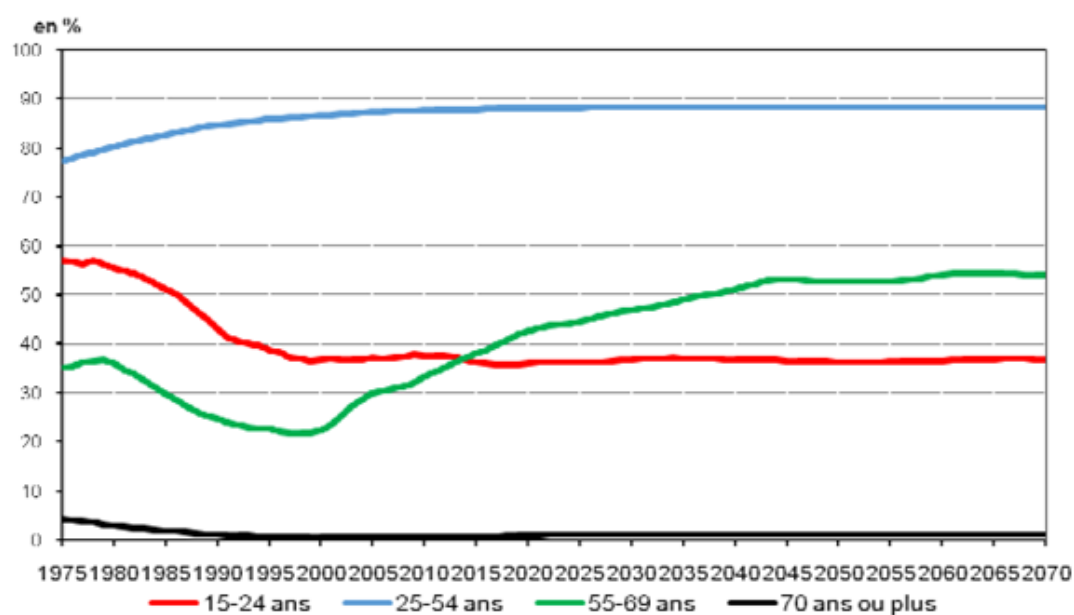


Champ : personnes de 15 à 69 ans en ménages ordinaires; France métropolitaine jusqu'en 1990, France hors Mayotte de 1991 à 2013, France à partir de 2014.

Source : projections de population active 2016-2070, Insee.

Figure 6. Taux d'activité tendanciel des 15-69 ans, par sexe (en %)

En revanche, la période devrait être marquée par une forte progression de l'emploi de la tranche d'âge de 55 à 69 ans, dans la continuité de ce à quoi on assiste depuis le début des années 2000. Les évolutions des autres tranches d'âge apparaissent marginales : ce phénomène apparaît nettement sur la figure 7 ci-dessous.



Champ : personnes de 15 ans ou plus en ménages ordinaires; France métropolitaine jusqu'en 1990, France hors Mayotte de 1991 à 2013, France à partir de 2014.

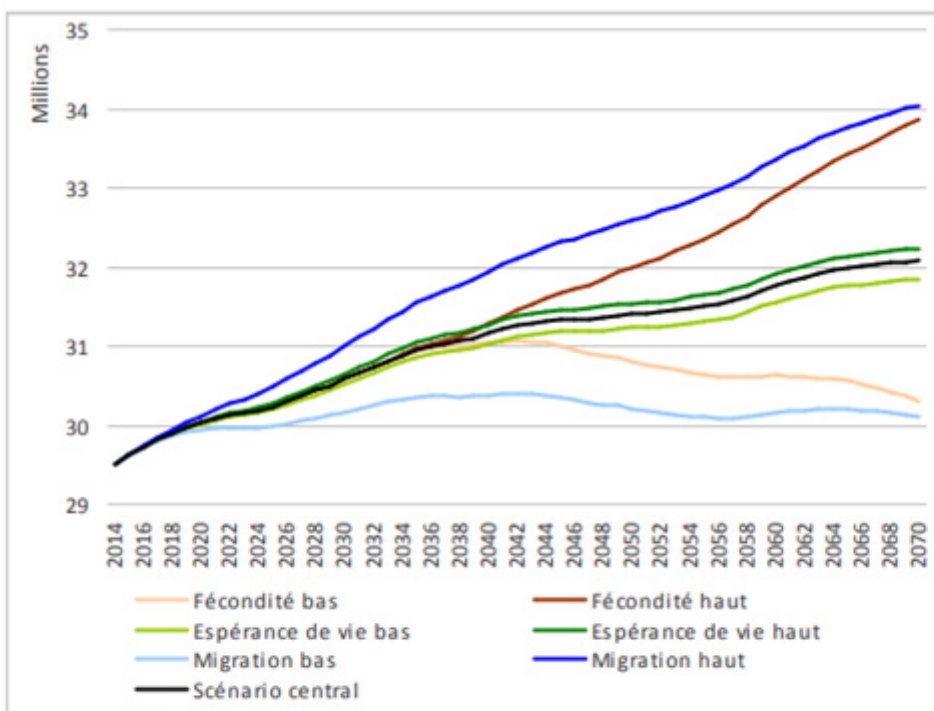
Source : projections de population active 2016-2070, Insee.

Figure 7. Taux d'activité tendanciels, par âge regroupé (en %)

Des découpages plus fins des tranches d'âge (non représentés ici) montrent que le phénomène touche plus particulièrement :

- Les femmes entre 55 et 59 ans dont le taux d'emploi dans les quinze ans à venir se rapproche sensiblement de celui des hommes, pour l'égaliser vers 2070 ;
- Les hommes plus que les femmes entre 60 et 64 ans : leur taux d'emploi voisin actuellement augmente significativement à l'horizon 2036, mais le phénomène est beaucoup plus marqué pour les hommes (de 42 % à 68 % pour eux contre 41 % à 60 % pour elles) ;
- De façon analogue, les hommes plus que les femmes entre 65 à 69 ans, avec une progression de 10 % à 18 % pour eux et de 7 % à 9 % pour elles à l'horizon 2036.

Comme indiqué précédemment, il s'agit là du scénario tendanciel établi par les chercheurs de l'Insee. Les variations extrêmes de trois paramètres ont été prises en compte dans des scénarios alternatifs : la fécondité, l'espérance de vie et les phénomènes migratoires. La figure 8 montre qu'ils peuvent agir de façon significative sur le nombre de personnes au travail (*a fortiori* si des tendances extrêmes se conjuguent dans le même sens). Il faut cependant noter que, pour la période prise en compte dans l'exercice de prospective (jusque 2036), seules les variations des tendances migratoires ont un effet significatif. Les taux de fécondité et d'espérance de vie (sauf survenue d'un phénomène catastrophique) ont une inertie suffisamment grande pour ne pas influencer significativement à court, voire à moyen terme.



Champ : personnes de 15 ans ou plus en ménages ordinaires; France métropolitaine jusqu'en 1990, France hors Mayotte de 1991 à 2013, France à partir de 2014.

Source : projections de population active 2016-2070, Insee.

Figure 8. Variantes démographiques : effets sur la population des 15 ans ou plus en ménages ordinaires.

L'étude insiste particulièrement sur le fait que les chiffres concernant la tranche d'âge 65-69 ans sont soumis à une incertitude relativement forte, liée au fait que la hausse des taux d'emploi est récente et qu'elle est fortement liée à des dispositifs réglementaires relativement récents à l'époque de l'étude (2017).

En conclusion, il apparaît que la population active devrait augmenter de façon significative d'ici 2036 (horizon de l'exercice de prospective). Cette augmentation est cependant conditionnée par une augmentation de l'activité de la tranche d'âge de 60 à 70 ans. Si cette condition est remplie, le seul paramètre susceptible de faire varier le volume de cette population active sont les phénomènes migratoires. Des variations dans les taux de fécondité ou de mortalité n'influent de façon significative qu'au-delà de 2045.

La formation professionnelle initiale

Sur les dix dernières années, le nombre d'élèves ou d'apprentis suivant une formation professionnelle dans le second degré (lycées professionnels et centres de formation des apprentis) a subi une baisse d'environ 10 %, dans un contexte de rénovation des diplômes organisés principalement autour des bacs professionnels et de certificats d'aptitude professionnels (CAP) préparés par apprentissage ou dans les lycées professionnels. Cette évolution est présentée ci-dessous⁵.

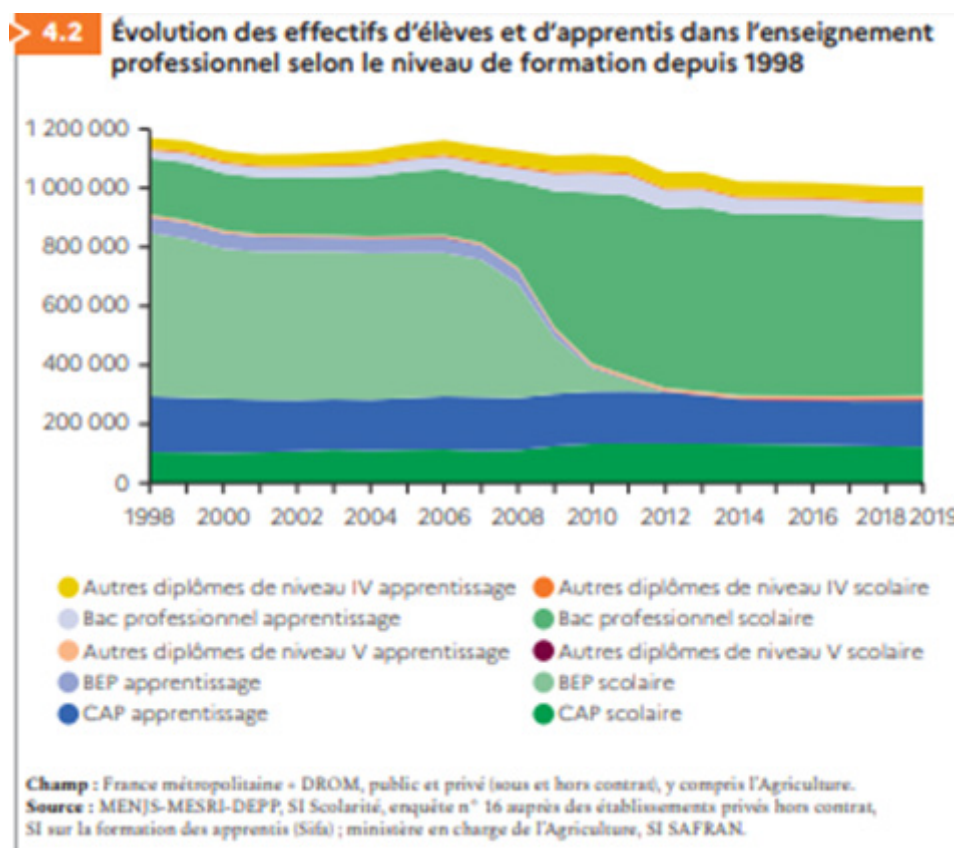


Figure 9. Évolution des effectifs d'élèves et d'apprentis dans l'enseignement professionnel selon le niveau de formation depuis 1998.

5. Direction de l'évaluation, de la prospective et de la performance, « L'état de l'école 2020 », ministère de l'Éducation nationale, décembre 2020. <https://www.education.gouv.fr/l-etat-de-l-ecole-2020-307185>

Les lycées professionnels forment les trois quarts des jeunes de la voie professionnelle. La proportion des jeunes sous statut scolaire est majoritaire parmi ceux préparant un baccalauréat professionnel (92 %). En revanche, les CFA (centres de formation d'apprentis) forment l'ensemble des jeunes préparant un brevet professionnel, plus de la moitié de ceux préparant un CAP (56 %) et plus globalement un diplôme de niveau V (57 %). En dépit des baisses constatées depuis 2009, les spécialités de la production restent plus souvent suivies par apprentissage que celles des services. Les spécialités des services sont majoritairement suivies sous statut scolaire. L'évolution de ces répartitions entre 2009 et 2019 apparaissent sur la figure 10.

Tout ceci s'inscrit dans un contexte où on observe une attrition des filières professionnelles dans l'enseignement secondaires, au profit des bacs généraux.

