

# **NOTE DE SYNTHÈSE**

## **CONTEXTES MONDIAL ET FRANÇAIS POUR LE DÉVELOPPEMENT ET LE DÉPLOIEMENT DE L'INTELLIGENCE ARTIFICIELLE**

# Sommaire

1.	<b>Introduction</b>	4
2.	<b>Définitions et contexte</b>	7
2.1	L'intelligence, une notion qui reste difficile à définir	8
2.2	Le travail, un sujet plus complexe que celui de l'emploi	10
2.3	Quand l'intelligence artificielle rencontre le travail	13
2.4	Déploiement massif de l'intelligence artificielle : croyances vs. réalités	17
3.	<b>Le développement et le déploiement de l'intelligence artificielle, un enjeu mondial</b>	22
3.1	Des initiatives internationales pour envisager le développement de l'IA	24
3.1.1	Travailler pour bâtir un avenir meilleur	24
3.1.2	Garantir une IA digne de confiance	26
3.1.3	PMIA et avenir du travail	28
3.2	Des initiatives pour le développement de l'IA à l'échelle de la France	32
3.2.1	Synthèse du rapport Villani	33
3.2.2	D'un EDEC pilote à un EDEC national	35
3.2.3	Repenser le dialogue social	37
4.	<b>Pour une IA inclusive et de confiance</b>	39
4.1	De l'inclusion	41
4.2	Opérationnaliser la notion d'inclusion	43
5.	<b>Conclusion</b>	47



001

# Introduction

Le ministère du Travail, de l'Emploi et de l'Insertion et Inria ont créé en novembre 2021, le LaborIA, un laboratoire (ou living lab) visant à construire une vision terrain des usages de l'IA et de ses effets sur le travail, l'emploi, les compétences et le dialogue social. Dans le cadre du LaborIA, Matrice mène un premier projet d'étude et d'expérimentation sur l'intelligence artificielle au travail en France, incluant la réalisation d'un baromètre de l'IA au travail, des séminaires de sensibilisation aux effets de cette technologie sur le travail et des expérimentations in situ dans les entreprises.

Considérant les nombreuses initiatives ayant déjà été, ou étant, menées sur le sujet "intelligence artificielle et travail" – parmi lesquelles celle du Partenariat mondial sur l'intelligence artificielle (PMIA) qui a créé un groupe de travail sur le sujet ; ou encore celles des EDEC (Engagement de développement de l'emploi et des compétences) avec, notamment, l'EDEC Perspectives IA – la toute première étape engagée par Matrice est la réalisation d'un état de l'art qui servira de support aux réflexions du LaborIA. Ce document ne prétend pas à l'exhaustivité, mais il est articulé autour de plusieurs problématiques d'intérêt en lien avec les objectifs globaux du LaborIA :

- Quel est l'impact de l'automatisation sur l'emploi en termes économiques, sociaux, humains, ... ?
- Comment parvient-on à introduire de l'IA dans les entreprises, pour leur propre bénéfice et celui de leurs parties prenantes, en premier lieu les travailleurs ?
- Quel dialogue social faut-il mener, quelles solutions et méthodes «by-design» faut-il créer pour limiter les impacts négatifs de l'automatisation sur l'emploi et favoriser un redéploiement des forces vives sur d'autres tâches / métiers humains ?

Ce rapport de synthèse dresse donc un panorama de l'état actuel du développement de l'intelligence artificielle au sein du monde du travail, en s'attachant à détailler, le plus précisément possible, l'ensemble des enjeux liés à cette technologie. Il se focalise notamment sur le contexte socio-culturel de l'intégration de l'intelligence artificielle au travail, retrace la vision et les initiatives internationales et françaises récentes (groupes de travail, cas d'usage, guides et propositions d'accompagnement, ...) puis, finalement, s'intéresse aux différents moyens de promouvoir une IA inclusive et de confiance en Europe, et plus spécifiquement en France.

Afin d'engager ce programme, Matrice a réalisé une analyse des différents travaux et initiatives déjà menés sur ce vaste sujet : "l'impact de l'IA sur le travail". Cet état de l'art propose une vision globale et synthétique, qui servira de base commune pour les réalisations du programme avec plusieurs objectifs :

- Chaque personne impliquée dans le programme doit pouvoir avoir accès aux informations clés sur le sujet et conduire sa démarche au moyen d'une base de réflexion commune.
- "Intelligence artificielle" et "travail" étant des concepts larges et complexes, il paraît important de s'accorder sur des définitions et un cadre commun. Ces éléments constituent le point de départ pour la mise en place du LaborIA et la conception d'un programme de recherche-action pertinent sur les enjeux actuels liés à l'IA et au travail. Cette démarche ne ferme pas la porte aux différents débats qui surviendront probablement sur ces concepts pendant le déroulement du programme.

- Matrice et les différents acteurs impliqués doivent pouvoir compléter la réflexion globale, déjà avancée, sur l'impact de l'IA au travail. Il est donc nécessaire d'identifier ce qui a déjà été fait afin de proposer de nouveaux axes d'études ou de les compléter.

Ainsi, ce document se veut accessible pour toute personne travaillant sur le programme LaborIA dans l'objectif de comprendre le sujet de l'IA au travail, son historique, son contexte et les interrogations posées.

Concrètement, cet état de l'art sera partagé et utile pour différentes parties prenantes :

- Aux partenaires du programme (ministère du Travail, de l'Emploi et de l'Insertion, Inria, Matrice) et plus spécifiquement à leurs collaborateurs. Cet état de l'art leur permettra de se constituer une base commune de connaissances en amont de l'opérationnalisation du programme.
- Dans un format adapté, aux entreprises participant aux séminaires et/ou expérimentations du programme. La lecture de l'état de l'art permettra à leurs collaborateurs de se sensibiliser à l'IA, avant de réfléchir à des projets concrets de mise en place de l'IA dans l'entreprise.
- Aux coordinateurs scientifiques du programme (membres du Comité scientifique et autres experts impliqués). L'état de l'art leur permettra de comprendre le cadre défini en début de programme et ainsi de proposer des orientations prenant en compte ce cadre ou en contraste avec ce point de départ.

002



# Définitions et contexte



## 2.1

## L'intelligence, une notion qui reste difficile à définir

Le concept d'intelligence reste un concept vague qui fait encore l'objet d'études approfondies.

Chez l'humain, ce que l'on appelle intelligence est pluriel : pour être vraiment rigoureux, il faudrait considérer une dizaine d'intelligences<sup>1</sup>. On peut néanmoins considérer trois grandes conceptions de l'intelligence<sup>2</sup> :

- L'intelligence héréditariste, qui est une donnée d'ordre biologique ;
- L'intelligence constructiviste, qui est le résultat d'une construction concrète et progressive ;
- L'intelligence perfectionniste, qui renvoie à l'exigence de se construire intelligent. Cette conception spécifique est fondée sur l'idée rousseauiste de perfectibilité.

Les deux dernières conceptions soutiennent une approche diversifiée et dynamique de l'intelligence humaine. En ce sens, cette dernière peut être vue comme une conquête individuelle et quotidienne pour parvenir à aiguiser sa propre capacité de distanciation et de conscience critique. Mais plus qu'une capacité réflexive, certains chercheurs<sup>3</sup> définissent l'intelligence humaine comme une aptitude à créer du nouveau en démontrant une faculté de sensibilité et d'adaptation

1. Gardner, H. & Hatch, T. (1989). Multiple Intelligences Go to School: Educational Implications of the Theory of Multiple Intelligences. *Educational Researcher*, 18 (8), 4-10. Disponible sur : <https://www.sfu.ca>.

2. Dubuc, B. (1994). Compte rendu de [Hadji, C. (1992). *Penser et agir l'éducation*. Paris : ESF.] *Revue des sciences de l'éducation*, 20 (4), 792-793. Disponible sur : <https://www.erudit.org>.

3. Romero, M., Lille, B. & Patiño, A. *Usages créatifs du numérique pour l'apprentissage au XXIe siècle*, Québec, Presses de l'Université du Québec, 2017.

au contexte socioculturel et d'empathie sur le plan intra et inter-psychologique aux différents acteurs. Ceci implique de comprendre la nature humaine et socio-historique pour être en mesure de porter un jugement sur ses propres processus et création, de manière autonome.

Pour la machine, parmi l'ensemble des définitions disponibles de l'intelligence artificielle (IA), on retiendra celle du Groupe d'experts de haut niveau sur l'intelligence artificielle - mandaté par la Commission européenne - formalisée dans le cadre de son travail de synthèse pour une compréhension commune de ce que sont les systèmes d'IA<sup>4</sup>.

Ainsi, les systèmes d'intelligence artificielle sont "des systèmes logiciels (et éventuellement matériels) conçus par des êtres humains et qui, ayant reçu un objectif complexe, agissent dans le monde réel ou numérique en percevant leur environnement par l'acquisition de données, en interprétant les données structurées ou non structurées collectées, en appliquant un raisonnement aux connaissances, ou en traitant les informations, dérivées de ces données et en décidant de la ou des meilleures actions à prendre pour atteindre l'objectif donné. Les systèmes d'IA peuvent soit utiliser des règles symboliques, soit apprendre un modèle numérique. Ils peuvent également adapter leur comportement en analysant la manière dont l'environnement est affecté par leurs actions antérieures."

En tant que discipline scientifique, l'IA comprend plusieurs approches et techniques, telles que :

- L'apprentissage automatique (dont l'apprentissage profond et l'apprentissage par renforcement sont des exemples spécifiques) ;
- Le raisonnement automatique (qui comprend la planification, la programmation, la représentation des connaissances et le raisonnement, la recherche et l'optimisation) ;
- La robotique (qui comprend le contrôle, la perception, des capteurs et des actionneurs, ainsi que l'intégration de toutes les autres techniques dans des systèmes cyberphysiques).

On peut néanmoins, dans la plupart des questionnements concernant le travail, prendre une définition plus large et plus inclusive de l'intelligence artificielle, et considérer que c'est le champ du numérique concerné par la simultanéité de données et d'algorithmes.

En confrontant ces éléments de définition, on observe une différence majeure entre intelligence humaine et intelligence artificielle. Chacune est multiple, cependant, l'une fait intervenir des composantes à la fois logiques, sensibles et réflexives, tandis que l'autre consiste en une forme acérée d'intelligence logique qui, sans intervention humaine, n'est pas en capacité de s'adapter à un contexte socioculturel ou à des formes de sensibilité spécifiques.

4. Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle constitué par la Commission européenne (2019). Définition de l'IA : Principales capacités et disciplines. Disponible sur : <https://digital-strategy.ec.europa.eu>.

# 2.2

## Le travail, un sujet plus complexe que celui de l'emploi



Selon le sociologue Georges Friedmann<sup>5</sup>, le travail est un trait spécifique de l'espèce humaine, en ce sens qu'il est un dénominateur commun et une condition de toute vie en communauté. En référence aux travaux de Karl Marx<sup>6</sup>, il donne une définition partielle du travail - rapportée à l'homo faber - comme étant l'ensemble des actions que l'homme, dans un but pratique, à l'aide de son cerveau, de ses mains, d'outils ou de machines, exerce sur la matière ; actions qui, à leur tour, réagissant sur l'homme, le modifient. Cependant, cette définition n'est que partielle car elle ne comprend pas, a priori, les activités de services. En outre, elle semble avoir été pensée uniquement via un prisme très technique et utilitariste alors que cette activité de travail peut être considérée sous des points de vue multiples : physiologique, psychologique, sociologique, économique, ... C'est pourquoi Friedmann insiste sur le fait qu'une définition du travail n'a de sens que si elle :

- Est corrélée à des données historiques, sociologiques et ethnographiques ;
- Fait référence à la variété de ses formes concrètes selon les sociétés, les cultures et les civilisations ;
- Inclut la manière dont le travail est vécu et ressenti par ceux qui l'effectuent.

Le terme français "travail"<sup>7,8</sup>, désigne trois axes, sur lesquels la notion s'est façonnée historiquement, et demeure dans une

5. Friedmann, G. (1960). Qu'est-ce que le travail ? Annales. Économies, sociétés, civilisations, 15 (4), 684-701. Disponible sur : <https://www.persee.fr>.

6. Marx, K. Le capital, Paris, Flammarion, 1883.

7. Casilli, A. A. En attendant les robots, Paris, Éditions du Seuil, 2019.

8. Méda, D. Le travail : une valeur en voie de disparition ?, Paris, Aubier, 1995.

triple relation : de l'individu au donné naturel (conception individualiste de l'être humain vis-à-vis du monde) ; de l'individu aux autres (conception de l'être humain vis-à-vis d'un collectif, d'une société) ; de l'individu à lui-même (conception individualiste de l'être humain vis-à-vis de lui-même).

Plus communément, le travail est une composante de l'identité individuelle, un lien aux autres, une source de développement personnel, un facteur potentiel d'épanouissement ; à condition toutefois qu'il soit porteur de sens, qu'il ne soit pas "pénible", et qu'il suscite une forme de reconnaissance<sup>9</sup>. Ainsi, d'un point de vue culturel, il est une valeur centrale de notre vivre-ensemble et, d'un point de vue substantiel, il est un élément prépondérant dans les trajectoires de vie des individus ainsi que dans la fabrique des sociétés<sup>7</sup>. Dans son ouvrage intitulé "Le travail : une valeur en voie de disparition ?", Dominique Méda<sup>8</sup> va même plus loin en décrivant le travail comme ce qui exprime au plus haut point notre humanité, notre condition d'être finis, créateurs de valeurs, mais aussi d'être sociaux.

Au quotidien, cette notion de travail se matérialise en partie - car il en existe des formes multiples - par l'emploi. Emploi pour lequel chaque individu se doit de produire un certain nombre d'éléments répondant à des objectifs - ou de prendre en charge un certain nombre de services - et en échange desquels il touche une gratification.

La question de l'emploi est, depuis plus d'une vingtaine d'années, une préoccupation majeure. Le progrès technologique, l'évolution du marché du travail, la mondialisation, (etc.) bouleversent l'ensemble du modèle salarial inventé au lendemain de la seconde guerre mondiale<sup>10</sup>. Dès le milieu du XXème siècle, l'introduction de l'ordinateur en entreprise modifie significativement la nature du travail tel qu'on le connaît alors. En outre, l'impact de l'ordinateur s'avère relativement hétérogène selon le secteur d'activité concerné : parfois, il permet aux employeurs d'engager moins de travailleurs, comme dans les industries automobiles ; d'autres fois, il nécessite d'embaucher des personnes dédiées, par exemple pour la saisie de données, qui demeure une tâche humaine comparativement au traitement de l'information rapidement effectué par les ordinateurs<sup>11</sup>. Dès cette époque, certaines observations sur les conséquences des premiers pas vers l'automatisation ont été formulées :

- Le travail est fragmenté en tâches qui se répartissent entre machines et humains ;
- L'utilisation de l'ordinateur en entreprise induit une accélération technique qui contraint les travailleurs à s'adapter à la machine et à travailler comme elle ;
- Les entreprises, dans une logique économique, mettent en compétition humains et machines dans le but de produire toujours plus et toujours moins cher.

9. Weil, T. (2018). Le travail en questions : quelles organisations, frontières, reconnaissances ?, The Conversation. Disponible sur : <https://theconversation.com>.

10. Piotet, F. (2005). La sociologie du travail depuis Georges Friedmann, SES ENS Lyon. Disponible sur : <http://ses.ens-lyon.fr>.

11. Guillaud, H. (2021). L'automatisation, une idéologie ?, Internetactu.net. Disponible sur : <https://www.internetactu.net>.



## Quand l'intelligence artificielle rencontre le travail

# 2.3

Alors qu'on assiste à la démocratisation de la micro-informatique à la fin du XXème siècle, certains auteurs<sup>12,13</sup> s'inquiètent de "la fin du travail" en expliquant que grâce à l'accroissement de la productivité et aux progrès technologiques, de moins en moins de main-d'œuvre est nécessaire pour produire autant. En suivant cette logique, il n'est plus possible de garantir un emploi pour l'ensemble des individus. Leur analyse est complétée par l'hypothèse de la nécessité d'un changement profond des modes de production du fait de l'évolution des contraintes économiques et de la transformation de la nature des besoins : les organisations ont désormais besoin d'être en capacité de s'adapter rapidement et efficacement aux changements, ce qui n'est pas réellement permis par le régime du salariat. Cependant, d'autres auteurs, tel David H. Autor<sup>14</sup>, semblent avoir une approche différente du sujet et insistent sur les deux pans du progrès technologique. En effet, la machine se substitue à l'humain sur un certain nombre de tâches, mais en faisant cela, elle est également un élément complémentaire de la main-d'œuvre humaine qui permet d'augmenter la productivité, les revenus et la demande de travail. De leur côté, Askenazy et Bach<sup>15</sup> vont même jusqu'à retourner la problématique du développement de l'IA en insistant sur le fait que cette technologie à usage général (GPT - General purpose technology) "n'emporte pas par nature une disparition du travail", même si elle "implique une progressive transformation d'activités (...) et une évolution de la formation des (futurs) actifs". En revanche, compte tenu des limitations technologiques actuelles, ils posent clairement la question de la productivité gagnée grâce à l'IA et de la possibilité réelle pour l'humain de se tourner vers des tâches à plus forte valeur humaine ajoutée pour répondre aux défis du XXIème siècle.

La question des effets technologiques sur les activités humaines traverse notre civilisation depuis ses origines. La peur de la destruction du travail par l'automatisation, et encore plus par l'intelligence artificielle, n'est que son incarnation la plus récente<sup>7</sup>. Malgré ce caractère récurrent, voire persistant, chaque progrès technologique reste immuablement accompagné de son lot d'inquiétudes et d'angoisses quant à l'impact qu'il pourrait avoir sur l'humain. Depuis le boom de l'IA dans les années 2010, les travaux de recherche se sont multipliés pour prévoir et comprendre quels impacts pourraient avoir l'intelligence artificielle, notamment sur les travailleurs. Les différents scénarios prospectifs laissent entrevoir six grandes tendances<sup>16</sup> :

1. Approche axée sur l'emploi - Le travailleur serait massivement remplacé par les systèmes d'IA et de lourdes pertes d'emplois seraient à déplorer. Certains chercheurs, tels que Frey et Osborne<sup>17</sup>, ont quantifié le degré d'automatisation des emplois et donc, leur destruction par l'IA. Selon eux, 47 % des emplois aux USA sont susceptibles de disparaître face à la vague d'innovation technologique basée sur l'apprentissage automatique et la robotique mobile. Cependant, d'autres comme Casilli<sup>7</sup> se font les critiques d'une telle étude et démontrent que la méthodologie utilisée pour arriver à cette quantification comporte des biais et des limitations de nature conceptuelle et de nature statistique, parmi lesquels :



12. Rifkin, J. La fin du travail, Paris, La Découverte, 1996.

13. Bridges, W. La conquête du travail, Paris, Village mondial, 1995.

14. Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. Journal of Economic Perspectives, 29 (3), 3-30. Disponible sur : <https://www.aeaweb.org>.

15. Askenazy, P. & Bach F. (2019). IA et emploi : une menace artificielle. Pouvoirs - Revue française d'études constitutionnelles et politiques, Le Seuil, 170, 1-7. Disponible sur : <https://hal.archives-ouvertes.fr>.

16. Groupe de travail sur l'avenir du travail, GPAI Montréal Summit, novembre 2020. Disponible sur : <https://gpai.ai>.

17. Frey, C. B., Osborne, M. (2017). The Future of Employment: How Susceptible are Jobs to Computerization? Technological Forecasting and Social Change, 114 (C), 254-280. Disponible sur : <https://www.sciencedirect.com>.

- Seul un sous-échantillon de 10 % des fonctions professionnelles est examiné par les auteurs ;
- Les transformations profondes de l'emploi, c'est-à-dire sa fragmentation en micro-tâches, ne sont pas prises en compte ;
- Les auteurs conçoivent l'innovation comme un processus transcendant les rapports sociaux de production et font l'impasse sur la dimension de négociation sociale qui encadre l'expansion du machinisme.

Dans ces conditions, il est difficile d'affirmer avec certitude que les robots vont rendre obsolètes la plupart des travailleurs.

2. Approche axée sur les tâches - Le travailleur serait assisté par les systèmes d'IA qui l'aideraient à réaliser ses tâches.

Cette vision du travailleur assisté met en avant l'aspect complémentaire du travail humain et du travail des machines<sup>14</sup>. Dans ce cadre, l'IA prendrait en charge les tâches routinières, répétitives, ennuyeuses, voire risquées pour l'humain. Ce dernier pourrait alors se concentrer sur des tâches plus complexes, novatrices et créatives, impliquant une part importante de travail en équipe et d'intelligence sociale. Dans cette perspective, l'OCDE (Organisation de coopération et de développement économique)<sup>18</sup> estime qu'en France, 16,4 % des emplois sont susceptibles de disparaître et 32,8 % des emplois pourraient être profondément transformés<sup>19</sup>. Néanmoins, l'organisation met également en évidence le fait que cette quantification de l'impact de l'automatisation sur l'emploi dépend fortement des territoires en termes de secteurs d'activité des industries prédominantes, de nature des métiers particulièrement présents, etc.

3. Approche axée sur l'autonomie - Le travailleur serait dominé par les systèmes d'IA, qui réduiraient drastiquement son pouvoir d'action.

Ce scénario renvoie à la crainte du développement d'une algocratie, c'est-à-dire d'un type de gouvernance au sein duquel les algorithmes établissent les lois et les règles qui structurent la façon dont les humains agissent. L'autonomie des travailleurs se retrouverait compromise à cause notamment d'une rationalisation accrue du travail<sup>20</sup>, du développement de la passivité du travailleur par une confiance excessive en l'IA ou par complaisance<sup>21,22,23</sup>, de la soumission à un algorithme par obéissance, peur de l'erreur et punition, etc. Les moyens proposés par ces auteurs pour surmonter ce risque d'algocratie consistent essentiellement à tenir compte des humains et de leurs besoins pour le développement de

systèmes d'IA, accompagner les travailleurs vers ces transformations du monde du travail via la formation professionnelle, engager le dialogue social avec les travailleurs avant même d'implanter un système d'IA au sein de l'entreprise (besoins, co-conception, transparence, explicabilité, ...) puis, concevoir les systèmes d'IA de manière éthique.

4. Approche axée sur la performance - Le travailleur disposerait d'un pouvoir d'action plus important grâce aux systèmes d'IA.

Cette approche se fonde sur l'hypothèse d'un partenariat entre l'humain et la machine de telle sorte que la valeur ajoutée des premiers serait accrue de façon significative. Forts d'un tel niveau de performance, les travailleurs pourraient disposer de davantage de temps pour se former, développer un esprit critique, laisser place à leur créativité, aux loisirs...<sup>24</sup>

5. Approche socio-économique - Les systèmes d'IA entraîneraient une polarisation du marché du travail et creuseraient davantage l'écart existant entre les différents profils de travailleurs.

Du fait du déploiement massif de l'IA, de nombreuses études suggèrent l'apparition de deux typologies d'emplois :

- Les emplois hautement qualifiés ayant un haut niveau de complémentarité avec l'IA qui seraient particulièrement stimulants et permettraient de laisser s'exprimer les capacités créatives de l'humain ;
- Les emplois peu qualifiés et précaires pour lesquels les travailleurs seraient cantonnés à des tâches peu intéressantes.

Les tâches relatives aux emplois intermédiaires, quant à elles, seraient probablement réalisées exclusivement par des machines.

6. Approche humaniste - Le déploiement massif de l'intelligence artificielle permettrait au travailleur de se concentrer sur des compétences proprement humaines.

L'automatisation des tâches et des emplois permettrait d'aboutir à une répartition optimale des actions humaines et des actions automatisées de sorte que l'humain serait en capacité de développer l'ensemble des capacités qui le différencient de la machine : créativité, dextérité manuelle, pensée abstraite, résolution de problèmes, adaptabilité, intelligence émotionnelle...<sup>24</sup>

À l'heure actuelle et malgré toutes les recherches sur le sujet, il reste difficile de déterminer avec précision et certitude ce que le déploiement massif de l'intelligence artificielle pourrait engendrer sur le monde du travail. Se situe-t-on toujours dans un cycle économique schumpétérien de destruction créatrice des activités économiques ?<sup>25</sup> Pour cette transition spécifique de l'ère

18. OCDE (2019), OECD Employment Outlook 2019 : The Future of Work, Éditions OCDE, Paris. Disponible sur : <https://doi.org/10.1787/9ee00155-en>.

19. Lepeltier, N. (2019). La robotisation va-t-elle faire disparaître des emplois ou bien en créer ?, Le Devoir. Disponible sur : <https://www.ledevoir.com>.

20. Head, S. Mindless: Why Smarter Machines are Making Dumber Humans? New York, Basic Books, 2014.

21. Ganascia, J.-G. Le mythe de la Singularité. Faut-il craindre l'Intelligence artificielle ?, Paris, Seuil, 2017.

22. Devillers, L. Des robots et des Hommes, Paris, Plon, 2017.

23. Carr, N. The Glass Cage: Automation and Us, New-York, W.W. Norton & Company, Inc, 2014.

24. Villani, C. (2018). Donner un sens à l'Intelligence Artificielle. Pour une stratégie nationale et européenne. Disponible sur : <https://www.aiforhumanity.fr>.