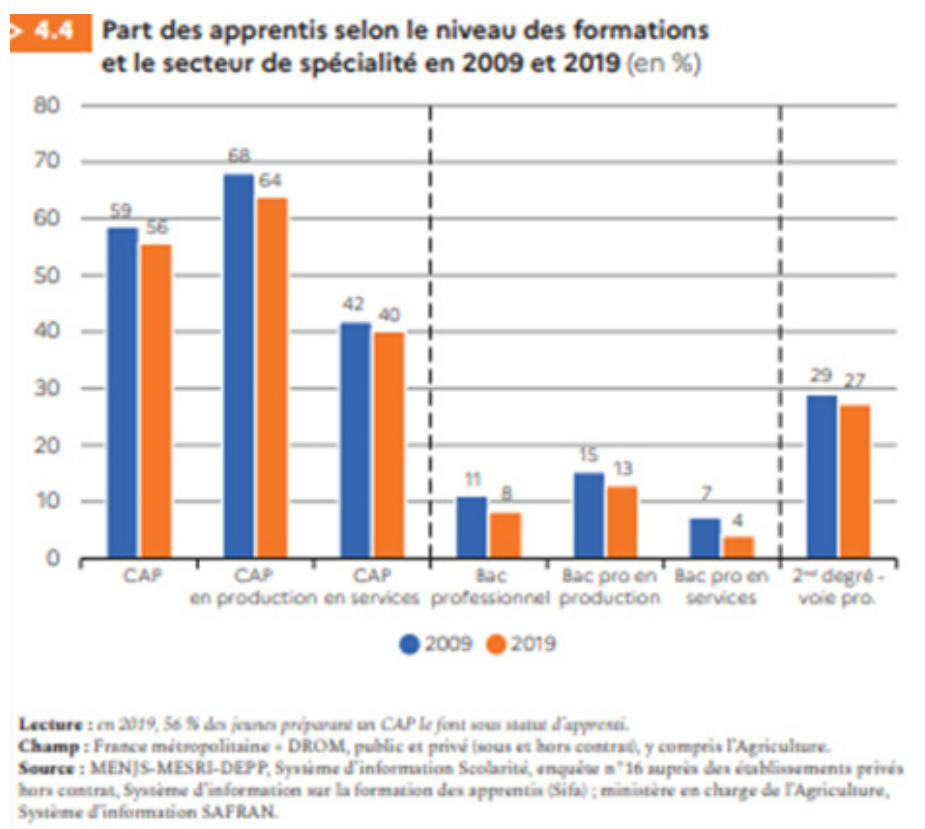


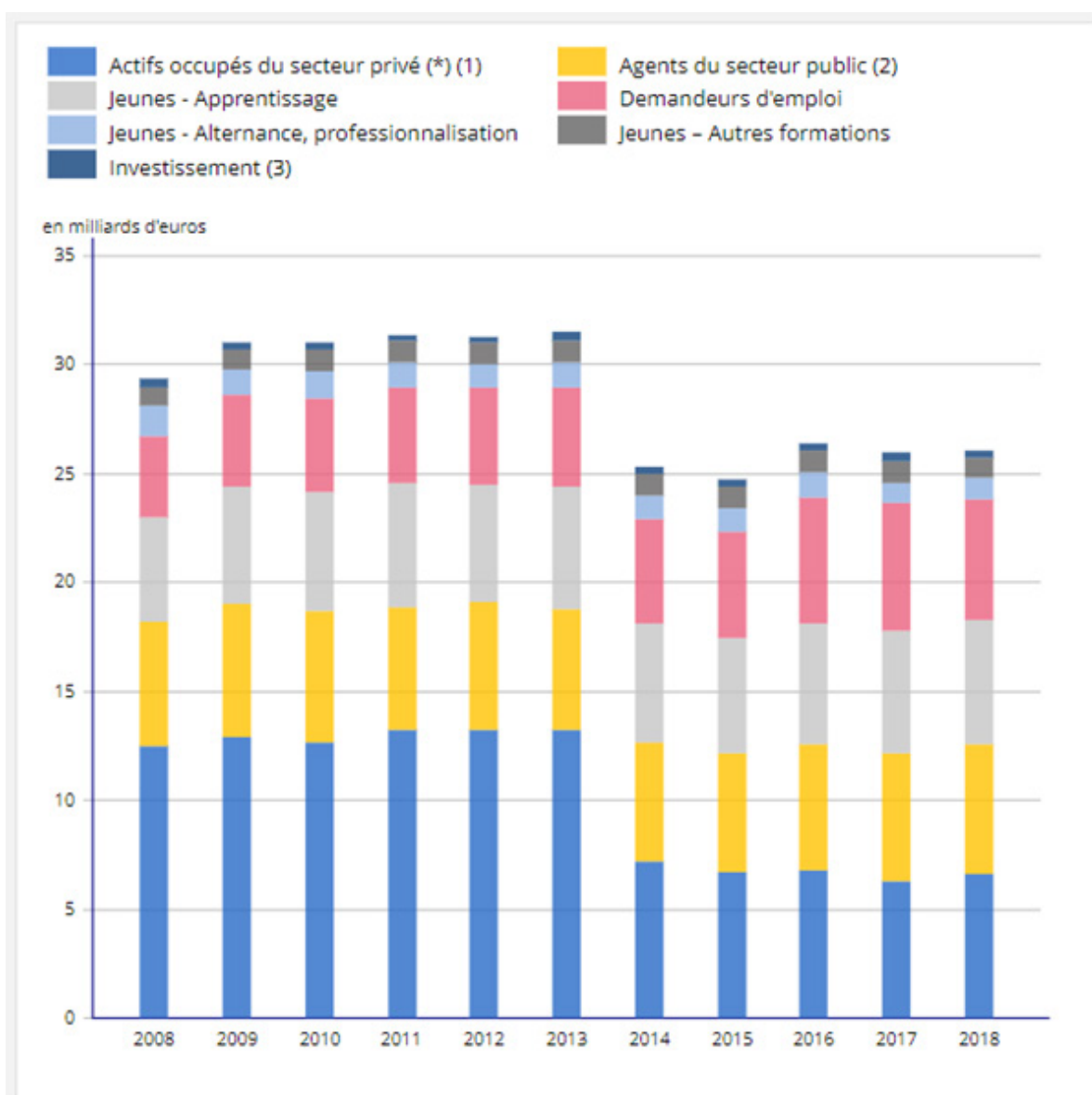
Figure 10. Part des apprentis selon le niveau des formations et le secteur de spécialité en 2009 et 2019 (en %)

	1995-96	2000-01	2005-06	2010-11	2015-16	2016-17	2017-18	2018-19	2019-20	Part des femmes 2019-20 (en %)	Évolution 2005-2019 (en %)
BTS / BTSA	12 539	27 800	35 345	49 965	60 095	62 830	67 401	72 608	79 226	35,0	+ 124,2
DUT	2 067	4 285	4 717	5 548	5 918	6 378	6 900	7 669	8 144	38,7	+ 72,7
Autres niveau III	667	3 468	4 171	6 561	7 304	7 118	7 899	8 274	8 490	47,6	+ 103,5
Total niveau III ¹	15 273	35 553	44 233	62 074	73 317	76 326	82 200	88 551	95 660	36,4	+ 116,7
Licence	56	692	5 392	11 943	16 612	18 122	20 009	20 907	22 851	44,4	+ 323,8
Maîtrise	577	1 837	1 489								
Autres niveau II	2 196	6 919	8 182	7 246	8 043	8 483	9 731	10 675	16 655	52,8	+ 103,6
Total niveau II ¹	2 829	9 448	15 063	19 189	24 655	26 605	29 740	31 582	39 506	47,9	+ 162,3
Diplômes d'ingénieurs	1 734	4 644	7 153	12 706	19 620	20 901	22 544	24 396	25 602	18,6	+ 257,9
DESS	193	1 162	411								
Master			2 999	9 522	14 907	16 165	17 816	19 394	21 463	54,1	+ 615,7
Autres niveau I	21	379	778	7 914	11 514	12 457	14 004	15 877	21 415	49,5	+ 2 652,6
Total niveau I ¹	1 948	6 165	11 341	30 142	46 041	49 523	54 364	59 667	68 490	39,4	+ 503,8
Total	20 050	51 106	70 637	111 405	144 013	152 454	166 304	179 800	203 846	39,6	+ 186,0

Le développement de l'apprentissage se fait surtout désormais dans l'enseignement supérieur, où il s'est fortement implanté depuis une bonne vingtaine d'années, comme on peut le voir dans le tableau précédent⁶.

La formation continue

Les sommes consacrées à la formation continue sont restées globalement stables au cours de la décennie qui vient de s'écouler⁷. La différence observée sur la figure II entre les années 2013 et 2014 tient à une rupture de série statistique : à partir de 2014, les sommes attribuées à la formation continue des actifs en emploi du secteur privé ne comprennent plus les dépenses directes des entreprises.



* : rupture de série. À partir de 2014, la dépense pour la formation continue des actifs occupés du secteur privé ne comprend plus les dépenses directes des entreprises.

(1) : une partie des dépenses bénéficiant aux jeunes sont classées avec les actifs occupés (plan de formation,...) ou les personnes en recherche d'emploi (jeunes bénéficiant de l'Aref, ...).

(2) : hors investissement.

(3) : y compris l'investissement pour les agents du secteur public.

Champ : France.

Source : Dares.

Ces dépenses sont réparties de façon très inégale en fonction de la taille des entreprises et du niveau de qualification des travailleurs. L'effort porte surtout sur les travailleurs les plus qualifiés et sur les travailleurs des plus grosses entreprises, comme le montrent les tableaux suivants⁸.

Taux d'accès à la formation professionnelle	
10-19 salariés	25,1
20-49 salariés	29,2
50-249 salariés	40,8
250-499 salariés	58,4
500-999 salariés	63,0
1000-1999 salariés	63,5
2000 salariés et plus	63,0
Ensemble	48,3

Accès des salariés à la formation professionnelle en fonction de la taille de l'entreprise (en pourcentage)

Situation sur le marché du travail*	
En emploi	51
Au chômage	34
Inactifs non retraités	9
Diplôme	
Diplôme du supérieur long	65
Diplôme du supérieur court	58
Baccalauréat	44
CAP-BEP	33
Brevet des collèges	25
Aucun diplôme ou certificat d'études primaires	15
Ensemble	41

Personnes âgées de 18 à 64 ans sorties de formation initiale ayant suivi au moins une formation professionnelle continue au cours des 12 derniers mois. *À la date de l'enquête.

Source : Insee - Données 2016 - © Observatoire des inégalités

Accès à la formation professionnelle en fonction de la situation sur le marché du travail et du diplôme (en pourcentage).

- Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, « État de l'Enseignement supérieur, de la Recherche et de l'Innovation en France n°14 », 2020. https://publication.enseignementsup/recherche.gouv.fr/eedr/FR/T260/l_apprentissage_dans_l_enseignement_superieur/
- Insee, « Dépenses pour la formation continue. Données annuelles de 2008 à 2018 », décembre 2020. <https://www.insee.fr/fr/statistiques/2388091>
- Observatoire des inégalités, « Les inégalités d'accès à la formation professionnelle », mai 2019. <https://www.inegalites.fr/Les-inegalites-d-acces-a-la-formation-professionnelle>

Considérations prospectives

Dans le cadre de cet exercice, la question sous-jacente qui a conduit au choix de cette variable « Démographie du travail (vieillesse, qualification, formation...) » est de savoir si les développements de l'intelligence artificielle peuvent éventuellement être influencés par des pénuries de main-d'œuvre, en particulier parmi les plus qualifiés. Les hypothèses ont été formulées en ce sens.

Hypothèses

Hypothèse 1.

Dans un contexte où les carrières se prolongent, où le recours à l'immigration est facile (en particulier pour les métiers en tension), où les fonds alloués aux formations (initiales et continues) se maintiennent, il n'existe aucune tension sur le marché de l'emploi. La main-d'œuvre est abondante et formée.

Hypothèse 2.

Le prolongement des carrières s'est révélé illusoire, les travailleurs vieillissants préférant partir en retraite avec une décote, quitte à compléter leurs retraites avec des emplois occasionnels à temps partiels, plus ou moins légaux. L'apprentissage s'est développé, mais surtout pour la poursuite d'études supérieures. En revanche, la formation continue s'adresse toujours aux mieux formés des plus grosses entreprises. Compte tenu du climat social et économique, l'immigration a été réduite, même pour les métiers en tension. Il existe des tensions fortes sur le marché de l'emploi, malgré un taux de chômage important.

Hypothèse 3.

L'emploi est devenu une priorité nationale, dans un contexte où la reconstruction écologique s'est traduite par un fort volontarisme technologique et un choix de relocalisation de certaines productions. L'accent est donc mis sur la formation du début à la fin des carrières, afin notamment de les prolonger de façon acceptable. Le recours à une immigration choisie est possible pour les métiers en tension, y compris en dispensant des formations complémentaires aux nouveaux arrivants.

Organisation du travail et statut des travailleurs

J. Clerté

Définition de la variable

Cette fiche vise à explorer les évolutions possibles des modes d'organisation du travail, des statuts des travailleurs, ainsi que les conséquences sur les collectifs de travail. Durant ces vingt dernières années, l'essor des TIC – et notamment de la collecte des données, accompagnée du développement des algorithmes – a donné lieu à l'émergence de nouvelles organisations du travail, comme les plateformes. Avec la crise sanitaire, ces développements technologiques sont également venus faciliter l'instauration du travail à distance à grande échelle. Parallèlement, la forte automatisation de certains secteurs a changé les modes de production et d'organisation des tâches.

Définitions des tendances observées

Robotisation industrielle et automatisation des services

La croissance de la robotisation dans les milieux industriels se mesure généralement par l'évolution du nombre de robots dans le secteur de l'industrie. L'automatisation touche aussi progressivement le secteur tertiaire : on peut citer le déploiement des distributeurs automatiques de billets pour le secteur bancaire ou les caisses automatiques dans le commerce^{1,2}. Les estimations sur l'impact destructeur ou créateur d'emplois varient selon les organismes. Au-delà de cet aspect, cette évolution a aussi un impact sur les formes de travail, sa localisation, ou le niveau d'autonomie des opérateurs.

1. C.B. Frey, M.A. Osborne, « The future of employment: How susceptible are jobs to computerisation? », *Technological forecasting and social change*, 2017.

2. M. Arntz, T. Gregory and U. Zierahn, « The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis », *OECD Social, Employment and Migration Working Papers*, No. 189, OECD Publishing, 2016. <http://dx.doi.org/10.1787/5j1z9h56dvq7-en>

Évolution de la structure et des statuts d'emploi

Les statuts d'emploi distinguent les personnes à leur compte des salariés et, au sein de ceux-ci, ceux qui travaillent respectivement dans les secteurs privé, public ou semi-public, ainsi que les différentes natures de contrat possible (CDI, CDD, intérim). Les catégories socioprofessionnelles, elles, permettent de classer une population selon l'activité professionnelle (agriculteur, ouvrier, etc.). Sous l'effet de l'automatisation et de la mondialisation, celles-ci ont subi de fortes mutations.