25. "Joseph Schumpeter (1883-1950), économiste et professeur en science politique autrichien naturalisé américain, explique que l'économie est gouvernée par un phénomène particulier : la "destruction créatrice". "C'est la donnée fondamentale du capitalisme et toute entreprise doit, bon gré mal gré, s'y adapter." La croissance est un processus permanent de création, de destruction et de structuration des activités économiques. En effet, « le nouveau ne sort pas de l'ancien, mais à côté de l'ancien, lui fait concurrence jusqu'à le nuire ». Ce processus de destruction créatrice est à l'origine des fluctuations économiques sous forme de cycles." - Source : <https://www.economie.gouv.fr>.

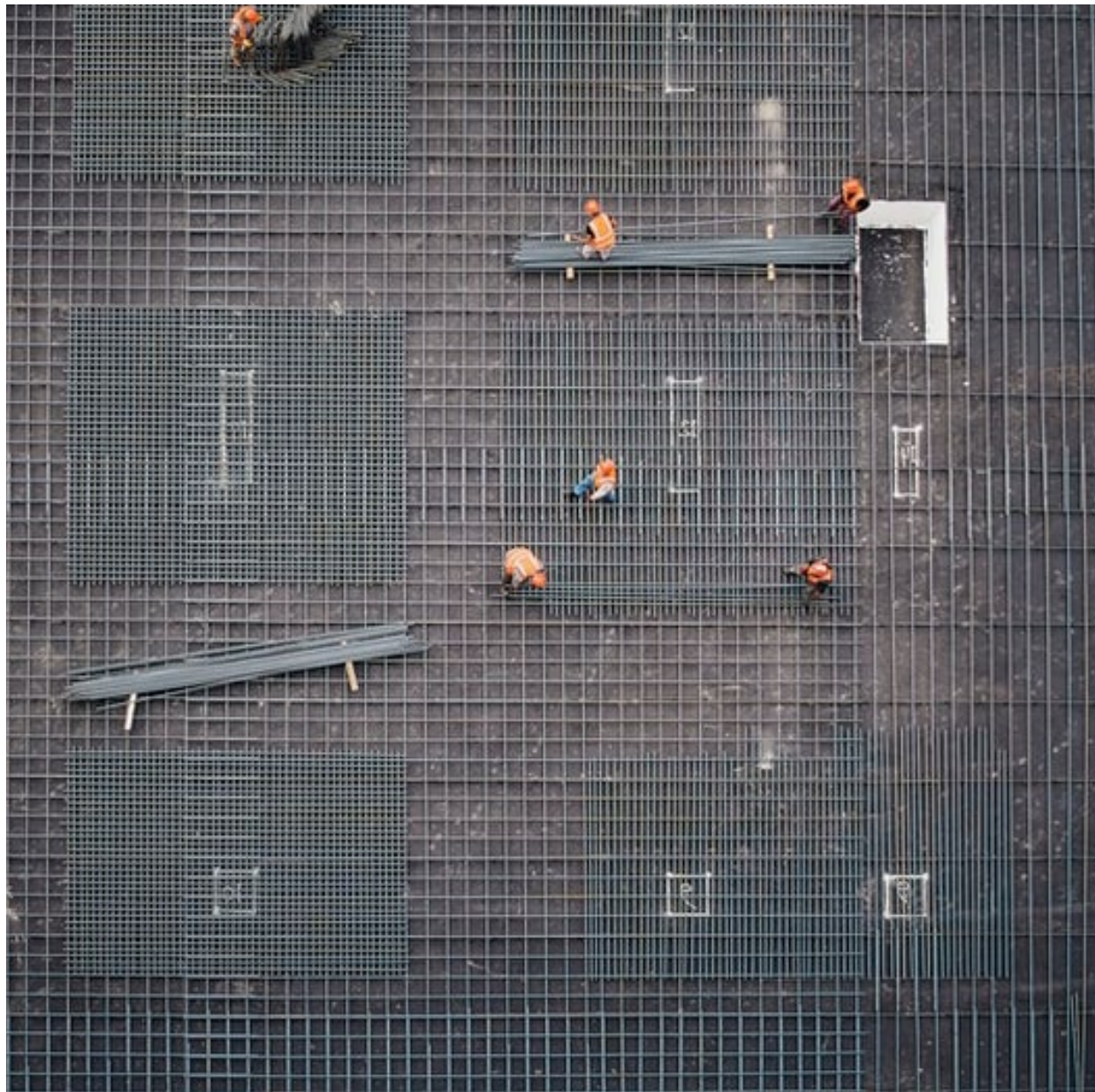


industrielle vers l'ère numérique, sera-t-il possible de créer de nouveaux emplois qui permettront de requalifier les travailleurs dont l'emploi a été détruit ? Ou est-ce que cette ambition est inatteignable et que tout doit être repensé en termes de politique et structure d'emplois ? Comment les travailleurs envisagent-ils l'automatisation, qu'est-ce qui est acceptable ? Se sentent-ils prêts pour une transition forte ? Quelles sont leurs aspirations ? Comment les intégrer dans la construction de la société numérique ? Comment les accompagner dans l'affirmation de leur identité, de leurs compétences et de leurs spécificités ?

Vaste flot de questions auxquelles seule la société dans son ensemble semble pouvoir répondre en mobilisant l'intelligence collective et le dialogue.

## Déploiement massif de l'intelligence artificielle : croyances vs. réalités

# 2.4

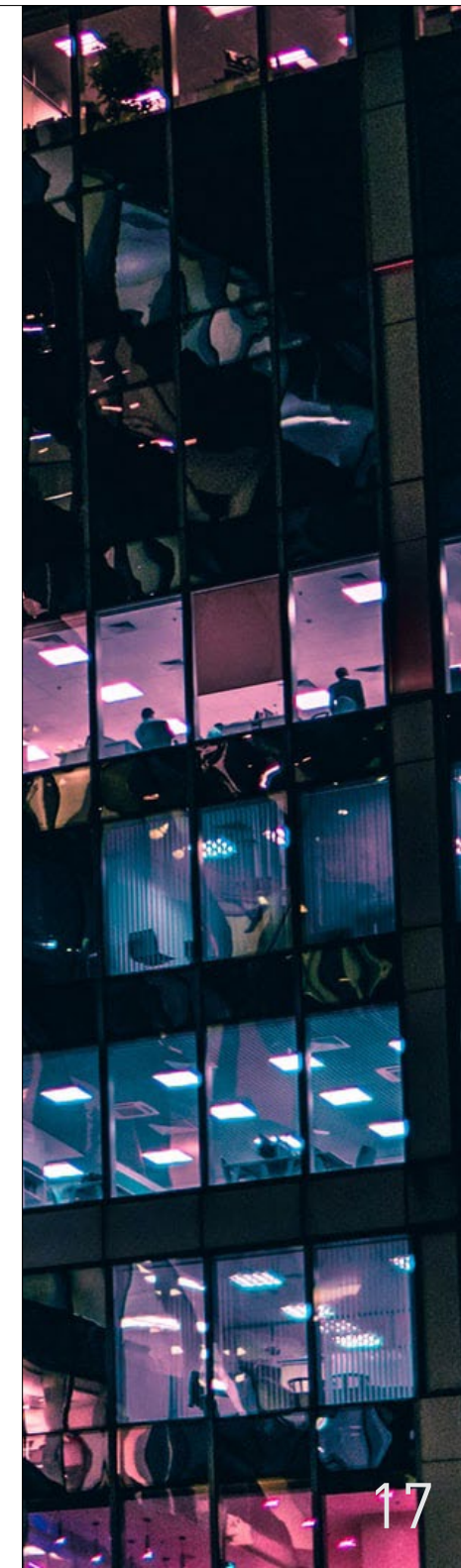


Selon le philosophe Gilbert Simondon<sup>26</sup>, toute technologie produit un "effet de halo" et emporte avec elle, au-delà de sa valeur d'usage et de sa valeur d'échange, une valeur symbolique - tantôt positive, tantôt négative - qui dépasse sa réalité propre et se répand dans les structures mentales des individus. L'ensemble des fantasmes et symboliques liés à la technologie empêche, le plus souvent, de la percevoir telle qu'elle est réellement, dans ses limites objectives et matérielles. L'observation des inquiétudes et espoirs exprimés de par le monde et suscités par l'arrivée des systèmes d'intelligence artificielle montre que ce type de technologie ne fait pas exception à la règle.

Parmi les inquiétudes les plus récurrentes, les acteurs critiques de l'IA envisagent la possibilité d'une disqualification de l'intelligence humaine, voire son remplacement, par des systèmes techniques omniscients et omnipotents. En 2014, Stephen Hawking<sup>27</sup>, physicien et cosmologiste britannique, dénonce le rythme effréné de développement des systèmes d'IA et le fait qu'ils deviendront vite incontrôlables au point de mettre l'humanité en péril. D'autres scientifiques renommés partagent son inquiétude et avertissent sur les dangers liés aux technologies d'apprentissage automatique. Selon eux, "les facultés d'apprentissage automatique des machines alimentées par le *Big Data* les rendront bientôt imprévisibles, puisque leur comportement ne résultera plus du programme que les hommes auront écrit, mais des connaissances qu'elles construiront elles-mêmes par induction automatique à partir d'informations glanées çà et là dans des bibliothèques

26. Simondon, G. (2014). L'effet de halo en matière technique : vers une stratégie de la publicité (1960). Dans : G. Simondon, Sur la technique (1953-1983) (pp. 279-293). Paris cedex 14, France : Presses Universitaires de France.

27. Griffin, A. (2014). Stephen Hawking: AI could be the end of humanity. The Independent. Disponible sur : <https://www.independent.co.uk>.





virtuelles, dans des entrepôts de données et dans le monde au cours de leurs pérégrinations. Couplée à cette imprédictibilité, leur autonomie croissante fait qu'elles nous échapperont alors et prendront un empire de plus en plus grand sur l'homme. Nous atteindrons alors un point de non-retour à partir duquel nous, humains, courrons tous à notre perte".<sup>28</sup> Cette crainte a d'ailleurs conduit nombre de chercheurs en intelligence artificielle à signer, en janvier 2015, une lettre ouverte pour alerter sur les dérives d'une IA dont les progrès techniques seraient réalisés en dehors de toute considération sur les progrès sociétaux qu'elle doit apporter. En outre, cette lettre ouverte<sup>29</sup> est complétée par un document explicitant les directions de recherche prioritaires<sup>30</sup> pour aider à maximiser les bénéfices sociétaux de l'IA.

Face à ces défiances fortes, d'autres acteurs adoptent des raisonnements différents. Pour Jean-Gabriel Ganascia<sup>21</sup>, une intelligence algorithmique est en train de se développer à grande échelle. Sur des tâches très spécifiques, cette technologie est si performante et s'améliore si vite, grâce essentiellement aux progrès en matière de puissance de calcul, qu'elle bouleverse la société qui se retrouve à vivre au temps des algorithmes<sup>31</sup>. Cependant, l'idée d'une intelligence artificielle forte qui dépasserait l'intelligence des humains n'est pas une idée vraie ou fausse, c'est à ce stade une croyance<sup>32</sup>, car elle n'est pas étayée par des arguments scientifiques. D'après certaines parties prenantes, il serait de l'intérêt de ceux qui dominent le marché du numérique de faire le croire et aux médias en quête d'audience de relayer cette croyance.

Revenant à la notion de halo symbolique, il s'avère qu'à l'heure actuelle les nouvelles technologies mettent en lumière un certain nombre de perspectives et de bouleversements et poussent les individus à s'interroger sur la société et son évolution<sup>33</sup> :

- Quelle idée se fait-on de la société ? Que devrait-elle ou ne devrait-elle jamais devenir ?
- Comment y travaille-t-on ? Comment y occupe-t-on son temps ?
- Comment y est-on en rapport avec les autres et avec l'environnement ?

Dans le cas de l'émergence de l'intelligence artificielle dans le monde du travail, ces questions peuvent se penser à l'aune des premières réalités expérimentées. Car si l'innovation technologique est encore loin de pouvoir engendrer une IA toute puissante qui remplacerait, voire asservirait les humains, les systèmes d'IA étroite<sup>34</sup> ont d'ores et déjà présenté des défauts importants liés aux

28. Extrait issu de : Ganascia, J.-G. Le mythe de la Singularité. Faut-il craindre l'Intelligence artificielle ?, Paris, Seuil, 2017.

29. Future of life institute (2015). An open letter - Research priorities for robust and beneficial Artificial Intelligence. Disponible sur : <https://futureoflife.org>.

30. Russell, S., Dewey, D. & Tegmark, M. (2015). Research Priorities for Robust and Beneficial Artificial Intelligence. AI Magazine, 36 (4), 105-114. Disponible sur : <https://ojs.aaai.org>.

31. Abiteboul, S. & Dowek, G. Le temps des algorithmes, Paris, Le Pommier, 2017.

32. Romero, M. (2018). Intelligence artificielle et pensée humaine. The Conversation. Disponible sur : <https://theconversation.com>.