Impact des technologies de l'information et de la communication sur les organisations du travail

Depuis les années 90, les technologies de l'information et de la communication ont connu une explosion qualitative et quantitative. Le recours à ces nouvelles technologies et l'accès à un nombre de données de plus en plus important a donné lieu à l'apparition de deux évolutions majeures du monde du travail : les plateformes numériques et le travail à distance.

Affaiblissement et transformation des organisations collectives et des collectifs de travail

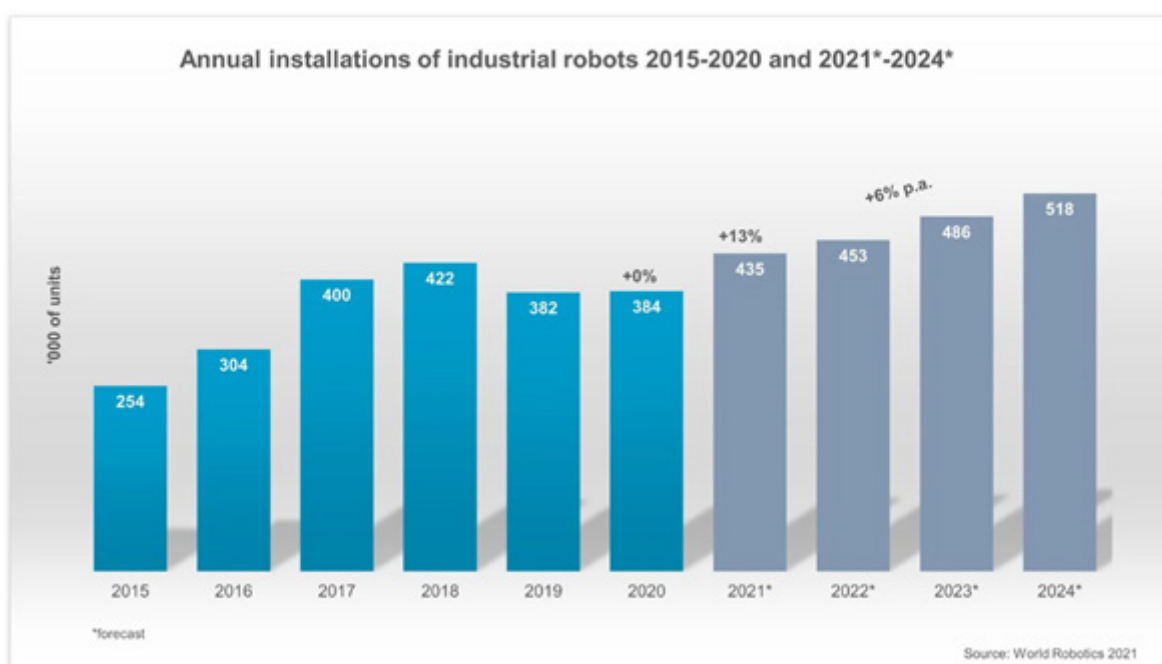
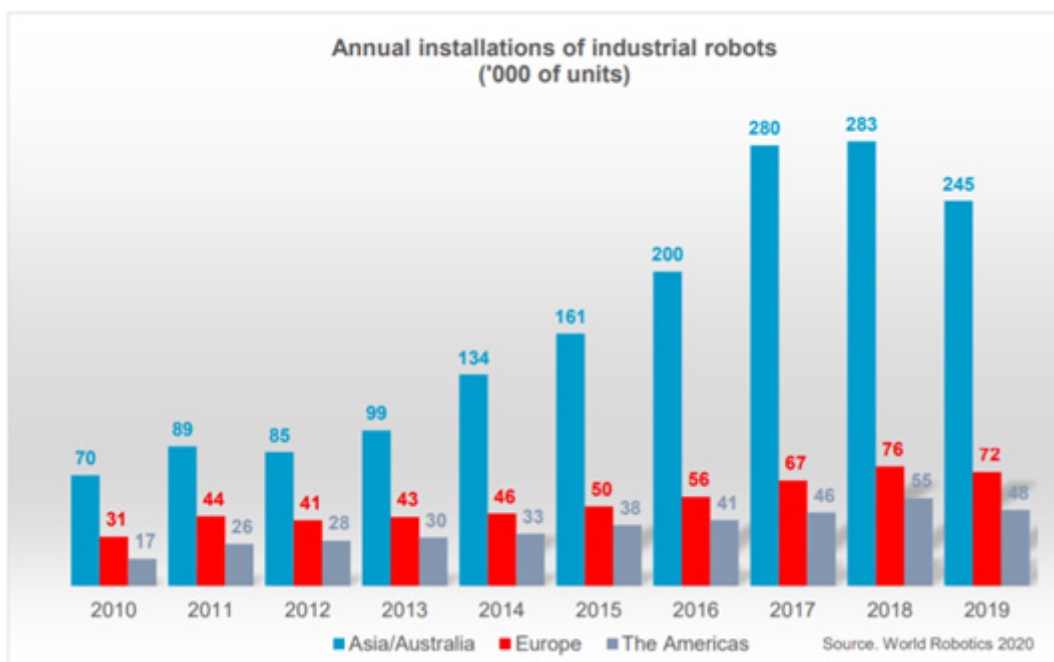
Les transformations de la structure des emplois et des organisations du travail ont eu un impact fort sur les organisations collectives de travailleurs. Cette tendance se mesure notamment par l'évolution du taux de syndicalisation. Affaiblies dans un premier temps, les organisations collectives tendent à renaître sous de nouvelles formes afin de s'adapter au contexte d'un nouveau monde du travail.

Rétrospective et tendances

Robotisation dans l'industrie et automatisation des services

En Europe, tout comme dans le reste du monde, la croissance de l'équipement en robots industriels n'a fait que croître : les données statistiques de l'IFR³ montrent une augmentation continue depuis 10 ans.

3. IFR : International Federation of Robotics.



© World Robotics

L'automatisation des secteurs tertiaire est plus difficilement quantifiable. Cependant, on peut citer plusieurs exemples de secteurs ou d'emplois directement touchés par ce phénomène : dans le secteur du commerce, nombreux sont les articles annonçant la prochaine disparition du métier de caissière. Même si ce phénomène ne s'est vérifié que sur le long terme^{4,5} – les clients n'étant d'abord pas favorables à leur usage –, on a constaté une accélération de l'intégration des caisses automatiques durant la crise sanitaire.

4. <https://www.lefigaro.fr/conso/les-caisses-sans-caissier-de-plus-en-plus-repandues-dans-les-supermarches-20190715>

5. <https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/na-49-automatisation-emploi.pdf>

Dans les banques, l'implantation des DAB (distributeurs automatiques de billets), qui sont passés de 5 000 sur le territoire en 1983 à près de 60 000 fin 2013, a conduit à une baisse des effectifs et à une modifications des tâches. Plus récemment, on a vu émerger l'utilisation de robots dans le secteur de la restauration, notamment pour faire face à la pénurie de main-d'œuvre⁶.

Les conséquences de cette révolution technologique sur le monde du travail sont multiples. À l'heure actuelle, l'augmentation de la robotisation ne semble pas être corrélée à une destruction d'emplois (voir tableau ci-dessous). Cependant, le débat persiste entre malthusiens et libéraux⁷, les uns prévoyant une prochaine vague de destruction d'emplois liée à l'automatisation des tâches dans les secteurs industriels comme dans ceux des services ; les autres estimant que les emplois détruits seront remplacés par de nouveaux métiers, selon le principe schumpeterien de « destruction créatrice ».

	Robots pour 1000 travailleurs	Taux de chômage (en %)
Corée du Sud	63,1	3,7
Allemagne	30,9	3,4
Japon	30,3	2,8
Suède	22,3	6,7
USA	18,9	4,4
France	13,2	9,2

Sources :

- Taux de chômage harmonisé – OCDE 2017
- IFR : International Federation of Robotics

Si la vague de destruction d'emplois n'est pas envisagée à moyen terme (la révolution NBIC⁸ et les enjeux environnementaux requérant de nouvelles compétences), l'automatisation généralisée du travail a tout de même un fort impact sur les organisations du travail, entraînant :

- **Une métamorphose des catégories socio-professionnelles** : les ouvriers, qui représentaient encore 30 % des actifs dans les années 80, en représentent aujourd'hui moins de 20 %. Encore plus fortement touchés, les agriculteurs ne forment plus que 1,5 % de l'emploi⁹. Parallèlement, les professions intermédiaires et les cadres supérieurs ont fortement augmenté, représentant 46 % des emplois.
- **Une organisation néo-tayloriste du travail** : le recours à la robotisation et à l'automatisation, bien que pouvant contribuer à supprimer certaines tâches pénibles ou répétitives, peut déposséder le salarié de la finalité de son activité. L'organisation du travail tend à se standardiser, de façon à optimiser le fonctionnement des robots et logiciels. L'autonomie des travailleurs en est fortement réduite. Plusieurs exemples viennent à l'esprit pour illustrer ce point : on peut citer les travailleurs de *call centers*

6. https://www.lexpress.fr/actualite/monde/asie/baisse-de-la-natalite-vieillesse-le-japon-mise-sur-les-robots-pour-remplacer-les-humains_2154134.html

7. A. Bueno, « Futur, notre avenir de A à Z », Flammarion, 2020.

8. NBIC : nanotechnologies, biotechnologies, informatique et sciences cognitives.

9. <http://www.observationsociete.fr/categorie-sociales/donnees-generales/comment-evoluent-les-categorie-sociales-en-france.html>

dont le travail est prescrit par une feuille de route stricte dictée par un logiciel de CRM¹⁰, les systèmes de commandes vocales utilisés dans le secteur de la logistique qui imposent un rythme élevé et ininterrompu aux opérateurs¹¹, ou encore le cas des travailleurs du clic entraînant des intelligences artificielles, dont les tâches répétitives et fractionnées perdent sens. Ces éléments avaient été objectivés dans une note de la DARES publiée en 2014¹².

- **Autonomie et réduction des risques physiques** : au contraire, dans certains cas, l'automatisation peut constituer un gain en autonomie ou encore une réduction, voire une suppression, des tâches dangereuses ou répétitives. C'est le cas dans l'industrie avec l'utilisation de cobots, qui prennent la relève des salariés pour des tâches répétitives

Polarisation de la structure et des statuts d'emploi

Depuis 30 ans, on assiste à une transformation profonde de la structure des emplois. Sous l'effet de la robotisation et de la mondialisation, les emplois du secteur secondaire se sont effondrés¹³, progressivement remplacés par ceux du tertiaire. Les emplois agricoles ont pour leur part connu la plus forte baisse.

Depuis les années 2000, ce sont surtout les emplois de service à la personne (santé, action sociale, services aux particuliers) qui ont fortement augmenté.

Graphique 2
Évolution en milliers de l'emploi par domaine professionnel entre 1982-1984 et 2012-2014



Champ : actifs occupés de France métropolitaine.

Source : enquêtes Emploi, Insee, moyenne annuelle sur les années 1982 à 1984 et 2012 à 2014 ; traitement Dares.

Selon les statistiques nationales, l'érosion des emplois peu qualifiés s'est faite davantage en faveur des professions très qualifiées que des professions peu qualifiées. Selon la note de France Stratégie¹⁴ publiée sur le sujet en 2020, la qualification peut se définir selon le niveau de diplôme, le salaire ou la catégorie socio-professionnelle. Elle conclue que la

10. INRS, « [Les centres d'appels téléphoniques](#) », dossier web.

11. INRS, [ED 135](#).

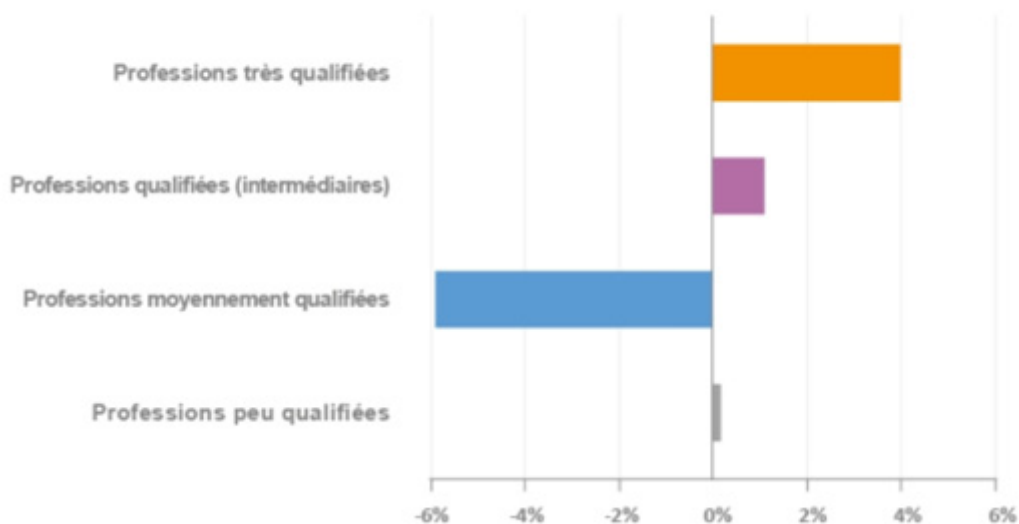
12. <https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/2014-049.pdf>

13. <https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/2017-003.pdf>

14. <https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/fs-2020-na-98-polarisation-marche-travail-decembre-ok.pdf>