33. Klein, E. (2021). Le "halo symbolique" des nouvelles technologies (nanos ou quantiques). YouTube. Disponible sur : <https://www.youtube.com>.

34. "Les systèmes d'IA étroite sont des technologies qui peuvent réaliser une seule tâche ou quelques tâches spécifiques. Ce sont les systèmes actuellement déployés." Source : Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle constitué

biais mais aussi au phénomène de boîte noire. Parmi les exemples les plus marquants, on peut notamment citer le système d'IA de recrutement d'Amazon, le chatbot de Microsoft ou encore les algorithmes prédictifs qui viennent en appui aux décisions de justice.

En 2017, l'entreprise Amazon a ainsi dû arrêter d'utiliser le système d'IA de recrutement<sup>35</sup> qu'elle avait développé en 2014. Le programme permettait de trier les CV des candidates et des candidats qui postulent à une offre d'emploi en leur attribuant une note de 1 à 5, élément indicatif de la qualité de la candidature estimée par le système. Après un an d'utilisation, l'entreprise s'est aperçue que les femmes étaient systématiquement moins bien notées que les hommes pour les offres d'emplois techniques. Ce biais de genre était en fait lié à la manière dont a été conçu le système : il a été entraîné avec les profils de personnes employées chez Amazon entre 2004 et 2014, qui étaient principalement des hommes. Avec de telles données d'apprentissage, l'algorithme d'intelligence artificielle a donc axé prioritairement ses recherches sur des profils masculins. Après une mise à jour du programme en raison des problèmes observés, l'équipe chargée du projet a finalement été dissoute car d'autres types de discriminations sont apparus.

En 2016, Microsoft lance Tay, une intelligence artificielle auto-apprenante, sous forme d'un chatbot<sup>36</sup>, capable de participer à des conversations sur les réseaux sociaux et les applications de messagerie, dont le public cible est la jeunesse américaine de 18 à 24 ans. Pour communiquer, cette technologie se base à la fois sur des données accessibles publiquement et sur des réponses pré-écrites par l'équipe de Microsoft en charge du projet. Une fois qu'elle a suffisamment échangé avec des utilisateurs, elle est également censée apprendre d'eux et personnaliser ses réponses en fonction du profil de la personne qui lui parle. L'expérience lancée par Microsoft avait pour but d'étudier la compréhension du langage. Néanmoins, un certain nombre d'internautes ayant tenté de tester ses limites, le système d'IA s'est rapidement mis à tenir des propos vulgaires et racistes. Après 8 heures de fonctionnement, Microsoft a préféré stopper son étude et retirer Tay des réseaux sociaux.

Dans le secteur de la justice, l'intelligence artificielle fait également son apparition depuis quelques années. On peut citer l'exemple d'un algorithme d'évaluation des risques<sup>37</sup> pour aider les juges américains à décider si un prévenu doit être, ou non, emprisonné avant son procès. Dans un contexte où ces professionnels ont des difficultés à évaluer les risques de récidive et le font souvent en quelques minutes avec très peu d'information, l'algorithme est censé permettre à ces derniers de s'appuyer sur des données sûres avant de prendre leur décision. Pour cela, le système d'IA attribue un score de risque de récidive à chaque individu en vérifiant son niveau d'éducation, s'il a un travail ou non, ou encore si l'un de ses parents a été envoyé en prison et en le plaçant dans un groupe d'appartenance. Outre le problème de la remise en cause du

par la Commission européenne en juin 2018, Définition de l'IA : Principales capacités et disciplines, avril 2019. Disponible sur : <https://digital-strategy.ec.europa.eu>.

35. Signoret, P. (2018). Amazon a dû désactiver une IA qui discriminait les candidatures de femmes à l'embauche, Numerama. Disponible sur : <https://www.numerama.com>.

36. Tual, M. (2016). À peine lancée, une intelligence artificielle de Microsoft dérape sur Twitter, Le Monde. Disponible sur : <https://www.lemonde.fr>.

37. De Maupéou, F. (2017). États-Unis : les algorithmes prédictifs, un danger pour la justice américaine ?, Éditions l'abécédaire des institutions. Disponible sur : <https://www.labecedaire.fr>.



droit à une justice individualisée, une étude<sup>38</sup> a révélé, après quelques années d'utilisation, que les scores donnés par l'algorithme étaient très peu fiables. En effet, selon les auteurs, deux individus de couleur de peau différente, ayant commis des crimes similaires, vont être notés différemment par l'algorithme. Ils prennent l'exemple d'une jeune femme noire de 18 ans et d'un homme blanc de 41 ans. Tous deux ont été accusés de cambriolage et de vol de matériel pour un montant avoisinant les 80 dollars. La jeune femme a un casier judiciaire pour des délits commis lorsqu'elle était mineure. L'homme, quant à lui, a déjà été reconnu coupable de vol à main armée et de tentative de vol à main armée, délits pour lesquels il a purgé cinq ans de prison, en plus d'une autre accusation de vol à main armée. Au moment de leur emprisonnement avant leur procès, l'algorithme d'évaluation des risques a calculé la probabilité que chacun des deux individus commette un crime futur : la jeune femme noire a été étiquetée comme cas à haut risque, tandis que l'homme blanc a été étiqueté comme cas à faible risque. De nombreux autres cas similaires ont été mis en lumière par cette étude. L'ensemble de ces éléments interroge sur la fiabilité de la prédiction et les risques associés à de tels usages.

Ces trois exemples illustrent une des limitations importantes des systèmes d'IA actuels : les biais<sup>4</sup> potentiels des algorithmes d'apprentissage machine. Tout d'abord, pour fonctionner dans des conditions optimales, ces algorithmes d'IA ont besoin d'énormes quantités de données d'apprentissage afin de discriminer les différentes réponses possibles. Sans cela, l'algorithme ne pourra pas atteindre l'objectif pour lequel il a été conçu ou alors fournira une réponse erronée. D'autre part, la qualité de ces données d'apprentissage est un élément crucial : dans le cas de l'algorithme de recrutement d'Amazon, des données non suffisamment équilibrées et inclusives ont conduit à une discrimination de genre. Ce phénomène semble également être à l'œuvre dans le cas de l'algorithme d'évaluation des risques de récidive utilisé par les juges. Ainsi, un système d'IA entraîné avec un ensemble de données d'apprentissage biaisées ne sera pas capable de procéder à une généralisation satisfaisante et prendra peut-être des décisions injustes, susceptibles de favoriser certains groupes par rapport à d'autres. Cependant, même si les biais sont révélés par les systèmes d'IA, leur origine ne provient pas toujours du fonctionnement de l'algorithme : il peut s'agir de biais sociaux<sup>39</sup> profondément ancrés chez l'humain et présents dans les données d'apprentissage utilisées. Ces biais humains sont des mécanismes de pensée qui agissent automatiquement et de manière inconsciente, et qui permettent à chacun de "faire le tri" dans les informations reçues en continu afin d'évaluer rapidement une situation et/ou d'engager un certain nombre d'actions. Ils sont le reflet de jugements socio-culturels historiquement construits à travers l'éducation, l'information publique, les institutions, etc. Les phénomènes de discrimination surviennent lorsque ces préjugés sont inclus dans la donnée - provenant elle-même d'une réduction voire d'un appauvrissement de l'information à un élément purement quantifiable. En effet, sortis de tout contexte et sans aucun lien avec les groupes sociaux au sein desquels s'ancrent ces préjugés, les biais sociaux sont algorithmiquement dupliqués lors de l'apprentissage machine. Pire encore, si le système d'IA est déployé massivement, on assiste à une généralisation des biais sociaux, et donc des discriminations, auparavant circonscrits.

38. Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). Machine bias - There's software used across the country to predict future criminals. And it's biased against blacks., ProPublica. Disponible sur : <https://www.propublica.org>.

39. Daoudi, L., López, A. & Cotillon, G. (2020). Une intelligence artificielle inclusive ? - Travail, inclusion sociale et processus de marginalisation, Rapport de mission de l'Université Paris 1 Panthéon Sorbonne, sous la direction d'Alain Assouline, en partenariat avec les écoles Webforce3.

Le phénomène de boîte noire<sup>4</sup>, quant à lui, concerne certaines techniques d'apprentissage automatique automatique - comme l'apprentissage profond - qui, bien qu'elles donnent de très bons résultats du point de vue de la précision, ne permettent pas à l'humain - qu'il soit utilisateur, développeur ou même concepteur - de comprendre complètement la raison de certaines décisions. L'humain peut juger des données entrantes et des résultats sortants mais ne sait pas ce qui se passe à l'intérieur du système d'IA. Cela s'avère problématique dans le cas de systèmes déployés dans des domaines tels que la banque ou l'aérospatial<sup>40</sup> au sein desquels de multiples réglementations exigent la compréhension de l'obtention des résultats à des fins de validation et de compréhension de leur applicabilité. En outre, le phénomène de boîte noire amplifie le phénomène de biais algorithmique, à partir du moment où l'opacité d'un système d'IA empêche de déceler les biais et leur origine technique.

Au regard de l'ensemble des éléments présentés jusqu'ici, il semble nécessaire de construire une stratégie de développement de l'intelligence artificielle qui soit en mesure de prendre en compte les enjeux économiques et concurrentiels, mais également les enjeux sociétaux et humains soulevés par cette transition d'une société industrielle vers une société numérique. En référence au [paragraphe 2.3](#), il serait tout d'abord pertinent de s'interroger sur les aspirations de l'ensemble de la société :

- Quelle(s) approche(s) souhaite-t-on favoriser pour construire l'avenir ?
- Sur quels fondements théoriques se base-t-on ?
- Que veut-on absolument éviter ?

Ensuite, la question de la mise en œuvre concrète de cette stratégie de développement de l'IA est à prendre en compte :

- Comment parvient-on à opérationnaliser l'ensemble des aspirations récoltées ?
- Quels éléments cruciaux sont à garder à l'esprit pour que l'IA soit une technologie bénéfique et acceptée par toutes les parties prenantes au sein de l'entreprise, notamment ?
- Quelles synthèses productives pertinentes, mêlant à la fois objets techniques et expériences utilisateurs, peut-on expérimenter sur le terrain ?

Loin de pouvoir répondre à toutes ces questions, la suite de ce document propose toutefois de présenter les éléments déjà disponibles et les principales initiatives lancées pour répondre aux différents enjeux liés à l'intelligence artificielle.

40. Georges, B. (2018). Les boîtes noires du "deep learning", Les Échos. Disponible sur : <https://www.lesechos.fr>.

003



# Le développement et le déploiement de l'intelligence artificielle, un enjeu mondial

Le déploiement de systèmes d'IA a déjà commencé à l'échelle mondiale, que ce soit en entreprise avec, notamment, l'utilisation de logiciels RH pour le recrutement ou dans la sphère privée avec la mise en œuvre d'assistants vocaux ou encore d'algorithmes de recommandation à travers les différents appareils électroniques devenus indispensables au quotidien. Dans l'immense majorité des cas, il s'agit de systèmes d'IA étroite spécialisés dans la réalisation d'une tâche unique ou d'un nombre réduit de tâches bien particulières.

À l'heure actuelle, la Chine et les États-Unis sont les deux grands leaders mondiaux<sup>41</sup> en matière d'intelligence artificielle avec une approche totalement différente en ce qui concerne la régulation des systèmes d'IA et l'exploitation du potentiel d'innovation de cette technologie :

- Les États-Unis ont une approche très libérale de la question<sup>42</sup> et énoncent des principes généraux assez vagues qui visent la mise en place d'une réglementation aussi réduite que possible ;
- La Chine, elle, adopte une approche très régulatrice<sup>43</sup> qui limite les possibilités des utilisateurs comme des entreprises et s'est récemment dotée d'un règlement pour la protection des données<sup>44</sup> sur Internet. L'autorité chargée du cyberspace a précisé que l'objectif à moyen terme est de réguler les algorithmes qui constituent une menace pour la sécurité intérieure, l'ordre social et l'ordre économique. Cependant, même si le pays s'empare de la question de l'utilisation des données par les entreprises, cette question n'est pas posée pour les administrations : c'est le pouvoir central qui a le dernier mot pour juger de ce qui est éthique ou non.

Face à ces deux puissances, tout l'enjeu de l'Europe et de la France est de parvenir à asseoir sa souveraineté et sa force économique dans le domaine en inventant un modèle spécifique en IA qui prenne en compte les spécificités de son territoire et de sa population.

41. La Cité de l'IA (2021). Livre blanc - 2021 Intelligence artificielle Hauts-de-France, IAADF. Disponible sur : <https://www.iahdf.org>.

42. Knight, W. (2020). White House favors a light touch in regulating AI, Wired. Disponible sur : <https://www.wired.com>.

43. Kerkour, T. (2021). Chine : le gouvernement édicte ses règles pour définir une intelligence artificielle "éthique", Le Figaro. Disponible sur : <https://www.lefigaro.fr>.

44. Le Figaro (2021). La Chine adopte une grande loi sur les données personnelles en ligne. Disponible sur : <https://www.lefigaro.fr>.



## 3.1

## Des initiatives internationales pour envisager le développement de l'IA

### ▸ 3.1.1 Travailler pour bâtir un avenir meilleur

Lors de la Déclaration du centenaire de l'Organisation internationale du travail (OIT)<sup>45</sup>, l'institution a renouvelé son engagement à façonner un avenir du travail équitable, inclusif et sûr, avec le plein emploi productif et librement choisi, et le travail décent pour tous. Selon elle, pour avancer et créer les conditions d'un avenir juste et durable, il est nécessaire d'investir dans les ressources humaines par le biais d'une approche centrée sur l'humain.

Dans ce cadre, l'OIT a réuni une Commission indépendante de 27 membres - personnalités mondiales du monde des affaires, des syndicats, des groupes de réflexions, d'organisations gouvernementales et non gouvernementales - qui a travaillé entre 2017 et 2019<sup>46</sup> sur les questions des nouvelles formes de travail, des ramifications institutionnelles de la nature changeante du travail, de l'apprentissage tout au long de la vie, de l'inclusion et l'égalité du genre, de la mesure de l'emploi et du bien-être des êtres humains ainsi que du rôle de la protection sociale universelle dans un avenir de travail stable et équitable. À l'issue de son mandat, la Commission ressort avec la proposition d'un programme d'actions centré sur l'humain pour l'avenir du travail qui renforce le contrat social en plaçant les hommes, les femmes et le travail qu'ils accomplissent au centre de la politique économique et sociale et des pratiques des entreprises. Ce programme est axé sur

45. OIT (2019). Cinq questions sur la Déclaration du centenaire de l'OIT. Disponible sur : <https://www.ilo.org>.

46. Commission mondiale sur l'avenir du travail (2019). Travailler pour bâtir un avenir meilleur, OIT. Disponible sur : <https://www.ilo.org>.

trois domaines d'actions complémentaires qui, ensemble, permettraient d'engendrer croissance, équité et durabilité pour les générations présentes et futures : accroître l'investissement dans le potentiel humain, accroître l'investissement dans les institutions du travail et accroître l'investissement dans le travail décent et durable. Pour atteindre cet objectif, la technologie (dont l'IA fait partie) est un outil au service du potentiel humain, des institutions du travail, des travailleurs et des employeurs, du travail décent et durable :

- Il est notamment question de renforcer la maîtrise du temps en accordant une plus grande autonomie aux travailleurs dans la gestion de leurs horaires de travail tout en répondant aux besoins des entreprises. Cela passerait par l'utilisation de la technologie pour gagner en flexibilité ainsi qu'en productivité et atteindre un équilibre entre vie professionnelle et vie personnelle, dans une société où ces deux vies sont de moins en moins compartimentées.
- Il s'agit également de gérer la technologie et de la mettre au service du travail décent. Cela sous-entend, d'une part, l'implication des travailleurs avec les gestionnaires dans la conception du travail et, d'autre part, le maintien de l'intelligence artificielle sous contrôle humain par l'adoption d'une approche qui assure que les décisions finales soient prises par des êtres humains et qui suppose l'établissement d'une réglementation en matière d'utilisation des données et de responsabilité dans l'emploi des algorithmes dans le monde du travail. En complément, la Commission propose la mise en place d'un système de gouvernance internationale pour les plateformes de travail numériques afin qu'elles et leurs clients respectent des socles de droits et de protections.
- Finalement, il est nécessaire de repenser le dialogue social et la représentation collective des travailleurs et des employeurs par l'utilisation de la technologie afin que les parties prenantes, y compris celles engagées dans l'économie des plateformes, puissent s'appuyer sur des organisations de travailleurs et d'employeurs légitimes.

Cette approche proposée par la Commission mondiale sur l'avenir du travail pose un cadre global intéressant pour envisager les initiatives focalisées sur le développement de l'IA, proposées par d'autres organisations, à l'aune des mutations en cours dans le monde du travail.

### ▸ 3.1.2 Garantir une IA digne de confiance

La perspective unique de l'Europe en matière d'IA tient au fait qu'elle s'efforce de placer les citoyens au cœur de son action. Ceci fait partie de son ADN et de ses valeurs fondatrices : les droits fondamentaux, la démocratie et l'état de droit.

En juin 2018, la Commission européenne a mandaté un Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle afin de travailler sur la stratégie de développement de l'IA selon une approche humaniste : l'IA digne de confiance. En deux ans, ce groupe de travail a produit un certain nombre de livrables dont l'ambition est de donner aux décideurs politiques nationaux et européens les premiers éléments actionnables leur permettant de se saisir des opportunités mises en lumière par l'IA et d'accélérer l'application de l'IA digne de confiance au profit des individus et des sociétés européennes. Il s'agit notamment :

- D'un ensemble d'éléments de définition communs<sup>47</sup> dans le domaine de l'IA ;
- De lignes directrices en matière d'éthique<sup>47</sup> pour une IA digne de confiance ;
- De recommandations globales de politique et d'investissement<sup>48</sup> pour une IA digne de confiance ;
- De considérations sectorielles sur les recommandations de politique et d'investissement<sup>49</sup> pour une IA digne de confiance ;
- D'une liste de critères pour une IA digne de confiance sous la forme d'un outil<sup>50</sup> qui traduit les lignes directrices en matière d'éthique en une liste de contrôle (auto-évaluation) accessible et dynamique. Cet outil peut être utilisé par les développeurs et les déployeurs d'IA qui souhaitent mettre en œuvre les exigences clés dans la pratique.

D'après le groupe d'experts, une IA digne de confiance présente trois caractéristiques qui fonctionnent en harmonie et se chevauchent tout au long du cycle de vie du système : la légalité via la garantie du respect des législations et réglementations applicables, l'éthique via l'adhésion à des principes et valeurs éthiques ainsi que la robustesse sur le plan technique et social. Si, du point de vue législatif, des exigences juridiques existent déjà et/ou sont en cours de construction (la Commission européenne travaille actuellement sur un nouveau règlement européen sur l'IA<sup>51</sup>, dont l'approche est basée sur les risques et les niveaux de risques) du point de vue éthique, il

47. Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle constitué par la Commission européenne (2019). Lignes directrices en matière d'éthique pour une IA digne de confiance. Disponible sur : <https://digital-strategy.ec.europa.eu>.

48. Independent high-level expert group on artificial intelligence set up by the European Commission (2019). Policy and investment recommendations for trustworthy AI. Disponible sur : <https://digital-strategy.ec.europa.eu>.

49. Independent high-level expert group on artificial intelligence set up by the European Commission (2020). Sectoral considerations on the policy and investment recommendations for trustworthy AI. Disponible sur : <https://futurium.ec.europa.eu>.

50. European AI Alliance. ALTAI - The assessment list on trustworthy artificial intelligence, European Commission. Disponible sur : <https://futurium.ec.europa.eu>.

51. Vie publique (2021). Intelligence artificielle : un nouveau règlement européen sur l'IA. Disponible sur : [www.vie-publique.fr](http://www.vie-publique.fr).

semble important de formaliser un ensemble de principes et de faire en sorte d'obtenir l'adhésion de la majorité des parties prenantes de l'IA.

De façon synthétique, les principes éthiques considérés par le groupe d'experts dans le contexte des systèmes d'IA sont les suivants :

1. Respect de l'autonomie humaine  
L'humain doit être en mesure de superviser et contrôler les processus de travail des systèmes d'IA. En outre, les systèmes d'IA doivent avoir pour objectif d'aider les humains dans leur environnement de travail et de créer des emplois qui ont du sens.
2. Prévention de toute atteinte  
Les systèmes d'IA doivent être conçus de sorte à ne jamais porter atteinte à la dignité humaine ni à l'intégrité mentale et physique des êtres humains. Plus largement, l'environnement naturel et l'ensemble des êtres vivants doivent être pris en compte dès la conception des systèmes d'IA.
3. Équité  
Les systèmes d'IA mis au point, déployés et utilisés doivent être équitables, tant sur un volet matériel - répartition égale et juste des bénéfices et des coûts, vigilance quant aux biais, égalité des chances pour l'accès à l'éducation, aux biens, aux services et à la technologie, etc. - que sur un volet procédural - capacité à contester les décisions prises par les systèmes d'IA et les humains qui les utilisent, mise en place d'un recours efficace à l'encontre des décisions, etc.
4. Explicabilité  
Le caractère explicable des algorithmes des systèmes d'IA doit permettre de renforcer et conserver la confiance des utilisateurs envers cette technologie, au moyen de processus de communication transparents, notamment.

Ces principes ont ensuite été déclinés en sept exigences clés auxquelles les systèmes d'IA doivent satisfaire pour être jugés dignes de confiance. On les retrouve dans la liste d'évaluation<sup>50</sup>, précédemment mentionnée, qui a pour objectif d'aider à vérifier l'application de chacune de ces exigences. Toutefois, en plus de la phase d'évaluation, le groupe d'experts fournit également toute une palette de méthodes techniques et non techniques qui doivent permettre de concevoir, déployer et utiliser des systèmes d'IA dignes de confiance.

L'ensemble de ces travaux sur les considérations éthiques constitue, pour le groupe d'experts, un premier pas crucial qui permet d'arbitrer sur le type d'IA souhaité et le type d'IA non souhaité en Europe. Ils sont ensuite complétés par des recommandations politiques et d'investissement<sup>48</sup> sur la manière dont une IA digne de confiance peut être développée, déployée, encouragée et mise à l'échelle, à la fois en maximisant ses avantages et en minimisant, prévenant ses risques. Ces recommandations sont adressées aux institutions européennes et aux États membres et sont, finalement, déclinées dans quatre secteurs stratégiques<sup>49</sup> : le secteur manufacturier et industriel de l'Internet of Things, l'administration numérique (e-government domain), la justice et l'application de la loi, le secteur de la santé.



### ▸ 3.1.3 PMIA et avenir du travail

Lancé en juin 2020, le Partenariat mondial sur l'intelligence artificielle (PMIA) est une initiative multipartite qui réunit des experts de premier plan de la science, de l'industrie, de la société civile, des organisations internationales et du gouvernement. En tant qu'initiative basée sur l'expertise, le PMIA travaille sur des questions spécifiques d'IA : l'IA responsable, la gouvernance des données, l'avenir du travail ainsi que l'innovation et la commercialisation.

Au sein du groupe de travail sur l'avenir du travail, l'objectif est de mener des analyses sur la manière dont le déploiement de l'IA peut affecter les travailleurs et leur environnement de travail, ainsi que sur la manière dont les travailleurs et les employeurs peuvent mieux concevoir l'avenir du travail. Dans son premier rapport de novembre 2020<sup>16</sup>, le groupe de travail présente les différents livrables envisagés pour répondre à l'enjeu de compréhension collective de l'incidence de l'IA sur le monde du travail à travers l'analyse de cas réels existants, mais également à travers un exercice de projection vers une vision du futur. Ces derniers peuvent être classés en six catégories :

- **Cas d'utilisation en milieu de travail**  
Ce livrable consiste à compiler et analyser différentes situations des systèmes d'IA : expériences en cours ou terminées, cas d'utilisation réels de l'IA à l'échelle de l'entreprise, aperçu de l'état actuel des interfaces d'IA et des processus alimentés par l'IA du point de vue des travailleurs.
- **Formation**  
Ce livrable consiste à recenser et élaborer les meilleures méthodes de formation techniques basées sur l'IA pour l'amélioration des compétences des travailleurs sur les emplois actuels et les emplois de l'avenir. Il s'agit de parvenir à former grâce à l'IA. En parallèle de cette dynamique, on peut observer que certains acteurs de l'IA s'emparent également du sujet de former à l'IA. Par exemple, IBM a lancé l'Académie IBM<sup>52</sup> dont l'objectif est de former des demandeurs d'emploi aux métiers de demain pour devenir développeurs d'applications spécialisés en Intelligence Artificielle. Les apprenants lauréats obtiennent un titre professionnel RNCP de niveau 6 (Bac+3) et 13 badges numériques de certifications IBM.
- **Collaboration homme-machine**  
Ce livrable consiste en l'analyse des possibilités techniques de collaboration homme-machine (IHM), de la coévolution et de la délégation automatisée des prises de décisions en milieu de travail ainsi que leurs incidences sur la santé mentale et physique des travailleurs et sur les organisations.
- **Gestion des biais**  
Ce livrable consiste à envisager des perspectives sur les biais et les inégalités engendrés par l'IA ainsi que des perspectives politiques, éthiques et techniques sur les mesures de correction.