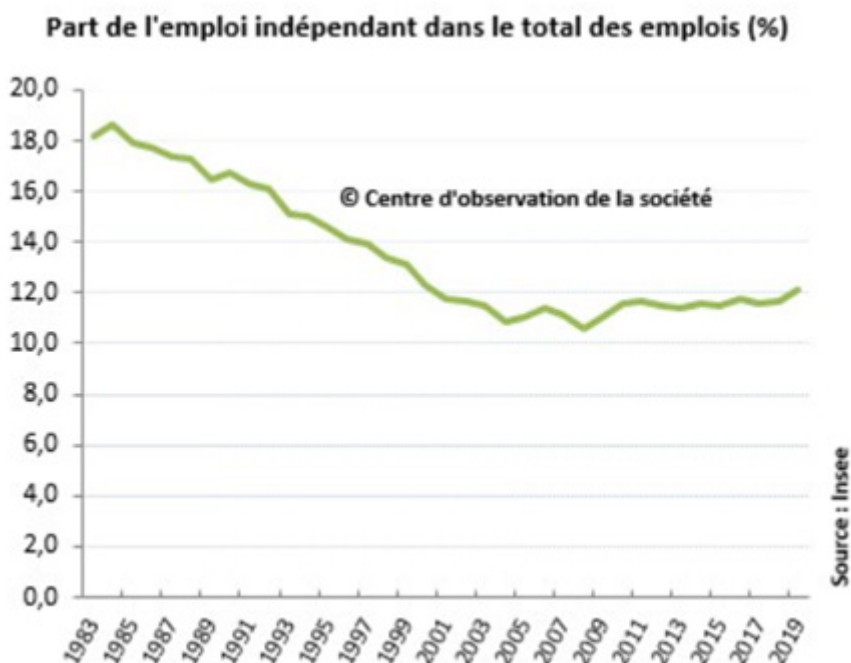
polarisation du marché du travail en France est marquée par un déplacement de l'emploi vers les qualifications les plus élevées, s'accompagnant d'une érosion accentuée par la crise de 2008 des métiers situés au milieu de la distribution des qualifications. Cependant, on constate que la part des emplois peu qualifiés se maintient – du fait de la croissance des emplois d'aide à la personne notamment.

Variation de la part des qualifications dans l'emploi entre 1996 et 2017, en points de pourcentage



Source : Insee-DGI, enquête sur les revenus fiscaux et sociaux rétropolés, 1996-2017.

Le groupe des travailleurs non-salariés est pour sa part en légère progression, du fait de la création en 2009 du statut d'auto-entrepreneur et des emplois d'exécution payés à la tâche, par exemple pour les plateformes de livraison de repas.



De façon globale, on assiste surtout à une polarisation des niveaux de vie entre les métiers de cadres et professions intellectuelles supérieures qui ont crû fortement, et les métiers d'employés peu qualifiés qui se sont maintenus, portés notamment par l'augmentation des effectifs d'aides ménagères, aides à domicile et assistantes maternelles – emplois qui ne sont ni délocalisables ni automatisables, et pour lesquels la demande est particulièrement dynamique.

Cette polarisation se lit au travers du creusement des écarts de revenus, qui s'est remis à progresser depuis 2009. En France, entre 2008 et 2018, la différence entre les valeurs des salaires (les 10 % plus élevés et les 10 % les plus bas) est passée de 2 065 à 2 486 euros, soit une augmentation de l'écart de 420 euros¹⁵ entre les moins bien payés et ceux qui le sont le plus.

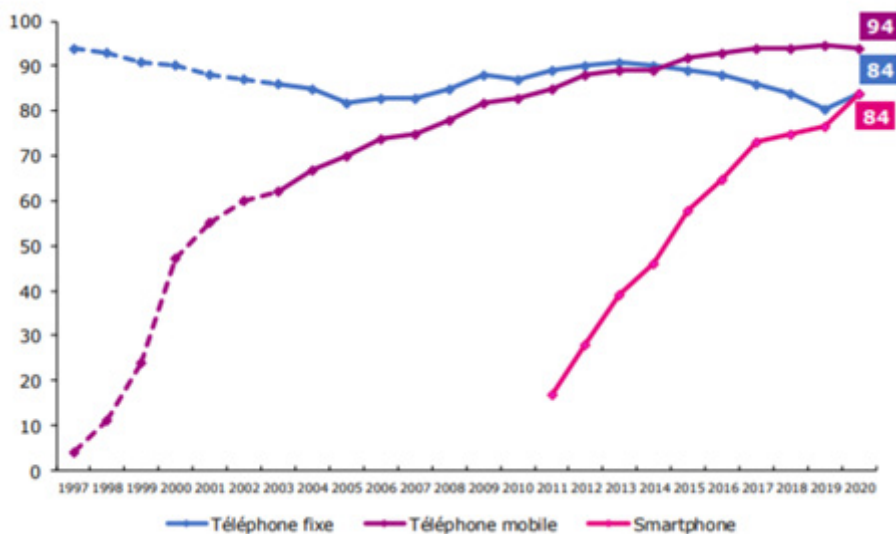
Le livre de Jérôme Fourquet et Jean-Laurent Cassely illustre ainsi la tendance actuelle de la classe moyenne à se scinder en deux, avec une segmentation par le haut pour une part de cette population pouvant accéder à des biens de consommation haut de gamme, et un segment inférieur qui amorce un décrochement – la France des *discount*, des promos et des bons plans¹⁶.

Impact des TIC et nouvelles formes de travail et d'emplois

Depuis, les années 90, le trafic Internet n'a cessé de croître (voir le rapport annuel Cisco¹⁷), de même que la vitesse de connexion. Ce trafic de données s'est vu encore augmenter avec le déploiement massif des ordinateurs, smartphones et tablettes. Cette révolution technologique a donné lieu à deux transformations majeures des organisations du travail.

La plateforme du secteur des services

L'émergence des plateformes de travail est liée à la croissance forte de l'équipement des ménages en objets connectés et à leur accès à Internet¹⁸, à l'origine de l'explosion des achats en ligne. On peut voir le taux d'équipement numérique (en pourcentage) ci-dessous :

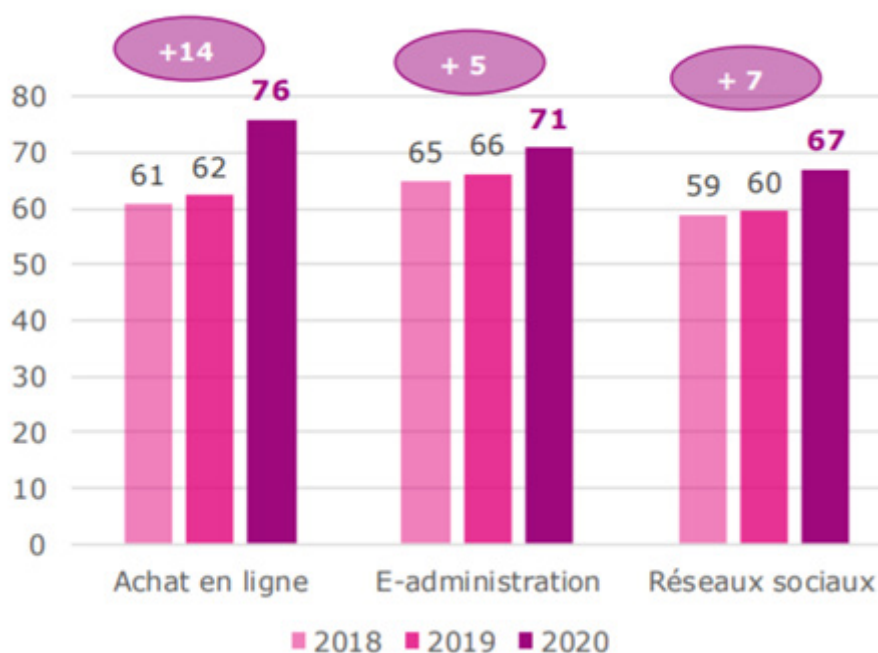


15. « Les inégalités de salaire remontent », <https://www.inegalites.fr/Les-inegalites-de-salaire-remontent>

16. J. Fourquet, J.L. Cassely, « La France sous nos yeux », 2021.

17. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>

18. <https://www.credoc.fr/publications/barometre-du-numerique-edition-2021>



Progression des achats en ligne, des réseaux sociaux, de l'e-administration (part des personnes interrogées).

Elles ont cependant connu leur vrai essor à la faveur de la récession de 2008-2009, en créant des opportunités d'emploi tout en permettant l'accès à des services à bas coût. Leur essor s'est ainsi accéléré depuis 2012. Elles bénéficient par ailleurs d'investissements dans l'économie numérique, qui leur permettent de développer de nouveaux modèles économiques dont la rentabilité n'est bien souvent pas avérée, mais qui déstabilisent des marchés.

Selon l'OIT¹⁹, entre 2010 et 2020, le nombre de plateformes a quintuplé, passant de 128 à 611 dans les pays du G20, soit 79 % des plateformes en activité dans le monde. On distingue deux types de travailleurs de ce nouveau modèle d'entreprises : ceux qui sont directement embauchés par la plateforme, et ceux dont le travail est intermédié par les plateformes (travailleurs indépendants). Leur nombre est difficile à estimer, cependant, un communiqué du ministère du Travail les estime à 100 000²⁰.

Ce statut de travailleurs de plateformes se distingue des professions indépendantes classiques par leur dépendance économique (certains parlent de travailleurs indépendants économiquement dépendants). Dans son rapport, l'OIT propose un portrait des travailleurs de plateformes et de leurs conditions de travail, sur la base d'une enquête conduite entre 2017 et 2020 auprès de 12 000 travailleurs indépendants issus de 100 pays. Souvent jeunes, peu diplômés, souvent immigrés, leur temps de travail est très élevé (jusqu'à 70 h par semaine), leurs revenus sont faibles et précaires. Ils ne bénéficient que de très peu de protection sociale, leur statut ne leur ouvrant aucun droit. Enfin, malgré leur statut d'indépendant, leur activité dépend étroitement du système de notation et d'évaluation algorithmique des plateformes.

19. OIT, « Digital platforms and the world of work in G20 countries: Status and Policy Action », 2021.

20. Ministère du Travail, de l'Emploi et de l'Insertion, « Les travailleurs indépendants des plateformes de mobilité pourront désigner leurs représentants en 2022 », <https://travail-emploi.gouv.fr/actualites/presse/communiqués-de-presse/article/les-travailleurs-indépendants-des-plateformes-de-mobilité-pourront-désigner>

En termes d'organisation du travail, les plateformes ont eu trois conséquences notables :

- Le management algorithmique vient transformer les organisations du travail traditionnelles en permettant une évaluation du travail et un calcul de sa répartition sur la base d'indicateurs de performance et de systèmes de notation. Progressivement, ce nouveau mode de management se diffuse au sein des entreprises traditionnelles et vient modifier le monde du travail dans sa globalité.
- Les plateformes reportent les coûts d'investissement dans les équipements de travail sur les travailleurs, qui en ont désormais la charge. Elles se déchargent également de la responsabilité légale de l'employeur en matière de santé et sécurité au travail.
- Enfin, elles contribuent à polariser le marché du travail entre salariés bénéficiant de droits collectifs et d'une protection sociale, et indépendants moins bien protégés et dépourvus de dispositifs de représentation collective.

Le travail à distance

Depuis 10 ans, les technologies de l'information et de la communication ont facilité le travail à distance, en rendant accessible un grand nombre d'outils de travail numériques hors des locaux de travail²¹.

En 2020, c'est l'accès à Internet et la puissance du trafic de données qui a permis le maintien de l'activité économique malgré les mesures de confinement des populations liées au contexte de pandémie. Cette mesure a d'ailleurs conduit à une augmentation du trafic de données de plus de 30 %²².

Ce nouveau mode de travail, , induit une transformation de l'organisation du travail et des statuts de travailleurs à plusieurs niveaux²³ :

- **Une fracture entre travailleurs** : cette évolution ajoute un nouveau niveau d'inégalités entre travailleurs qualifiés et peu qualifiés. Les emplois qualifiés, majoritairement positionnés dans le secteur tertiaire et la tech, sont facilement télétravaillables, tandis que les moins qualifiés – plus présents dans les secteurs des services et aide à la personne et de la logistique – ne peuvent pas en bénéficier. Les conditions de travail de ces deux groupes se distinguent aussi de plus en plus, les premiers bénéficiant dans l'ensemble, grâce au travail à distance, de plus d'autonomie, les seconds restant soumis à des contraintes organisationnelles plus strictes en termes d'horaires, de lieu de travail et de pression managériale.
- **Des inégalités entre télétravailleurs** : une autre fracture tend à se renforcer entre générations. Les salariés les plus jeunes bénéficient d'une part de moins bonnes conditions de travail à domicile que les plus âgés (ayant généralement de plus grands logements) ; et ils ont moins facilement la possibilité de se créer un réseau comme les plus anciens ou de se familiariser avec les *process* de travail, et peuvent ainsi moins facilement défendre leurs intérêts.
- **Une nouvelle manne d'emplois délocalisables** : selon, une étude américaine de 2013 sur la « délocalisabilité » (*offshorability*) des emplois²⁴, plus les emplois sont qualifiés et bien payés, plus ils sont délocalisables. Les auteurs estiment que 21 à 27 % des emplois aux États-Unis sont délocalisables, mais cette proportion grimpe à 35

21. <https://dares.travail-emploi.gouv.fr/sites/default/files/pdf/2018-029.pdf>

22. https://www.arcep.fr/uploads/tx_gspublication/rapport-etat-internet-edition-2021-juil2021-presentation.pdf

23. Marc Malenfer, « Deux jours par semaine, et après ? », Futuribles.

24. <https://www.jstor.org/stable/10.1086/669061>

voire 40 % concernant les emplois nécessitant un diplôme. Jusqu'à présent épargnés par les délocalisations vers des pays à faible coût de main-d'œuvre, les emplois qualifiés des techs pourraient s'avérer menacés. Au Royaume-Uni, les prémices de cette délocalisation d'emplois de cols blancs commencent à se faire sentir et se traduisent par une baisse moyenne de 25 % des revenus des professionnels devenus indépendants²⁵.

- **Menace de désengagement des travailleurs et affaiblissement des collectifs de travail** : la distanciation physique des salariés tend à affaiblir le lien d'adhésion aux valeurs des entreprises, déjà en cours depuis quelques années. L'affaiblissement de la cohésion d'équipe menace également la capacité d'innovation. Pour prévenir le risque de désengagement, certaines entreprises ont mis en place des systèmes de surveillance de leurs salariés à distance; d'autres ont voulu les obliger à un retour sur site à plein temps. Les organisations de travailleurs en font également les frais. Les organisations de représentation du personnel sont confrontées à une démultiplication d'attentes variables et contradictoires des salariés concernant le travail à distance, qui rend complexe la définition d'une ligne stratégique. Ils ne sont par ailleurs pas en mesure d'effectuer pleinement l'analyse des risques, qui fait partie de leur prérogative, lorsque le travail se déroule en grande partie dans la sphère privée. Déjà affaiblis, ils rencontrent de nouvelles difficultés pour mobiliser les travailleurs et exercer leur mission, et voient ainsi leur pouvoir s'amenuiser.

Évolution et transformation des collectifs de travail

Depuis 50 ans, les taux de syndicalisation n'ont fait que décroître, pour se stabiliser depuis une dizaine d'années autour de 10 %, avec un grand écart entre secteur privé et public, ayant respectivement un taux de 8,4 % et 19,1 %²⁶. Cette évolution est étroitement liée à l'évolution structurelle des emplois depuis les années 50.



25. Vidéo de Xerfi : https://www.xerficanal.com/economie/emission/Alexandre-Mirlicourtois-Le-teletravail-accelere-l-uberisation-des-cadres_3749114.html

26. <http://www.observationsociete.fr/travail/donnees-generales-travail/une-france-tres-peu-syndiquee.html>

La désindustrialisation, la stabilisation du taux de chômage à un niveau élevé et la précarisation des emplois (croissance du nombre de CDD) ont conjointement contribué à la chute de la syndicalisation.

Les nouvelles formes de travail et d'emploi décrites précédemment tendent pour leur part à individualiser les tâches et les situations de travail, rendant plus complexe la constitution d'organisations collectives. En ce qui concerne les travailleurs de plateformes, leur droit à la négociation collective est même remis en question par la nature de leur statut.

Face à ces difficultés, de nouvelles formes d'organisations voient le jour²⁷. Les travailleurs commencent par se constituer en collectifs, dont l'objectif principal est de faire porter leur voix au travers des réseaux sociaux et auprès des institutions politiques, leurs donneurs d'ordres refusant d'amorcer le dialogue avec eux (cf. le collectif CLAP²⁸).

Le modèle des coopératives, pour sa part, leur permet d'aller plus loin en tentant de reprendre le pouvoir face à des entreprises restées sourdes à leurs revendications²⁹. En décembre 2020, 21 collectifs s'étaient déjà constitués en coopératives afin de lancer leur propre service de livraison. Les coursiers bordelais ont initié le mouvement en 2017, et le modèle se développe depuis dans plusieurs villes de France, telles que Dijon, Besançon ou Strasbourg. Ces structures permettent aux travailleurs d'accéder au statut de salarié et ainsi de socialiser leurs revenus, de prendre directement part à leur gestion, et de remettre les problématiques de prévention et de protection au centre des préoccupations.

De par leur fonctionnement, elles permettent de contourner les règles institutionnelles pour répondre aux besoins de sécurité des travailleurs. En cela, elles dépassent les logiques traditionnelles d'action. Elles proposent un nouveau modèle d'entreprise de travailleurs autonomes libérés du lien de subordination, permettant d'aller au-delà de la dichotomie d'un monde économique scindé entre salariés subordonnés et indépendants précaires³⁰.

Les coopératives sont d'ailleurs considérées comme une solution crédible pour parvenir à réguler les plateformes numériques de travail (Uber, Deliveroo, etc.) – comme cela est suggéré dans le rapport Frouin³¹ paru en décembre 2020.

27. J. Clerté, M. Héry, « Économie du digital : quand les travailleurs s'organisent pour préserver leurs conditions de travail », Travailler au Futur.

28. Collectif des Livreurs Autonomes de Paris.

29. « Livraison à vélo : comment les coopératives s'organisent », Les Échos, 26 décembre 2020, <https://www.lesechos.fr/industrie-services/conso-distribution/livraison-a-velo-comment-les-cooperatives-sorganisent-1276580>

30. « Coopératives de travailleurs autonomes : du jeu dans les règles au jeu sur les règles, éléments de comparaison Belgique, France et Italie », Futuribles, avril 2021. <https://www.futuribles.com/fr/bibliographie/notice/cooperatives-de-travailleurs-autonomes-du-jeu-dans/>

31. <https://www.vie-publique.fr/rapport/277504-reguler-les-plateformes-numeriques-de-travail-rapport-frouin>

Synthèse : tendances lourdes et interrogations

Tendances lourdes

- Démocratisation du télétravail
- Poursuite de la tertiarisation des emplois
- Extension de l'automatisation au secteur tertiaire

Interrogation majeures

- Extension de la plateformes aux activités du tertiaire
- Délocalisation des emplois de cols blancs
- Quel avenir pour le statut de salariés ?
- Évolutions des collectifs de travail : vers une disparition ou une renaissance sous d'autres formes

Hypothèses

Hypothèse 1. Organisations du travail hétérogènes et inégalités statutaires

Cette première hypothèse se construit sur une tendance à la démocratisation du télétravail pour l'ensemble des emplois de cols blancs, renforçant les inégalités entre les travailleurs. Cette tendance est soutenue par les décideurs politiques pour des raisons d'enjeux environnementaux ou sanitaires, et par les chefs d'entreprises qui y voient une opportunité de réduction de leur charge immobilière.

Les plus qualifiés sont de plus en plus autonomes, tout en bénéficiant d'un système de protection sociale du fait de leur statut de salariés. Les emplois de première et seconde lignes ne peuvent pas bénéficier du télétravail et sont soumis à des contrats courts, le plus souvent externalisés ou plateformes. La polarisation entre emploi qualifié et peu qualifié s'accroît.

Hypothèse 2. Tous prestataires de services (fin du salariat)

Dans cette hypothèse, sous l'effet des évolutions technologiques et du développement de l'IA, la plateformes de l'économie se poursuit et s'étend aux professions intellectuelles. Certaines fonctions administratives sont progressivement remplacées par des IA. Les entreprises, cherchant à être de plus en plus agiles pour rester dans la course à l'innovation, recourent à des contrats courts ou à une main-d'œuvre externalisée embauchée sur la

base de contrats de projet. Certains emplois hautement qualifiés et facilement externalisables se voient soumis à la concurrence de travailleurs étrangers, mettant ainsi sous pression les catégories socio-professionnelles supérieures et entraînant une baisse de leur rémunération.

Hypothèse 3. Flexisécurité et émergence de nouvelles formes d'organisation du travail

Dans cette dernière hypothèse, afin de freiner les effets négatifs sur l'emploi des nouvelles technologies et de préserver le système de protection sociale, l'État met en œuvre un certain nombre de réformes, de façon à ouvrir l'accès à plus de protection aux travailleurs indépendants et aux plus précaires. Parallèlement, les nouvelles formes d'organisations telles que les coopératives prennent de l'ampleur et permettent aux travailleurs indépendants de profiter des mêmes droits et protection que les travailleurs salariés, sans être soumis au lien de subordination. Le modèle d'entreprise libérée se développe parallèlement à celui du télétravail, du fait de la mutation des fonctions de management. Les espaces de coworking se développent également pour permettre aux travailleurs de reconstruire un collectif de travail hors de l'entreprise. Les innovations liées à l'IA contribuent à abaisser les coûts de production et permettent de relocaliser un certain nombre d'industries, en accord avec les problématiques environnementales de plus en plus fortes, engendrant une légère croissance de l'emploi ouvrier.

Automatisation, cobotisation et équipements du travailleur

M. Sarrey

Cette fiche vise à décrire le contexte de la digitalisation du monde du travail. Elle abordera principalement les thèmes de la mécanisation autonome (secteur agricole et industriel), l'automatisation de tâches administratives, de services ou d'aide à la décision (secteur tertiaire), la collaboration Homme-machine (prise en charge partielle des tâches d'un poste de travail), la surveillance des situations de travail et les équipements de travail numériques.

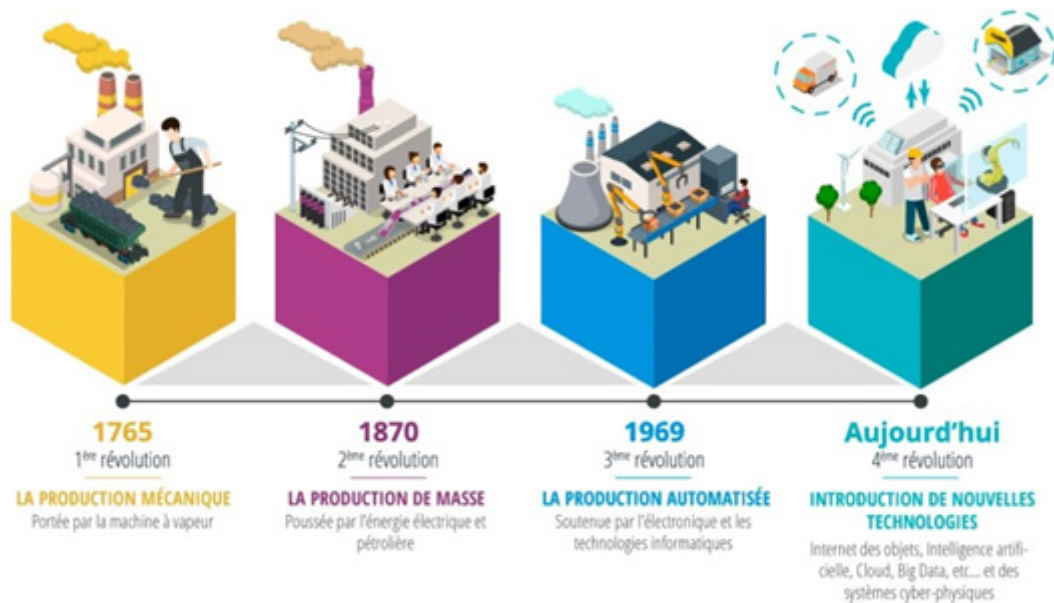
Définition et contexte

L'automatisation des métiers n'est pas nouvelle, elle est à l'œuvre depuis des décennies. Cependant, cette automatisation prend aujourd'hui de nouvelles formes : autrefois cantonnée aux tâches ouvrières dans les usines, elle se retrouve dorénavant dans tous les domaines et pour de nombreuses tâches. Grâce à l'automatisation, les métiers se voient exécutés – tout ou partie – par des automates interagissant dans le monde physique ou cybernétique.

Qu'est-ce que l'automatisation ?

Selon le CNRTL (Centre National de Ressources Textuelles et Lexicales), l'automatisation est la « substitution d'une ou de plusieurs machines à l'Homme pour réaliser de manière automatique un programme déterminé d'opérations »¹. Elle se distingue de la mécanisation par la programmation déterminée des opérations.

1. <https://www.cnrtl.fr/definition/automatisation>



Pour l'industrie, l'automatisation est fléchée « Industrie 3.0 ». Elle a fait suite à la mécanisation électrique (industrie 2.0) dans le courant des années 1970. Cette automatisation de l'industrie est souvent symbolisée par la robotisation des chaînes de production, principalement dans le secteur automobile. Néanmoins, on distingue l'automatisation de la robotisation, car les robots industriels de l'époque s'apparentaient à des automates numériques sans aucun degré d'autonomie.

Dans le monde agricole, la mécanisation du début du XX^e siècle n'a pas initié d'automatisation au sens industriel. C'est l'arrivée de machines autonomes (robots de traite pour la production laitière, robots mobiles agricoles), au XXI^e siècle, qui « automatiseront » le secteur.

Dans les services, l'automatisation remplace l'Homme pour l'exécution de tâches administratives routinières. Elle a recours aux logiciels. Les premières tâches automatisées du domaine tertiaire² sont sans doute celles liées au secrétariat. Nous verrons que cette automatisation « tertiaire » s'étend aujourd'hui à toutes les branches du secteur ; commerce, transport, administration, enseignement, santé, etc. Dans le secteur des services, la nature des tâches remplacées par les systèmes automatiques est principalement d'ordre cognitif.

Ainsi, pour les secteurs tertiaire et primaire, l'automatisation s'apparente à la robotisation.

Qu'est-ce que la robotisation ?

La robotisation est le « remplacement de l'Homme ou d'une machine traditionnelle par une machine entièrement automatique, résultat de cette action »³. Cette définition apporte la notion d'autonomie à la machine, qui devient alors « entièrement automatique », c'est-à-dire capable de prendre des décisions sans intervention humaine. Notons que cette définition s'applique aussi bien aux systèmes physiques (robots agricoles, industriels, de service) qu'aux systèmes cybernétiques (*bot* conversationnel, logiciel d'aide à la décision, *bot* de relation clients, *trading*, etc.). Majoritairement, ces derniers font appel à des programmes apparentés à de l'intelligence artificielle.