52. <https://formation.needemand.com/academie-ibm/>

- **Conditions de travail**  
Ce livrable consiste à analyser la manière dont des conditions de travail décentes et positives peuvent être encouragées par les organisations dans le cadre de situations de travail caractérisées par une croissance de l'utilisation des systèmes d'IA.
- **Recherche ouverte et innovation**  
Ce livrable consiste en la mise en place d'un laboratoire de recherche ouverte en innovation sur l'avenir du travail sous la forme d'une plateforme Web, d'un lieu physique ou d'un réseau pour l'échange sur des expériences de systèmes d'IA aux niveaux individuel et organisationnel.

La suite de ce paragraphe se focalise sur l'analyse des cas d'utilisation en milieu de travail. Pour mener à bien ce travail, une méthode de sondage commune a été conçue, avec comme contrainte la prise en compte des différences culturelles existantes à travers le monde. Cette méthode prend la forme d'un questionnaire portant sur une application professionnelle donnée de l'IA, et non sur une technique d'IA. Le groupe de travail a fait le choix de collecter des cas d'utilisation les plus diversifiés possibles pour parvenir à saisir ce qui se passe dans le contexte réel du milieu du travail et constituer un catalogue à valeur empirique pour alimenter les réflexions des chercheurs du groupe de travail. Au terme de deux mois de travaux (octobre-novembre 2020), le catalogue comportait 53 cas d'utilisation, répartis sur huit pays. Ces cas d'usage proviennent principalement de directeurs, de gestionnaires et de concepteurs. Il est également à noter que les répondants sont majoritairement des hommes.

Les premières conclusions tirées de cette recherche sont les suivantes :

- Deux principales fonctions de l'IA ressortent dans ces 53 cas d'utilisation : une fonction d'aide à la décision grâce à laquelle l'IA aide le travailleur à comprendre les problèmes qu'il doit résoudre et une fonction d'outil grâce à laquelle les tâches des travailleurs sont simplifiées, rendues plus accessibles, moins fastidieuses, plus rapides et plus efficaces.
- Les systèmes d'IA pourraient être utilisés pour augmenter l'humain en venant en support sur trois champs différents :
  - L'amélioration des organisations en produisant des connaissances nouvelles et en complétant les tâches humaines, notamment grâce aux technologies d'apprentissage automatique (simplification des tâches afin qu'elles puissent être effectuées sans beaucoup de formation, compensation du manque d'expertise humaine, conception de programmes qui détectent de subtiles différences ou anomalies, ...)
  - L'amélioration de l'expérience par l'ajout de nouvelles applications en matière de ressources humaines (gestion de carrière, mesure du bonheur, gestion des compétences, meilleur accès à l'information, formation professionnelle mieux adaptée aux besoins des employeurs, ...)
  - Le renforcement des travailleurs par un système d'IA qui les assiste lorsqu'ils effectuent leur travail (nouvelle division du travail entre humain et machine, amélioration des interactions humain-machine, synthèse du meilleur de l'humain et de la machine).

- La signification d'un système d'IA "centré sur l'humain" dépend du contexte de déploiement des cas d'utilisation :
  - Les acteurs métiers (travailleurs, managers, ...) sont essentiels à la conception et à l'amélioration des systèmes d'IA.
  - L'utilisation des systèmes d'IA nécessite une forte intervention humaine du fait de son approche probabiliste ou empirique. La connaissance de l'environnement et des contextes étant limitée pour les systèmes d'IA, il est nécessaire que les humains fournissent ces renseignements et corrigent les erreurs éventuelles afin de prendre une décision.
  - Dans les cas recensés par le groupe de travail, les travailleurs participent toujours à la conception des systèmes d'IA alors que les partenaires sociaux sont rarement pris en considération.
- Les problèmes éthiques liés aux systèmes d'IA sont interprétés différemment :
  - Il semble que le niveau de maîtrise technique des facteurs éthiques par les répondants soit très bas. La question de la transparence par exemple est interprétée comme la question de l'explicabilité et de l'interprétabilité des systèmes d'IA, ou encore comme un élément concernant leur propre pratique. En revanche, les questions sur les données personnelles et la notion de responsabilité semblent mieux appréhendées.
  - De nombreux clients n'ont pas d'exigences particulières sur la transparence des systèmes d'IA, ils sont uniquement axés sur les résultats.
  - Les avis sur les aspects éthiques de l'IA sont contrastés.  
 Dans le secteur privé, les répondants estiment que le questionnaire éthique est une préoccupation intellectuelle qui n'émerge pas au niveau opérationnel : à ce niveau, l'utilité et l'efficacité du système d'IA sont les deux éléments à considérer en priorité. En outre, certains répondants jugent également ces questions inappropriées car pour eux, le paradigme de l'IA probabiliste sous-entend d'abandonner certaines formes de contrôle de sorte à ce que les systèmes d'IA apportent une réelle valeur ajoutée aux organisations. Enfin, se distingue une dernière catégorie de répondants pour lesquels le questionnaire éthique dans le domaine de l'IA est disproportionné et injuste alors qu'il peut être pratiquement inexistant dans d'autres industries.  
 Dans le secteur public, les normes d'éthique sont particulièrement élevées avec, au centre des préoccupations, les questions de discriminations, notamment. Ainsi, dans ce secteur, les problèmes éthiques sont mieux compris comparativement au secteur privé.
- Les systèmes d'IA ont une incidence sur le nombre d'emplois et la qualité de vie au travail :
  - Dans pratiquement tous les cas recensés, le remplacement du travail humain n'est pas la motivation première au déploiement d'un système d'IA ; il s'agit plutôt d'apporter un soutien aux travailleurs.
  - Des incidences indirectes sur l'emploi commencent à être observées, telles que la suppression des embauches de travailleurs temporaires, l'élimination de certains emplois chez les fournisseurs externes, la réorientation de certains travailleurs de l'entreprise au

détriment de nouvelles embauches.

- Les systèmes d'IA entraîneraient le plus souvent une amélioration des conditions de travail en automatisant les tâches les plus fastidieuses, ennuyeuses ou physiquement exigeantes. Cependant, il existe tout de même des cas où la satisfaction au travail peut se détériorer et la charge de travail augmenter.
- Selon de nombreux répondants, l'intégration de systèmes d'IA entraîne une réaffectation du travail humain vers des tâches plus gratifiantes à plus forte valeur ajoutée ou vers des tâches liées au développement de systèmes d'IA.
- Finalement, il a été observé que l'incidence des systèmes d'IA sur l'expertise humaine est hétérogène. Dans certains cas, ils renforcent la position des experts en orientant le travail humain vers des tâches à forte valeur ajoutée. Dans d'autres cas, ils sont utilisés pour compenser les limites de certains travailleurs, ce qui peut avoir pour conséquence d'affaiblir le statut d'expert.

Suite à l'analyse de ce premier échantillonnage de cas d'utilisation de l'IA, les experts du groupe de travail ont mis en évidence certaines limites à dépasser. Tout d'abord, outre la question du nombre de cas d'utilisation recensés, le catalogue initié est très déséquilibré dans le sens où il ne reflète pas suffisamment l'expérience des employés, des usagers ou encore des représentants du personnel. Il leur paraît donc essentiel de travailler sur la représentativité des répondants afin de dresser un portrait complet des effets sociaux des systèmes d'IA en milieu de travail et au sein des organisations. Enfin, aucune analyse culturelle n'a été menée sur les 53 cas d'utilisation recensés. Pour ce faire, il est nécessaire de récolter un très grand nombre de cas provenant de chaque espace culturel. Sans cette analyse, les spécificités culturelles dans la façon de mettre en œuvre l'IA en milieu de travail et au sein des organisations ne pourront pas être relevées et cela conduirait à des généralisations inexactes et des stéréotypes.

Sur l'année 2021, le groupe de travail sur l'avenir du travail s'est concentré sur un projet de mise en place d'une plateforme d'observation de l'IA sur le lieu de travail, qui intègre le catalogue empirique des cas d'utilisation développé en 2020, ainsi que sur un projet opérationnel permettant de proposer des principes et processus de l'IA équitable. Les résultats relatifs à ces deux projets ont été présentés lors du deuxième sommet du PMIA qui a eu lieu à Paris les 11 et 12 novembre 2021.



# 3.2

## Des initiatives pour le développement de l'IA à l'échelle de la France

La prise de conscience par les entreprises de l'importance de l'intelligence artificielle se situe dans les années 2015. Depuis cette date, un nombre croissant d'entre elles met en œuvre de l'IA<sup>53</sup> : (i) le nombre d'entreprises dans le monde mettant en œuvre l'IA en 2019 a augmenté de 270 % par rapport à 2015 et ce nombre a triplé au cours de l'année 2018 ; (ii) en France, 30 % des organisations avaient déployé de l'IA début 2020 et 37 % avaient intensifié les explorations de projets portés par un système d'IA. Cependant, on peut également observer des niveaux de maturité disparates de l'IA, qui dépendent des secteurs considérés : les secteurs de la santé, du transport & logistique, de la banque & assurance sont les plus avancés tandis que le niveau de maturité de l'IA pour le tourisme ou l'agriculture reste relativement faible.

La stratégie nationale pour l'IA a, quant à elle, été initiée en 2017 avec comme première étape la réalisation d'un état des lieux de l'IA en France<sup>24</sup>, effectué par Cédric Villani. De cette étude, le gouvernement a pu établir un plan d'investissement à horizon 2022<sup>54</sup> : la stratégie française pour l'IA "AI for Humanity". Les différentes initiatives présentées dans la suite de ce paragraphe s'inscrivent dans cette stratégie.

53. Perspectives IA (2020). Synthèse des études disponibles sur le domaine de l'intelligence artificielle. Disponible sur : <https://perspectives-ia.fr>.

54. <https://www.aiforhumanity.fr/>

### ▸ 3.2.1 Synthèse du rapport Villani

En 2018, Cédric Villani, mathématicien et député de l'Essonne, terminait sa mission parlementaire et publiait, avec son équipe, le rapport<sup>24</sup> "Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne".

Dans ce contexte, une réflexion a été menée avec son équipe pour penser une stratégie d'intelligence artificielle, à la fois dans un cadre national et international, mais également avec tous les acteurs de la société (sciences exactes et humaines, administrations, entrepreneurs, journalistes, auteurs de science-fiction, etc.). Il en ressort une proposition globale qui couvre des champs multiples : politique, recherche, social, économie, éducation, écologie, éthique, ou encore inclusion.

Afin de rééquilibrer les rapports de force entre les très grands acteurs de l'IA, et du numérique en général, et les autres acteurs, le groupe de travail recommande la mise en place d'une politique articulée autour de la donnée et combinée à un objectif de souveraineté. En effet, les données doivent avant tout servir les intérêts de la France et de l'Europe et une position ferme doit être assumée en ce qui concerne le transfert de données hors de l'Union européenne (UE). En outre, l'idée est que l'État soit en mesure d'impulser une dynamique forte en IA sur les secteurs qui constituent un avantage comparatif et une niche d'excellence nationale : santé, environnement, transport-mobilités et défense-sécurité.

Cette politique va de paire avec l'ambition d'une recherche agile et diffusante. La recherche française en IA et en mathématiques occupe une place de choix au niveau mondial. Cependant, on assiste également à de profondes mutations du paysage de la recherche en intelligence artificielle : tous les grands acteurs du domaine ont ouvert des centres de recherche fondamentale richement dotés, implantés dans des zones favorables au développement scientifique, où l'on trouve des chercheurs et des étudiants performants. À la difficulté historique de la recherche publique française pour transformer ses avancées scientifiques en applications industrielles et économiques s'ajoute donc une problématique de fuite des cerveaux, non seulement vers les institutions académiques étrangères, mais plus récemment également vers les grandes entreprises mondiales. Pour répondre à cet enjeu, le groupe de travail propose de fédérer les acteurs de la recherche autour d'Instituts interdisciplinaires d'intelligence artificielle (3IA) autonomes et coordonnés, répartis sur l'ensemble du territoire national, avec un accent fort sur l'interdisciplinarité, notamment vis-à-vis des sciences humaines et sociales. Le document inclut :

- Des propositions de dotations conséquentes pour attirer les scientifiques français et étrangers ainsi que les partenaires privés et rendre plus attractives les carrières dans la recherche publique (salaires ou compléments de salaires, moyens de calcul de pointe, accès à un cloud sur mesure et adapté aux spécificités de l'IA, etc.) ;
- Des propositions de dispositifs administratifs spécifiques aux 3IA (allègement drastique des formalités administratives du quotidien, aides pour l'amélioration de la qualité de vie, modes de collaboration spécifiques, simples et rapides avec les partenaires privés, etc.).