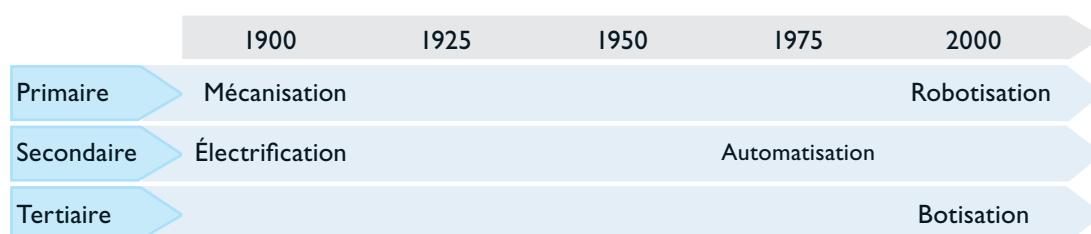
2. <https://www.insee.fr/fr/metadonnees/definition/c1584>

3. <https://www.cnrtl.fr/definition/robotisation>

Depuis quelques années, la robotique « industrielle » s'ouvre vers des domaines où l'on n'imaginait pas intégrer de robots auparavant. Issus du monde de la recherche et des start-ups, de nouveaux robots sont apparus : plus légers, plus simples à intégrer, plus aptes à collaborer avec des opérateurs humains. Il s'agit des cobots, pour « *collaborative robot* ». Ces cobots sont destinés à remplacer une partie des tâches physiques d'un métier. Ne remplaçant qu'une partie des tâches d'un opérateur humain, le cobot devra être capable de s'intégrer dans un environnement prévu pour les personnes, sans engendrer de risques, pour collaborer ou *a minima* partager l'espace avec les opérateurs.

Synthèse de l'automatisation par secteur

Pour cette fiche, nous réunirons sous le vocable « automatisation », la robotisation agricole, l'automatisation industrielle et l'usage de bots et de cobots dans les services.



L'Internet des objets / Internet of Things (IoT)⁴

L'automatisation ne se limite pas à l'exécution automatique de tâches (physiques ou cognitives), elle s'étend également à l'acquisition de grandeurs. C'est l'apport de l'IoT. Cette technologie réunit des objets hétérogènes par leur capacité à renvoyer des données sur l'Internet. En fonction des applications, les data générées sont traitées en temps réel, analysées, consolidées ou encore affichées. Dans le monde du travail, les exemples d'équipements IoT sont nombreux : enregistreurs de mesures physiques, caméras de surveillance de situation de travail⁵...

Le cas particulier des « équipements à porter »

Le développement du numérique a fait émerger une nouvelle famille d'objets techniques : la *wearable technology*⁶. Il s'agit d'équipements tels qu'une montre connectée, des lunettes connectées, des textiles spéciaux, etc.

Plus que de simples interfaces entre une machine et la personne qui la porte, ces objets remplissent des tâches de surveillance, d'alertes ou d'acquisition de données. Ils peuvent être munis de capteurs de pression atmosphérique, de température, de pulsation cardiaque, d'accélération, etc. Ces objets font aujourd'hui partie de notre quotidien et intègrent peu à peu le monde du travail.

4. <https://www.alexandria.unisg.ch/252999/1/s12599-015-0383-3.pdf>

5. <https://yomain.fr/>

6. <https://www.usine-digitale.fr/wearable-tech/>

Certains équipements de protection individuels (EPI) sont munis de capteurs ou d'avertisseurs, et d'une connexion à Internet. Ce sont les EPI connectés⁷.

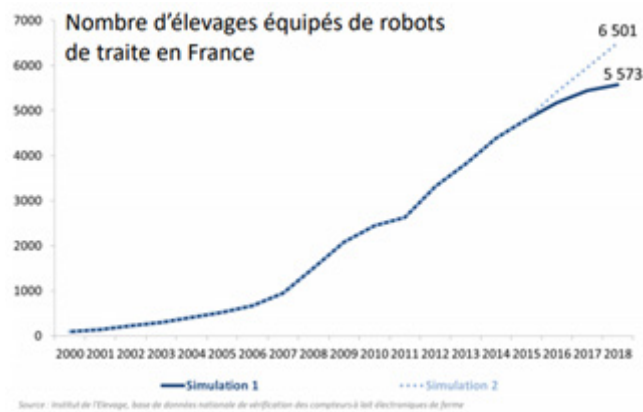
Des équipements à porter tels que les exosquelettes actifs (comme l'orthèse Roam Robotics⁸) rentrent également dans cette catégorie. Dans le monde du travail, ces orthèses constituent aujourd'hui le lien ultime entre le robot et l'humain.

Des travaux scientifiques ou artistiques sur l'implantation⁹ d'objets dans le corps humain sont en cours. Ces travaux appartiennent au courant de pensée transhumaniste. Les notions de cyborg et d'humain augmenté existent dans la littérature depuis les années 60.

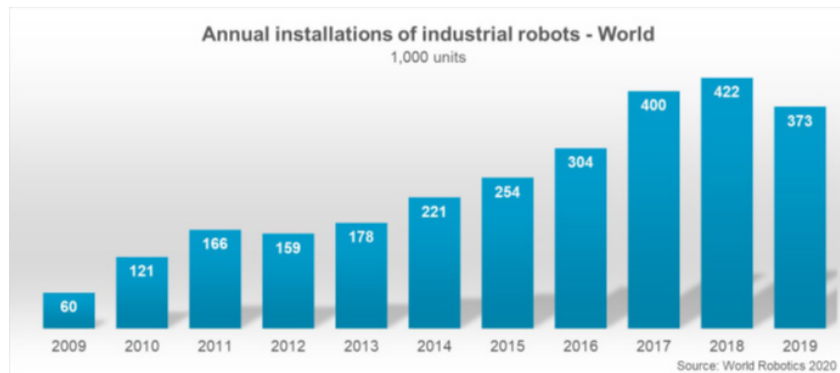
Indicateurs

Pour quantifier les évolutions de l'automatisation, nous allons nous intéresser aux marchés des équipements d'automatisation. Le volume de ventes ou d'installations de ces équipements sera un indicateur factuel de l'évolution de l'automatisation.

Secteur primaire :



Robots industriels :

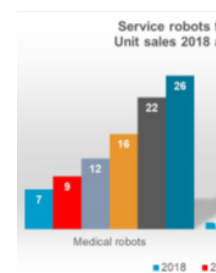
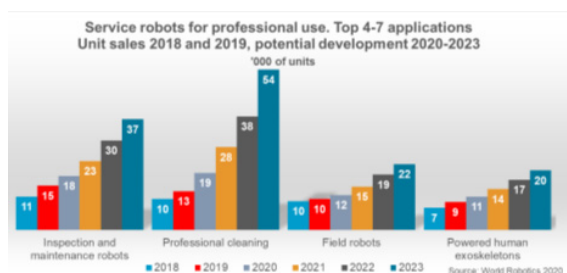
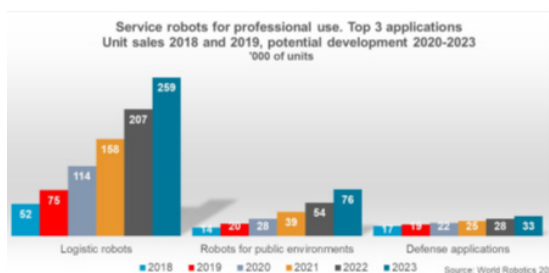


7. <https://dati-plus.com/dispositif-dati-pti/epi-connectes/>

8. <https://www.usine-digitale.fr/article/roam-robotics-presente-une-orthese-intelligente-pour-soulager-les-douleurs-de-genou.N1092079>

9. <https://mbamci.com/transhumanisme-humain-augmente-grace-aux-puces-electroniques/>

Robots de service :



Secteur tertiaire : les bots

Pas de données significatives.

Synthèse historique

Évolutions majeures

Dans l'industrie manufacturière

Comme montré précédemment, l'automatisation est à l'œuvre depuis les années 1970. Au passage du siècle, les tâches faciles à automatiser l'étaient déjà en grande partie. Et, selon le principe de Pareto, l'automatisation – qui avait alors atteint 80 % – nécessitait 80 % des efforts pour devenir complète. Ainsi, l'automatisation de l'outil industriel est apparu moins rentable. Les gains de productivité des usines françaises se sont centrés sur deux axes principaux : la qualité, les cadences et la disponibilité (quantifiées par le TRS¹⁰) ; et la délocalisation de la production en LCC¹¹. Cette dernière, qui présentait le plus grand gap de productivité, a été largement employée.

Dans les années 2010 apparaissent les premiers signes de déclin de la rentabilité du modèle « production délocalisée en LCC ». Plusieurs facteurs l'expliquent : l'augmentation du coût de main-d'œuvre qualifiée lié à l'augmentation du niveau de vie des pays producteurs ; l'augmentation du coût des transports (réelle ou perçue) ; le déficit d'image des produits *made in LCC*. À cela s'ajoutent en 2020 les considérations environnementales et la conscience du risque de rupture d'approvisionnement (crise Covid).

De manière concomitante, dans le courant des années 2010, des progrès techniques ont permis d'étendre les capacités d'automatisation à des coûts devenus pertinents. C'est l'arrivée des robots collaboratifs, de l'IoT, des technologies mobiles, du Cloud et de l'IA.

Ainsi, aujourd'hui, relocalisation et nouvelles technologies rendent l'automatisation de l'outil industriel profitable et, par conséquent, croissante.

10. Taux de rendement synthétique

11. Low Cost Country

Qu'en est-il dans l'agriculture ?

Deux avancées technologiques ont impacté l'automatisation du monde agricole durant ces vingt dernières années. En premier lieu, les robots de traite, dont la démocratisation débute vers 2000¹².

« La France comptait 40 exploitations laitières équipées d'un robot de traite en 2000 et plus de cent fois plus en 2018, d'après l'Idel. En effet, l'an dernier, 5 573 élevages laitiers étaient équipés de robots de traite, et cela ne devrait pas s'arrêter : 50 % des nouvelles installations de traite sont des robots. Au total, 17 % des vaches seraient traitées par un dispositif automatisé aujourd'hui. »¹³

La robotisation de l'agriculture passe également par la robotique mobile. La France dispose d'environ 2 000 robots agricoles (hors robots de traite). Ces robots pratiquent du désherbage mécanique ou la gestion de l'alimentation et des effluents d'élevage. Cela porte à 10 000 le nombre de robots pour l'ensemble du territoire¹⁴ et ce marché est en progression rapide [ndlr : nombre d'exploitations agricoles françaises = ~430 000].

Quelle évolution dans l'automatisation des services ?

Banque / Assurance : des systèmes d'agents conversationnels basés sur l'intelligence artificielle ont commencé à montrer une pertinence économique pour le service client à partir de 2017¹⁵. IBM Watson prétend (en 2020) être implanté dans 70 % des institutions bancaires mondiales¹⁶. Cette solution d'IA gère entre 50 et 60 % des questions posées sur l'agent conversationnel en ligne de GM Financial, et répond à 90 % des requêtes de manière autonome¹⁷. Cette automatisation du service client a montré son efficacité lors de la crise Covid (période pendant laquelle les agents se réorganisaient en travail à domicile). GM Financial prévoit d'étendre les fonctions de Watson à une plateforme téléphonique. En France, la branche assurance (maison et véhicule) du Crédit Mutuel utilise également IBM Watson¹⁸.

Dans la grande distribution, l'automatisation a pris la forme de caisses enregistreuses automatiques, permettant à un caissier de surveiller 4 à 6 bornes de paiement utilisées directement par le client. Ces automates sont apparus dans les grandes enseignes françaises en 2004. En 2017, 57 % des grandes surfaces en étaient pourvues¹⁹. La généralisation de ces caisses est freinée par l'accueil mitigé de la clientèle, les vols et le coût social. *A contrario*, la crise Covid a montré la pertinence de cette technologie pour le respect des gestes barrières sanitaires.

Le monde de la logistique est en plein bouleversement, l'automatisation en est à la fois la cause et la conséquence. La cause, car le commerce en ligne – sorte d'automatisation de l'achat – a créé une forte demande de logistique B2C. C'est la société américaine Amazon qui a initié le mouvement avec la création de sa filiale française en 2000. Le e-commerce pèse 43 milliards d'euros en 2020 après un bond de 24 % dû à la crise Covid²⁰. Par conséquent, le secteur a été contraint de s'automatiser pour être capable de s'adapter

12. <https://www.cairn.info/revue-reseaux-2020-2-page-253.htm>

13. <https://www.web-agri.fr/robot-de-traite/t403>

14. <https://www.lesechos.fr/2018/03/le-robot-watson-creuse-son-sillon-dans-la-banque-985864>

15. <https://www.ibm.com/fr-fr/watson>

16. <https://www.ibm.com/blogs/watson/2020/08/gm-financial-uses-ibm-watson-assistant-to-develop-a-secure-and-powerful-ai-assistant/>

17. <https://www.ibm.com/watson/stories/creditmutuel>

18. <https://www.lci.fr/social/caisses-en-libre-service-les-consommateurs-se-convertissent-2127404.html>

19. <https://www.lefigaro.fr/flash-eco/amazon-moins-performant-que-le-marche-du-e-commerce-en-france-en-2020-20210317>

20. [book-square-mecanisation-automatisation-robotisation-et-cobotisation-1.pdf](#)

à ces volumes d'échanges jusqu'alors inatteints. Comme pour l'industrie, la logistique s'automatise grâce aux robots collaboratifs ou mobiles, aux technologies mobiles, au Cloud et à l'IA. On parle alors de logistique 4.0²¹.



Les services du personnel ont également automatisé une partie de leurs tâches, et notamment grâce à l'IA. En 2014, la société Amazon a expérimenté une IA pour évaluer les candidats de une à cinq étoiles. Cette expérience a tourné court en 2015 quand ils ont observé un biais de genre²². Les CVs des candidates étaient systématiquement moins bien notés.

L'éducation n'est pas épargnée par l'automatisation. Sans écarter le professeur, l'automatisation proposée par des plateformes comme Moodle²³ permet de déployer les capacités d'enseignement à l'échelle planétaire. Moodle a été créé en 2002 ; en 2004, ils enregistraient leurs 1 000 premiers sites ; en 2008, 500 000 utilisateurs ; en 2015, 18 millions ; aujourd'hui, 190 millions d'utilisateurs au travers de 145 000 sites.

Hypothèses.

Évolutions majeures

Ceci étant posé, imaginons trois hypothèses :

- Nous commencerons par une première hypothèse très optimiste (du point de vue de l'automaticien), portée par la technologie et de grands enjeux planétaires. L'ensemble des métiers se retrouvent aidés par l'automatisation et aucun frein d'ordre éthique, sociétal ni de ressources ne vient entraver son déploiement.
- Ensuite, nous ferons l'hypothèse d'une automatisation modérée, tirée par

21. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/>

22. <https://moodle.com/the-moodle-story/>

23. <https://entrepriseiotinsights.com/20180427/channels/fundamentals/the-psychology-of-industry-50-tag99>

l'économie, mais ralentie par des préoccupations de société, de sécurité, de ressources ou encore technique.

- Finalement, nous envisagerons une décroissance de l'automatisation globale des métiers. Deux raisons principales seront explorées : les technologies ne tiennent pas leurs promesses, et il y a un rejet massif des métiers pour les automates dépourvus d'« humanité ».

Hypothèse I. L'automatisation va grand train

Primaire

L'interdiction mondiale du glyphosate provoque la généralisation du désherbage mécanique. La souveraineté alimentaire de la France impose de confier les grandes cultures à des tracteurs agricoles autonomes. De plus, le pays aux 360 fromages développe sa production laitière pour produire les protéines nécessaires à l'alimentation humaine et ainsi compenser la chute irrépressible de consommation de viande. L'usage massif d'objets de mesure connectés (associés à la *blockchain*) assure la sécurité alimentaire. La robotisation de la traite se généralise.

Les conséquences sur le secteur primaire se mesurent par l'émergence d'un écosystème dense et diffus de concepteurs, constructeurs et concessionnaires de machines agricoles robotiques. Un tissu de services se constitue autour des technologies numériques agroalimentaires. La densité d'agriculteurs reste stable, la concentration du secteur ayant déjà eu lieu.

Secondaire

L'industrie 4.0 a cédé la place à l'industrie 5.0²⁴. La simplification et les coûts de la robotisation des tâches de l'industrie manufacturière ont permis une automatisation proche des 90 %. Les 10 % des tâches de production restantes sont assurées en « home office » par des télémanipulateurs conduisant leur machine derrière un réseau à faible latence, et à distance de l'environnement sévère de la production industrielle. En dehors de la production, les personnes restant dans l'atelier sont aidées dans leurs efforts musculaires par des orthèses mécatroniques, leurs signaux biologiques sont analysés par leurs vêtements « intelligents ».

Ces quelques pourcentages d'automatisation supplémentaires (par rapport à 2022) renforcent encore la rentabilité de la relocalisation de la production sur le territoire. L'activité industrielle et l'emploi salarié du secteur augmentent mécaniquement.

Tertiaire

Les dossiers de solvabilité pour les emprunts bancaires, l'analyse des facteurs de risques pour les cotisations d'assurance, l'adéquation candidat/profil recherchée, la recherche dans la jurisprudence, etc. sont instruits par des intelligences artificielles. Les personnes

24. <https://enterpriseiotinsights.com/20180427/channels/fundamentals/the-psychology-of-industry-50-tag99>

responsables de ces fonctions s'appuient sur les résultats produits automatiquement. L'Homme garde la maîtrise de la décision et en assume la responsabilité. Les employés de ces secteurs se voient accompagnés dans les tâches purement administratives : leur métier et leur poste ne sont pas réellement mis en cause. Pas d'évolution majeure due à l'automatisation en termes de nombre d'emplois.

La logistique et la grande distribution se robotisent à grande échelle. La chaîne de tâches allant de la commande en ligne à la livraison au consommateur final est totalement assurée par des équipements autonomes. Cette chaîne est constituée d'un bot conversationnel de conseil à la vente, d'un paiement virtualisé, d'un ordre de livraison, d'une constitution de colis robotisée par transstockeur, AMR²⁵, robot palettiseur, et d'une livraison par un véhicule autonome. La grande distribution est quant à elle robotisée par des chariots-caisses-enregistreuses ou par des caisses automatiques. Malgré une augmentation des volumes transportés, la rationalisation du secteur provoque une baisse des emplois liés à la logistique et au commerce. L'écosystème des services autour de la logistique se renforce.

Pour l'éducation, la santé, l'aide à la personne, etc., la nécessité d'échanges interhumains limite l'automatisation à l'augmentation de performances des outils métiers.

Hypothèse 2. Des verrous technologiques et économiques

Les cyberattaques répétées des robots agricoles de plein champ²⁶ ont eu raison de la volonté d'extraire les paysans de la cabine de leur tracteur. Le développement de cette technologie ne présente pas d'intérêt économique supérieur aux risques encourus. En revanche, l'automatisation des tâches administratives et de suivi des marchés continue, ainsi que la robotisation des salles de traite.

Dans les ateliers, la collaboration des machines et des opérateurs n'apporte pas de bénéfice notable. Les contraintes techniques imposées pour assurer la sécurité des opérateurs humains réduisent drastiquement les gains de productivité. Les tâches complètement robotisables sans collaboration humaine sont robotisées, le coût de cette industrialisation est réduit. Les opérateurs réalisant des tâches physiquement éprouvantes les téléopèrent depuis un *dispatching* proche. Ils assurent ainsi une surveillance « au plus proche » sans prendre de risque physique et sans être soumis à l'environnement de production. Ainsi, l'usage d'orthèse mécatronique ou d'IoT portable d'analyse de constantes biologiques n'est pas nécessaire.

Les bots ne tiennent pas leurs promesses. Les agents conversationnels ne comprennent pas toutes les subtilités du langage humain. La diversité d'accents des francophones rend la compréhension des systèmes automatiques efficace à 80%. Les 20% restant répondent à la loi de Pareto, le gap technologique annoncé tarde. L'adoption de ces technologies par les collaborateurs humains et par les clients stagne, l'automatisation asymptote. La logistique – et ses chauffeurs routiers – plafonne au niveau 3²⁷ de l'autonomie des véhicules (niveau qui laisse un chauffeur dans la cabine du camion).

Les milliards d'euros investis dans ces technologies dans les années 20 pèsent sur les résultats, des opérateurs humains pallient les défauts des IA, le secteur s'oriente vers des voies où la technologie aide les opérateurs et ne vise pas à les remplacer.

25. Autonomous Mobil Robot

26. <https://technplay.com/robot-desherbage-pirate-hacktivistes/>

27. <https://www.matmut.fr/assurance/auto/conseils/vehicule-autonome-differents-niveaux-autonomie>

Hypothèse 3. Le rejet du robot : l'Homme, un animal social

L'accident de trop !

- L'enfant du paysan percuté par un robot agricole en plein champ.
- Le téléopérateur d'une usine déverse accidentellement des milliers de tonne d'« agent orange » dans la Garonne, tuant toute vie animale et végétale sur des centaines de kilomètres.
- Un e-agent de *trading* haute fréquence vend automatiquement des centaines de milliers de bonds souverains, provoquant une crise économique majeure.

Ou le ras-le-bol généralisé d'un monde déshumanisé

- « Mon père appelait ses vaches par leur prénom, je ne vois les miennes que derrière un tableur, je ne les connais qu'au travers d'hectolitres de lait annuel. »
- Dans son *dispatching*, loin de l'atelier, l'ouvrier ne participe plus à un collectif de travail destiné à produire des « biens », il perd le goût de ce métier dépourvu de « geste métier ».
- Les clients des banques, assurances, supermarchés, etc. boudent les enseignes proposant des agents robotiques au profit des petites agences, commerces de proximité, où une personne les accueille derrière un comptoir.

Dans cette hypothèse, c'est un refus généralisé de l'automatisation qui conduit au retour à des tâches automatiques entièrement sous contrôle humain.

L'usage d'algorithme d'IA se limite à l'aide à la décision (par la consolidation de données, par exemple). La cobotisation se limite à l'usage d'orthèses pour augmenter la force ou la précision. Et l'IoT à la mesure, mais pas à la surveillance.

Responsabilités et modes de gestion de la prévention

J. Munoz

Cette fiche vise à explorer les questions de responsabilité des acteurs (employeurs, travailleurs, donneurs d'ordres) en matière de santé et sécurité au travail, et par conséquent les modes de gestion des risques professionnels qui seront adoptés, notamment en prévention. L'IA pose de sérieuses questions en matière de responsabilités juridiques : les processus de décision autonome supportés par des systèmes d'apprentissage profond sont de nature à mettre à l'épreuve l'édifice institutionnel de gestion des risques professionnels et des responsabilités afférentes.

Nous rappellerons les principes qui prévalent dans le domaine de la prévention et la sécurité du travail (responsabilité de l'employeur) ; la place du droit européen dans ce domaine, notamment à partir de la directive 89/391/CEE du Conseil du 12 juin 1989 ; et enfin son évolution, à travers le principe d'obligation de résultat vers le principe d'obligation de moyens.

Rétrospective

Rappel du compromis du 9 avril 1898 et naissance du droit social et du travail

À l'aube de la société industrielle, la question des accidents du travail a occupé une place centrale dans les débats sur la responsabilité. Les échanges tournaient autour de la question de l'imputabilité de tels accidents, car ils se produisaient dans un espace fabriqué par l'homme, tout en posant la question du caractère « imprévisible » de tels phénomènes. La montée en puissance des mouvements ouvriers et l'instauration de la Troisième République vont favoriser l'élaboration d'un compromis. Si dès le 13 juin 1893 une loi est promulguée sur l'hygiène et la sécurité des travailleurs, celle-ci sera rapidement amoindrie par le compromis autour de la loi du 9 avril de 1898 sur les accidents du travail. Cette loi

créé la notion de risque professionnel. Sa définition est la suivante : « l'idée de risque ne désigne pas ce qui serait la cause d'un dommage, mais la règle selon laquelle on en répartira la charge. L'assurance propose une règle de justice qui n'a plus comme référence la nature, mais le groupe, une règle sociale de justice que le groupe est libre de fixer »¹.

À travers ce compromis est introduite une responsabilité sans faute de l'employeur. Cela suppose une transformation dans le mode de qualification de la situation, processus qui a été très bien analysé par Alain Supiot lorsqu'il souligne que la responsabilité « fondée sur l'idée du risque et non de faute est la conséquence directe de l'objectivation du corps humain dans le contrat de travail, celui-ci tire le corps humain du côté des choses. Tirant ainsi le corps humain du côté des choses et non des personnes, le contrat de travail tire du même côté les principes de la responsabilité qui cesse d'être une responsabilité du fait des personnes pour devenir une responsabilité du fait des choses. On dédramatise les rapports patron/ouvrier en objectivant la responsabilité, en supprimant toute idée de faute »².

Cependant, cette responsabilité sera limitée à la quotité prévue dans le cadre d'une assurance sociale contractée par l'employeur. Si le cadre législatif conservait la possibilité de réclamer la faute inexcusable de l'employeur ou du salarié, une telle éventualité était extrêmement rare, voire inexistante. Par ailleurs, l'édifice ainsi mis en place dominera amplement pendant plusieurs décennies, occultant les questions de prévention jusqu'à la création de la Sécurité sociale à la fin de la Seconde Guerre mondiale.

En effet, le contexte politique plus favorable à un rééquilibrage du rapport de force facilitera l'introduction de modifications importantes dans la gestion de risques professionnels. La gestion des accidents du travail et des maladies professionnelles était assurée par des compagnies d'assurance privées. Cette gestion sera transférée à la Sécurité sociale afin d'éviter le fort taux de refus et les coûts de gestion parfois exorbitants des compagnies privées. Quant à la faute inexcusable, à part une évolution dans sa définition à la suite d'un arrêt de la Cour de cassation toutes chambres réunies le 15 juillet 1941, il n'y aura pas d'évolution significative jusqu'aux années 1980.

Du droit d'hygiène et de sécurité vers un droit de la santé et de la sécurité

Des années 1970 à 1980, les questions liées aux risques professionnels connaîtront quelques évolutions, mais il va falloir attendre l'intervention du cadre européen pour observer des modifications majeures dans le secteur de la prévention en termes de responsabilité.

Si les années 1970 ont connu les jugements de « juges rouges » ayant incarcéré des employeurs pour « troubles à l'ordre public », de telles situations ne seront guère observées ultérieurement. La loi du 6 décembre 1976 introduit une série de mesures en matière d'accidents du travail, mais également dans le cadre de la responsabilité en termes de faute inexcusable de l'employeur, car désormais l'employeur peut non seulement s'assurer contre sa faute inexcusable, mais également pour celle de son subordonné.