Du point de vue de l'impact de l'IA sur le travail et l'emploi, il est nécessaire d'anticiper les grandes transformations à venir pour s'y préparer et expérimenter de nouveaux dispositifs qui seront plus à même de répondre aux besoins de la société. En ce sens, le groupe de travail propose la création d'un lab public de la transformation du travail qui s'apparenterait à une tête chercheuse à l'intérieur des politiques publiques de l'emploi et de la formation professionnelle et ferait le lien entre des prévisions macro-économiques et d'analyse des mutations des usages, d'une part, et des expérimentations concrètes de nouvelles logiques de transition professionnelle, d'autre part. Les éléments issus de ce dispositif permettraient d'alimenter un certain nombre de réflexions concernant les formations initiale et continue tout au long de la vie pour parvenir à outiller réellement l'ensemble de la population sur les questions techniques et numériques.

En plus des éléments précédents, la stratégie de développement de l'IA proposée par Villani et son équipe doit s'inscrire de façon systémique dans les grands enjeux mondiaux actuels :

- Du point de vue de la soutenabilité, il est nécessaire de penser l'IA sur ses aspects économiques, mais également écologiques et de, potentiellement, pouvoir utiliser cette technologie pour penser l'impact de l'humain sur son environnement. Ainsi, les systèmes d'IA devraient être conçus de façon à s'intégrer dans le cadre des objectifs de l'ONU sur le développement durable (ODD), de l'Accord climat et du Pacte mondial pour l'environnement.
- D'un point de vue éthique, il est indispensable de faire émerger des technologies d'IA qui soient conformes aux valeurs et normes sociales françaises et européennes. On retrouve ici des propositions qui entrent en résonance avec le travail réalisé par la Commission européenne (voir [paragraphe 3.1.2.](#)) et des notions telles que l'explicabilité des algorithmes, l'auditabilité des systèmes d'IA, penser l'éthique dès la conception et l'intégrer à l'ensemble du cycle de vie des technologies d'IA, etc.
- Finalement, du point de vue de l'inclusion, le groupe de travail recommande d'agir pour l'égalité avec notamment un fort besoin d'atteindre la parité hommes / femmes au sein des métiers du numérique et de l'IA. Il invite également à développer la médiation numérique et l'innovation sociale pour faire en sorte que l'IA bénéficie à tous avec, notamment, le maillage d'un "réseau du prendre soin" sur l'ensemble du territoire français.

Sur ces derniers aspects, tout l'enjeu est de faire en sorte que toutes les parties prenantes de l'IA, dont le consommateur / utilisateur soient des acteurs proactifs, capables de contribuer au verdissement des technologies numériques intelligentes, à même de questionner les principes éthiques d'un système d'IA et de poser un regard critique sur les résultats obtenus, ou encore conscients de la responsabilité collective à porter pour s'assurer qu'à l'avenir, personne ne soit mis de côté dans la société numérique.

### ▸ 3.2.2 D'un EDEC pilote à un EDEC national

En 2019, l'État, la région Hauts-de-France, le MEDEF Lille Métropole et l'OPCO (Opérateur de compétences) Akto ont décidé de lancer une expérimentation régionale portée dans le cadre d'un engagement de développement de l'emploi et des compétences (EDEC). Ce pilote IA-HDF (pour Intelligence artificielle Hauts-de-France) avait pour objectif d'impulser dans la région une dynamique collective autour de ce sujet et d'anticiper toutes les évolutions professionnelles induites.

Après avoir commencé par une étude prospective<sup>55</sup> concernant l'impact de l'IA sur les emplois et les compétences en Hauts-de-France, un certain nombre d'actions<sup>41</sup> ont été mises en œuvre pendant deux ans pour servir les objectifs de l'expérimentation.

La Cité de l'IA, un hub créé à l'initiative du MEDEF Lille Métropole, a permis de fournir un lieu, un réseau et des services autour de l'intelligence artificielle. Elle est un acteur central pour la production de webinars, d'études, de réflexions sur des thématiques fondamentales, de veille sur les actualités du monde de l'IA, d'événements de sensibilisation, de worklabs #IA, de services et d'accompagnement aux entreprises, etc. Plus spécifiquement, elle propose un dispositif d'accompagnement aux PME et PMI des Hauts-de-France pour qu'elles puissent bénéficier, grâce à un cofinancement public, de sept jours d'expertise afin d'intégrer l'IA dans leurs stratégies<sup>56</sup> et process. Des sessions de sensibilisation gratuites d'une demi-journée à destination des comités de direction, des managers, des salariés et des branches professionnelles sont également proposées. À mi-chemin entre ces deux dispositifs, les worklabs #IA permettent d'aborder une thématique spécifique, sur la base de l'expérience d'un grand témoin, via une approche résolument opérationnelle. Enfin, réalisés en partenariat avec EuraTechnologies, des ateliers IA permettent à des start-ups de présenter aux TPE-PME des solutions simples, opérationnelles et facilement intégrables. Parallèlement à cela, des groupes de réflexion travaillent sur des thématiques cruciales en intelligence artificielle et alimentent les différents dispositifs présentés avec les informations fondamentales à retenir et à partager sur chacune des thématiques :

- Pour une IA éthique et responsable ;
- De l'idée à la mise en œuvre ;
- Collaborations entreprises et labos de recherche, mode d'emploi ;
- IA et RH.

En complément, les acteurs du projet IA-HDF ont sélectionné l'IMT Lille Douai pour créer la première certification nationale Chef de projet IA. La première promotion a été lancée le 13 avril 2021. Le parcours de formation certifiée des chefs de projet en intelligence artificielle maîtrisant le pilotage opérationnel de projets depuis l'analyse du besoin jusqu'à l'accompagnement à la conduite

55. La Cité de l'IA (2019). Étude prospective IA-HDF 2019. Impact de l'intelligence artificielle sur les emplois et les compétences en Hauts-de-France, IA-HDF. Disponible sur : <https://www.iahdf.org>.

56. Sur le sujet de la stratégie IA des entreprises, un guide concernant le développement d'une stratégie de valorisation des données a également été créé par le Mouvement des Entreprises de France. Il met en évidence les domaines de valorisation des données au sein de l'entreprise, précise les axes à considérer avant d'engager une telle démarche et explicite les différentes étapes nécessaires à son implémentation au sein de la structure. Ce guide est disponible sur : <https://www.iahdf.org>.



du changement. Finalement, en ce qui concerne les actions de la formation professionnelle, l'OPCO Atko accompagne les entreprises dans la structuration et la mise en œuvre de leur plan de développement des compétences pour faire face aux mutations technologiques. À ce titre, de nombreuses formations en intelligence artificielle sont financées par cet opérateur. Elles s'adressent aussi bien aux métiers purement techniques de l'IA qu'aux autres fonctions de l'entreprise (RH, commercial, finances, gestion, logistique, ...) également impactées par l'arrivée des systèmes d'IA.

Dans la poursuite de cette action IA-HDF, le ministère du Travail, de l'Emploi et de l'Insertion a lancé, en 2020, une démarche nationale avec la Coalition française pour les compétences numériques (French Digital Skills and Jobs Coalition, représentée par le MEDEF), sous le label PERSPECTIVES IA pour accompagner les dirigeants d'entreprises, leurs collaborateurs et le grand public vers une meilleure connaissance de l'intelligence artificielle, de ses usages et des opportunités qu'elle représente.

Outre plusieurs guides d'accompagnement pour lancer une démarche IA<sup>57,58,59,60,61,62</sup>, un outil d'auto-diagnostic<sup>63</sup> est mis à disposition des entreprises pour qu'elles identifient, dans leur contexte, les actions qui sont à conduire pour bénéficier efficacement des effets positifs de l'intelligence artificielle. De plus, le dispositif propose des fiches claires et synthétiques<sup>64</sup> sur des projets d'IA déjà expérimentés en entreprise, représentatifs des différents usages possibles, transposables d'un secteur et d'une entreprise à une autre, afin de donner un panorama des applications concrètes.

À titre d'exemple, le cas d'usage "Veille et analyse concurrentielle"<sup>65</sup> décrit les principales fonctionnalités d'un système d'IA ayant pour objectif d'analyser le contenu des offres de la concurrence afin de réaliser une veille concurrentielle et de détecter rapidement les nouvelles tendances. Il indique également les technologies utilisées dans le système d'IA et leur niveau de maturité, les bénéfices attendus, les ressources requises et conditions d'implémentation pour déployer la solution ainsi que les secteurs économiques et fonctions de l'entreprise pour lesquels elle est applicable. Au sein de ce cas d'usage, une analyse de l'impact de la solution sur les métiers, missions et compétences des travailleurs est fournie et finalement, une entreprise témoin livre, selon elle, les facteurs clés de succès à prendre en compte pour un déploiement optimal et une bonne acceptabilité de la solution.

57. Perspectives IA (2021). Guide d'accompagnement pour lancer une démarche d'IA. Disponible sur : <https://perspectives-ia.fr>.

58. Perspectives IA (2021). L'intelligence artificielle au sein du marketing, commercial et de la relation client. Disponible sur : <https://perspectives-ia.fr>.

59. Perspectives IA (2021). L'intelligence artificielle au sein des fonctions de direction. Disponible sur : <https://perspectives-ia.fr>.

60. Perspectives IA (2021). L'intelligence artificielle au sein de la finance et de la comptabilité. Disponible sur : [perspectives-ia.fr](https://perspectives-ia.fr).

61. Perspectives IA (2021). L'intelligence artificielle au sein des fonctions logistique et achat. Disponible sur : [perspectives-ia.fr](https://perspectives-ia.fr).

62. Perspectives IA (2021). L'intelligence artificielle au sein des ressources humaines. Disponible sur : [perspectives-ia.fr](https://perspectives-ia.fr).

63. <https://perspectives-ia.fr/auto-diagnostic/>.

64. [https://perspectives-ia.fr/recherche/?type\\_de\\_publication=cas-dusage](https://perspectives-ia.fr/recherche/?type_de_publication=cas-dusage)

65. Perspectives IA (2020). Veille et analyse concurrentielle. Disponible sur : [perspectives-ia.fr](https://perspectives-ia.fr).

Au total, 17 cas d'usages sont disponibles dans des secteurs variés - commerce, santé, industrie, construction, tertiaire supérieur, trans-sectoriel, etc. - pour répondre aux besoins actuels du monde économique - conduite du changement, données, performance, etc.

### ▸ 3.2.3 Repenser le dialogue social

Comme observé précédemment, l'IA bouleverse le travail dans son ensemble et pose des questions inédites en matière de compétences, d'éthique, de transparence, d'usage des données, de création, de captation et de partage de la valeur produite. Répondre à ces questions nécessite de passer par le dialogue social afin de parvenir à réunir autour de la table toutes les parties prenantes de l'intelligence artificielle pour co-construire un nouveau cadre de travail et de nouveaux outils technologiques qui répondent pleinement aux besoins des entreprises et des travailleurs. Néanmoins, sur un sujet aussi pointu que l'intelligence artificielle, le dialogue à mener est autant social que technologique<sup>66</sup>, et les citoyens, salariés et agents publics, comme leurs représentants, ne possèdent pas forcément toutes les clés pour réellement prendre part aux discussions tournant autour de ces enjeux.

Dans ce contexte, l'Ugict-CGT a établi un guide intitulé "Intelligence artificielle & algorithmes : pour quelle 'robolution' ?"<sup>67</sup> qui a pour objectif d'apporter des éléments de compréhension sur ce que sont les systèmes d'IA et les algorithmes ainsi que sur les changements bénéfiques qu'ils peuvent potentiellement apporter dans le monde du travail puis plus largement la société. Tout au long de ce guide, les auteurs montrent que la conception, les conditions de mise en place, et l'utilisation des systèmes d'IA nécessitent un véritable débat sociétal. En outre, ils considèrent l'intelligence artificielle et les algorithmes comme "une occasion (...) de revoir profondément les relations sociales et le dialogue social en entreprise et dans l'administration qui sont en panne aujourd'hui."

Le syndicat français de l'encadrement (la CFE-CGC)<sup>68</sup> semble faire le même constat. En effet, il a identifié le besoin de "réinventer" le dialogue social afin qu'il s'empare de deux dimensions clés :

- La question des données et de l'IA comme mode de production, appelant à un partage renouvelé de la valeur économique créée ;
- La question des données et de l'IA comme moteur de développement des compétences des métiers et des professionnels de proximité.

Pour y répondre, le projet SéCoIA Deal, cofinancé par la Commission européenne, a été lancé en avril 2021 pour une durée de deux ans. Il vise à regrouper une communauté d'acteurs suffisamment large pour pouvoir incarner les différents intérêts impactés par le développement de l'économie de la donnée et de l'IA et être en mesure de travailler de façon systémique sur

66. Binaire (2019). Dialogue technologique, dialogue social, Le Monde. Disponible sur : <https://www.lemonde.fr/blog/binaire>.

67. Ugict-CGT (2020). Intelligence artificielle & algorithmes : pour quelle "robolution" ?. Disponible sur : <https://syndicoop.fr>.