Néanmoins, le véritable bouleversement en matière de gestion de risques professionnels viendra de la directive-cadre 89/391/CEE du Conseil du 12 juin 1989 qui vise à promouvoir l'amélioration de la sécurité et la santé des travailleurs au travail. Parmi les innovations

1. F. Ewald, « L'État providence », 1986.

2. A. Supiot, « Critique du droit du travail », 2011.

introduites par la directive-cadre, la notion d'évaluation de risques occupe une place centrale. Ainsi, à travers cette idée est développé le principe d'une obligation générale de prévention et à la source de risques. Cette obligation est à la charge de l'employeur, tenu comme responsable de la sécurité et de la santé de ses subordonnés. Le texte souligne l'importance de protéger la « santé de travailleurs » en ne faisant pas de différenciation entre la santé physique et mentale. La directive-cadre ne se contente pas de fixer un certain nombre de principes pour favoriser la prévention, mais défend l'idée de « l'adaptation du travail à l'homme ».

L'ensemble de ces principes sera transposé dans le droit français dans le cadre de la loi du 31 décembre 1991. Ces textes vont aboutir à une réécriture du Code du travail, et l'article L 4121-1 de celui-ci va traduire clairement l'esprit du nouveau cadre, selon lequel « l'employeur prend les mesures nécessaires pour assurer la sécurité et protéger la santé physique et mentale des travailleurs. Ces mesures comprennent des actions de prévention des risques professionnels, des actions d'information et de formation, et la mise en place d'une organisation et de moyens adaptés. L'employeur veille à l'adaptation de ces mesures pour tenir compte du changement des circonstances et tendre à l'amélioration des situations existantes ». L'idée générale d'une telle réécriture est clairement orientée vers une obligation de prévention. À la veille du nouveau millénaire, l'idée d'obligation était déjà contenue dans les principales modifications issues de la réglementation européenne. La responsabilité de l'employeur se trouve amplement modifiée dans ce contexte.

En réalité, ce qui se joue derrière ces évolutions est une transformation importante dans la façon d'aborder les questions de conditions de travail. En effet, on passe d'une conception de la prévention hygiéniste où le droit s'assurait que les conditions matérielles pour réaliser une activité soient correctes, à une conception de la prévention fondée sur le bien-être et la santé des subordonnés. Dans cette dernière perspective, il y a une rupture, dans la mesure où le champ englobé est bien plus vaste que les aspects purement matériels. Cela va se cristalliser avec la loi du 17 janvier 2002 et les arrêts de la Cour de cassation de février 2002.

Une accélération de l'évolution de la responsabilité en matière des risques professionnels

La loi du 17 janvier 2002 et les arrêts sur l'amiante de février 2002 introduisent des modifications importantes dans la conception de la responsabilité. En se référant à la directive-cadre, cette loi met en place le document unique d'évaluation de risques. Un tel dispositif vient accentuer l'obligation de prévention en matière de risques professionnels. En parallèle, et depuis la fin des années 1990, le contentieux autour de l'amiante va pousser les magistrats de la Cour de cassation à rendre un arrêt formalisant le principe d'obligation de résultat en matière des risques professionnels.

Les arrêts du 28 février 2002, dans le cadre de l'amiante en matière de responsabilité à la suite d'une procédure de faute inexcusable de l'employeur, vont introduire un vieux principe du domaine de la responsabilité. Les magistrats vont mobiliser la responsabilité contractuelle pour fonder l'arrêt introduisant l'obligation de résultat. Ainsi, ledit arrêt souligne qu'« attendu qu'en vertu du contrat de travail le liant à son salarié, l'employeur est tenu envers celui-ci d'une obligation de sécurité de résultat, notamment en ce qui concerne les maladies professionnelles contractées par ce salarié du fait des produits fabriqués ou utilisés par l'entreprise ; que le manquement à cette obligation a le caractère d'une faute inexcusable, au sens de l'article L.452-1 du Code de la sécurité sociale, lorsque l'employeur avait ou aurait dû avoir conscience du danger auquel était exposé le salarié, et qu'il n'a pas pris les mesures nécessaires pour l'en préserver ».

Cet arrêt aura une jurisprudence assez constante, notamment à partir de 2005, sans pour autant provoquer une avalanche de procédures en matière de faute inexcusable de l'employeur, les affaires étant presque exclusivement limitées à l'amiante. La responsabilité en matière de risques professionnels apparaît ainsi forte, puisque l'employeur peut estimer que le moindre accident du travail ou maladie professionnelle impliquera une condamnation devant les juridictions civiles. Toujours en matière de réparation, la faute inexcusable se cantonne à une série de préjudices (esthétiques, *pretium doloris* ou perte de promotion professionnelle). Dans un arrêt du 18 juin 2010, le Conseil d'État va renforcer la réparation de préjudices dans le domaine de risques professionnels en matière de faute inexcusable. Ainsi, l'obligation de résultat introduite par la Cour de cassation en 2002 aura des effets sur l'ensemble de l'édifice en matière de responsabilité.

Toutefois, ces dernières années, la même Cour de cassation a commencé à dessiner un tournant en matière jurisprudentielle avec une série d'arrêts, notamment à partir de 2015 et plus précisément avec l'arrêt du 22 octobre 2015. Dans cet arrêt, qui portait sur les effets d'une restructuration en cours dans une grande entreprise, les magistrats n'ont pas retenu l'obligation de résultat comme un élément pour justifier leur décision et ont rejeté le pourvoi de syndicats. En effet, l'entreprise, en entamant la restructuration, avait mis en place différents dispositifs pour accompagner les salariés dans cette structuration (cellule d'écoute, formations, congés, etc.). Autrement dit, elle a mobilisé des moyens afin de réduire les risques psychosociaux. C'est ainsi que le revirement s'opère pour exiger des employeurs de mettre en place « une obligation de moyens » afin de répondre à l'impératif de la protection des salariés. Certains juristes considèrent que les juges du fond de la Cour de cassation ont pu montrer une certaine sévérité envers les employeurs, ce qu'a pu « les amener à croire que, quoi qu'ils puissent faire en la matière, leur condamnation était inéluctable, ce qui, à terme, peut être de nature à contrarier la prévention elle-même »³. En effet, la Cour de cassation n'établit pas une absence de risques. La restructuration, dans sa nature même, peut produire des risques pour les salariés, et il est du devoir de l'employeur de protéger ses salariés. Ainsi, l'obligation de résultat ne disparaît pas totalement, mais elle est remplacée par une obligation de moyens renforcée.

L'employeur n'a plus une responsabilité automatique comme c'était dans le cas de l'obligation de résultat. On peut alors penser que l'employeur échappe à toute responsabilité. En réalité, l'obligation de moyens renforcée considère, sur la base de l'article L. 4121-1 et L. 4121-2 du Code du travail, que l'employeur doit faire la démonstration qu'il a pris toutes les mesures prévues par le Code du travail en matière de prévention. Or, l'évolution en matière de prévention allant vers la santé au travail, elle englobe des dimensions qui vont bien au-delà des contraintes classiques industrielles, telles que le port de charges ou les gestes répétitifs.

Ces évolutions jurisprudentielles peuvent aussi avoir des répercussions dans le domaine de la sous-traitance ou encore de l'analyse du travail de plateformes. Plus récemment, durant la pandémie, certains employeurs ont tenté de remettre en cause la couverture d'accidents du travail et de maladie professionnelle dans le cadre du télétravail, arguant le fait que les salariés ne se trouvaient plus dans l'espace de travail. C'est oublier, qu'en matière de responsabilité, au titre des accidents du travail, le lieu n'est pas le seul critère, mais que le temps et la subordination le sont également.

En dépit de l'apparition de rapports promouvant la restructuration de l'ensemble de l'organisation de la gestion de risques professionnels⁴, la gestion des risques professionnels – entendue ici au sens d'assurance sociale – n'a pas été radicalement bouleversée par

3. P.Y. Verkindt, « Clinique du travail et évolutions du droit », 2021.

4. C. Lecocq, « Rapport sur la santé au travail », 2018.

l'évolution de la responsabilité en matière d'accidents du travail et de maladies professionnelles.

Incertitudes majeures

- La question de la responsabilité se trouve confrontée à l'évolution des organisations du travail et de technologies. Comment, dans un contexte de plateformes des emplois, la question de la responsabilité en santé et travail va évoluer ? Comment la question de la co-activité sera traitée dans ces circonstances ?
- La récente loi de « renforcement » de la prévention (loi du 2 août 2021) introduit un principe de traçabilité des expositions : comment ce principe affectera la jurisprudence en matière de responsabilité ?
- L'introduction de nouveaux dispositifs techniques permettant de développer le télétravail va-t-elle modifier la responsabilité ?
- Quels sont les risques financiers afférents à une obligation de moyens renforcée ?

Hypothèses

Hypothèse 1. Obligation de résultat

L'IA est intégrée dans le cadre d'une obligation de résultat. Dans cette perspective, la responsabilité de l'employeur est dite « objective », à savoir engagée *ipso facto* à partir du moment où un accident du travail ou une maladie professionnelle est déclaré. La mise en place de l'IA aurait dû faire prendre conscience à l'employeur des effets d'un tel dispositif technique.

Hypothèse 2. Obligation de moyens renforcée de l'employeur

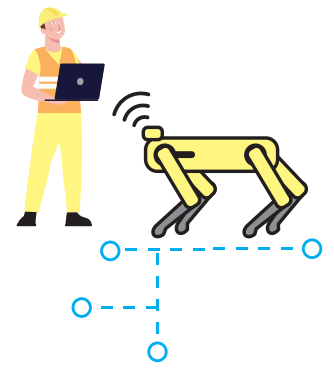
Dans le champ juridique, l'IA peut permettre aux acteurs – et notamment aux collectifs de victimes – de disposer d'un outil permettant d'objectiver ou de tracer des expositions. Une telle éventualité favoriserait les recours et les actions en termes de responsabilités, notamment dans le cadre de procédures telles que la faute inexcusable de l'employeur. L'IA, selon le type de critères choisis pour son fonctionnement, pourrait faciliter les recours, car elle mettrait à la disposition des victimes des informations difficiles agrégées ou compilées de manière individuelle.

Hypothèse 3. Obligation de moyens a minima

Dans un contexte d'organisation du travail de plus en plus fragmentée, les contentieux suite aux accidents et maladies se sont complexifiés et il est de plus en plus dur d'établir des responsabilités. Afin de désengorger les tribunaux et dans un souci de limiter les contraintes pour les entreprises de sécuriser les initiatives innovantes, la réglementation évolue vers un allègement de la responsabilité de l'employeur. Dans cette optique, il lui est essentiellement demandé de recourir à des technologies qui répondent à des normes, d'assurer les inspections périodiques de ses équipements et d'assurer un minimum d'informations et formations des travailleurs. Le simple respect de ces quelques exigences protège l'entreprise d'éventuelles poursuites en cas d'accident.

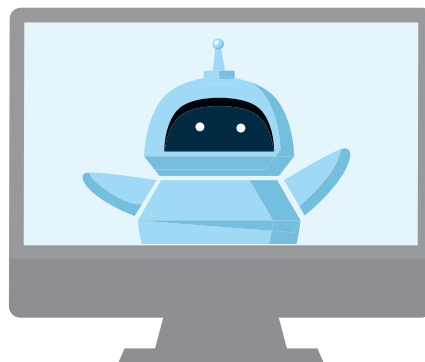
Bibliographie

1. François Ewald, « L'État providence », Paris, Ed. Grasset, 1986.
2. Loïc Lerouge, « L'obligation de sécurité : construction, réception et portée », In Nicolas Chaignot Delage et al., « Clinique du travail et évolutions du droit », p. 143-176, 2017.
3. Michel Miné, « L'obligation de sécurité de résultat de l'employeur », In Nicolas Chaignot Delage et al., « Clinique du travail et évolutions du droit », p. 523-539, 2017.
4. Matthieu Poumarède, « Intelligence artificielle, responsabilité civile et droit du travail », Droit social, p. 152, 2021.
5. Alain Supiot, « Critique du droit du travail », Paris, Ed. Puf, 1994.
6. Pierre-Yves Verkindt, « Pratique du juge et interprétation des normes : regards historique et contemporain sur la construction du droit de la santé au travail », In Nicolas Chaignot Delage et al., « Clinique du travail et évolutions du droit », p. 63-90, 2017.
7. Bulletin Cour de cassation, Décision n° 2010-8 QPC du 18 juin 2010.



Dans le cadre de son activité de prospective, l'INRS a animé une réflexion collective visant à explorer comment des systèmes mobilisant de l'intelligence artificielle pourraient être utilisés afin d'améliorer la santé et la sécurité au travail à l'horizon 2035.

Quelles sont les opportunités et menaces pour la prévention des risques professionnels ? Dans quels domaines des avancées sont-elles possibles ? À quelles conditions ? Comment les acteurs peuvent-ils se préparer ?



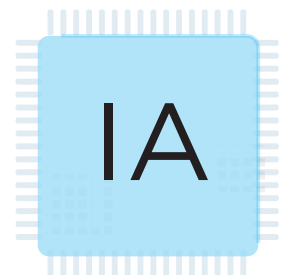
Institut national de recherche et de sécurité
pour la prévention des accidents du travail et des maladies professionnelles
65, boulevard Richard-Lenoir 75011 Paris • Tél. 01 40 44 30 00 • info@inrs.fr

Édition INRS PV 19

1^{re} édition • novembre 2022 • 978-2-7389-2812-2

Conception éditoriale : Mathilde Cocquelin

▶ L'INRS est financé par la Sécurité sociale – Assurance maladie/Risques professionnels ◀



www.inrs.fr