68. CFE-CGC (2021). SéCoIA Deal : servir la confiance dans l'IA par le dialogue. Disponible sur : <https://www.cfecgc.org>.



plusieurs thématiques :

- Contribuer à une meilleure compréhension et appropriation des enjeux et impacts des données et de l'IA ;
- Préfigurer une négociation collective nouvelle, associant les différentes parties prenantes ;
- Formuler des recommandations génériques et diffusables visant à développer un dialogue social enrichi.

Un site internet dédié devrait être mis en ligne prochainement pour permettre au plus grand nombre de suivre les activités du projet de s'en approprier les axes de réflexion et de conclusions.

004



**Pour une  
IA inclusive  
et de  
confiance**

Suite à tout ce qui a été présenté précédemment, l'ambition de ce dernier paragraphe est de se focaliser sur la notion d'IA inclusive et de confiance, cœur des stratégies française et européenne, et de réunir un ensemble d'éléments qui pourront venir nourrir de futures réflexions sur cette thématique.

La notion d'IA de confiance a été développée dans les paragraphes précédents. Pour l'Europe et la France, il s'agit de matérialiser un domaine technologique de pointe par la conception, le développement et la mise en œuvre de systèmes qui répondent à des exigences réglementaires :

- Ils doivent être accessibles et compréhensibles par les êtres humains (principes de transparence, d'explicabilité et d'interprétabilité) ;
- Leur design doit inclure la prise en compte ainsi que le respect de l'environnement et de l'ensemble des êtres vivants (principe de prévention de toute atteinte).

Dans le même temps, ces systèmes d'IA - conçus, développés et mis en œuvre - doivent être utiles et répondre à un certain nombre d'usages :

- Ils doivent constituer un support opérationnel dans la transformation du monde du travail vers la création d'emplois porteurs de sens et à forte valeur humaine ajoutée (principe de respect de l'autonomie humaine) ;
- Ils doivent permettre de lutter contre les disparités et les fractures sociétales existantes dans un cadre qui assure à chaque être humain la capacité de poser un regard critique sur, voire de contester, les résultats et décisions des machines (principe d'équité).

La notion d'inclusion est également portée par les différentes initiatives internationales et nationales présentées jusqu'ici. Dans son rapport<sup>24</sup>, Cédric Villani déclare à ce sujet qu'il "est de notre responsabilité collective de s'assurer que personne ne soit mis de côté". Néanmoins, cette notion même est ambivalente, au point que chacun semble avoir sa propre définition.

## De l'inclusion

# 4.1

Étymologiquement, le terme inclusion provient du latin et s'est construit sur la racine cludere qui signifie "fermer", "interdire l'accès à". Il est donc intimement lié au terme exclusion que l'on peut définir, dans le contexte qui nous intéresse, comme le fait de rejeter un individu jugé comme incompatible par rapport à un autre<sup>69</sup> dans le cadre d'un rapport de force entre l'individu rejeté et la société.

Ces dix dernières années, la notion d'inclusion s'est substituée à celle d'intégration ou de diversité<sup>70</sup> dans l'espace public et médiatique, révélant ainsi un changement de paradigme dans le rapport à l'autre et à la société. En effet, alors que l'intégration s'inscrit dans une dimension socio-économique - il s'agit principalement de permettre l'insertion sociale et professionnelle d'une partie de la population, ou schématiquement de faire entrer l'individu en marge dans le cercle de la société - le principe d'inclusion va plus loin et suppose un effort de la société afin que la place de chacun y soit garantie de droit, autrement dit, pour que le cercle s'agrandisse de lui-même. Parallèlement à cela, la notion de diversité repose sur la reconnaissance des différences entre les individus par rapport à un modèle dominant tandis que la notion d'inclusion dépasse ce cadre en remettant ce modèle en question et en cherchant le(s) moyen(s) de la supprimer. Dans une situation idéale, le principe d'inclusion ouvre donc la voie à un vivre-ensemble basé sur la singularité, la participation et la promotion de la production du commun.

Malgré tout, la notion d'inclusion fait l'objet d'une double

69. Pascal, B. *Pensées*, Paris, A. Pougin, 1835.

70. Assaad, I. (2019). *Pourquoi vaut-il mieux parler d'inclusivité plutôt que d'intégration ou de diversité ?*, Medium. Disponible sur : <https://medium.com>.





conception<sup>71</sup> dans la recherche en sciences sociales. L'ensemble des éléments esquissés précédemment s'inscrit dans une première approche qui considère les processus de participation, d'émancipation, de reconnaissance sociale des individus et des sociétés. Une seconde approche envisage cette notion par le questionnement de l'autorité, la violence symbolique, la domination et les injonctions de participation. Dans ce cadre, l'inclusion se matérialise par une perpétuation de l'ordre établi<sup>72</sup> par des modes d'inscription des rapports de domination dans le psychisme ou dans le corps et semble interroger plus largement le degré de liberté de l'individu, sa faculté à poser un regard critique sur l'organisation de la société ainsi que son pouvoir de décision quant à la façon dont il veut mener son existence.

Récemment, un groupe de travail, regroupant plusieurs étudiantes de l'Université Paris 1 Panthéon Sorbonne, a mené une étude<sup>39</sup> sur la possibilité d'une intelligence artificielle inclusive dans le cadre d'une mission réalisée en partenariat avec les écoles Webforce3. L'enjeu de ce projet était de déterminer si la notion de diversité peut permettre de lutter contre les biais algorithmiques discriminatoires - ce que l'on peut traduire par : éviter de créer de l'exclusion grâce à une IA régulée et, ainsi, "respectueuse" - et, plus globalement, si le développement d'une certaine forme d'IA dans le monde du travail peut contribuer à inclure dans l'emploi des personnes en marge du système en France. Dès l'initiation du projet, l'ambivalence de la notion d'inclusion a été relevée et une approche de philosophie sociale techno-critique a été adoptée pour la dépasser. Ainsi, les auteures envisagent leurs recherches à partir du postulat suivant : "À condition de ne pas penser l'inclusion comme subordination à une technologie totalisante, l'IA est un outil qui peut être solidaire d'un idéal de justice sociale". En concluant sur leurs travaux, qu'elles jugent elles-mêmes comme une amorce de réflexion philosophique appliquée à l'IA inclusive, elles soulignent le fait que leur enquête de terrain auprès des acteurs et utilisateurs de l'intelligence artificielle a fait émerger deux acceptions distinctes de la notion d'inclusion. La première consiste à prendre en compte, dès la conception, les personnes marginalisées et leurs problématiques, dans le but de favoriser leur accès à l'emploi. Ces dernières ont ainsi l'opportunité de devenir des protagonistes de l'innovation et d'enrichir l'intelligence collective d'un groupe de travail de par leur expérience de vie différente de celle considérée comme la norme. La seconde consiste à considérer une production juste des systèmes d'IA pour laquelle l'ensemble des tâches est réparti équitablement tout au long de la chaîne de production parmi toutes les parties prenantes dans une logique de coopération et d'interdépendance de profils au sein d'un collectif de travail.

71. Wolff, V. (2019). Éditorial. L'inclusion : entre injonction et participation ? Pensée plurielle, 49 (1), 7-9. Disponible sur : <https://www.cairn.info>.

72. Genel, K. L'inclusion sociale, entre autorité, reconnaissance et justification dans l'École de Francfort et la sociologie (de la) critique, dans [Reconnaissance, identité et intégration sociale, Nanterre, Presses universitaires de Paris Nanterre, 2009]. Disponible sur : <https://books.openedition.org>.

## Opérationnaliser la notion d'inclusion

# 4.2

Dans le contexte actuel de développement de l'intelligence artificielle, les initiatives se multiplient pour se doter de moyens réels qui permettraient de favoriser l'émergence de systèmes basés sur une technologie d'IA se voulant à la fois inclusive et de confiance. Loin d'offrir un panorama exhaustif, la suite de ce paragraphe regroupe un ensemble d'éléments opérationnels déployés ou envisagés pour atteindre cet objectif. Ces éléments et ceux cités dans les paragraphes précédents sont centralisés dans la [Figure 1](#).

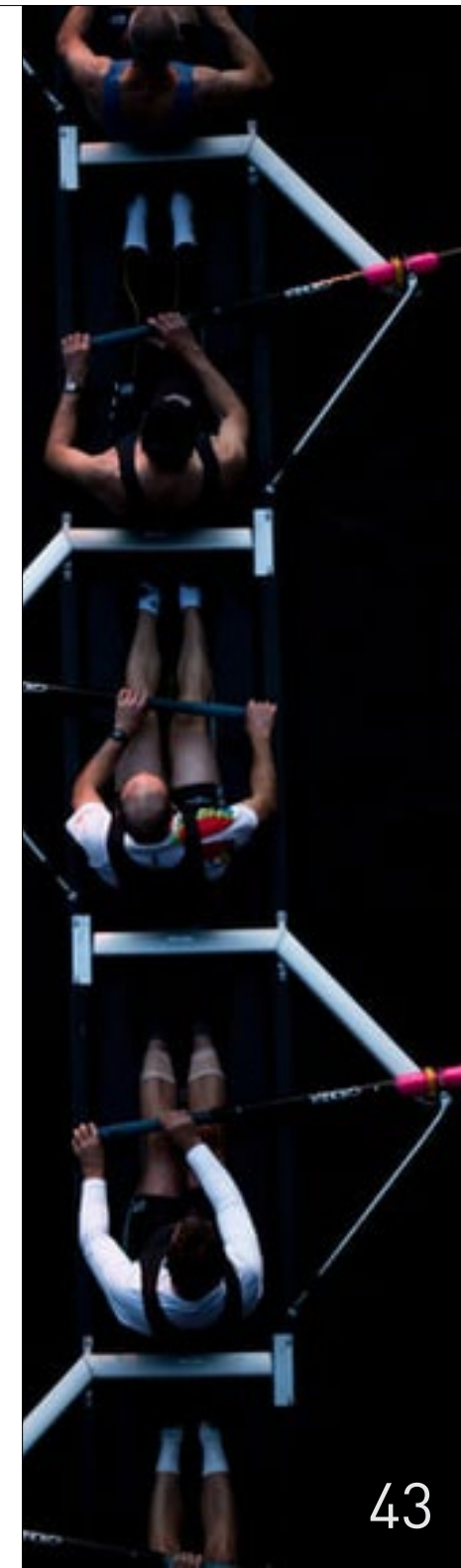
Du point de vue des éléments juridiques et institutionnels pouvant servir l'objectif d'une IA inclusive et de confiance, différentes propositions et initiatives plus ou moins récentes viennent s'ajouter à celles présentées dans le [paragraphe 3](#). Parmi elles, on note notamment :

- La création par Arborus et Orange d'une charte internationale<sup>73</sup> pour une IA inclusive, aujourd'hui signée par une centaine d'entreprises à travers le monde. Ce type de dispositif permet aux entreprises de formaliser leurs engagements en parallèle d'un cadre juridique en construction ;
- La proposition d'une réflexion sur les politiques d'éthique des entreprises<sup>74</sup> et leur formalisation en se basant sur un cadre réglementaire commun.

À ces éléments, il est nécessaire d'articuler des éléments qui vont permettre de promouvoir la collaboration au sein d'un collectif de travail. Tout d'abord, un nouveau paradigme du dialogue social. En référence au projet SéCoIA Deal (voir

73. <https://charteia.arborus.org/>

74. Blons, E. (2019). L'IA peut-elle améliorer le dialogue social ?, Forbes. Disponible sur : <https://www.forbes.fr>.



[paragraphe 3.2.3.](#)) et dans la continuité des guides produits par Perspectives IA ([voir paragraphe 3.2.2.](#)), il semble nécessaire d'encourager les directions des entreprises à mettre autour de la même table l'ensemble des parties prenantes avec les concepteurs et développeurs d'IA avant même que les systèmes soient pensés et créés. Pour que le dialogue puisse se faire et être réellement constructif, plusieurs éléments de réflexion doivent être travaillés :

- Comment réussir à créer une base de connaissances et un langage commun entre des profils d'acteurs différents, aux sensibilités, compétences, métiers et cultures variés ?
- Dans quelle mesure la stratégie IA d'une entreprise doit-elle être co-construite avec l'ensemble des parties prenantes ? Quelles sont les conditions de succès d'un tel projet ? Quels sont les garde-fous à mettre en place ? Quelles méthodes envisager pour générer des synergies et permettre de co-concevoir et co-développer des outils techniques intelligents par une approche interdisciplinaire ?
- Comment anticiper et penser de nouvelles trajectoires professionnelles pour les salariés dont les missions seraient impactées par le déploiement de systèmes d'IA ? Quels éléments clés pour un dialogue transparent, ouvert et constructif avec toutes les parties prenantes sur ce sujet ?
- Dans le cas d'entreprises dont le métier consiste à développer et commercialiser des systèmes d'IA, comment s'assurer que chaque solution proposée s'adapte au contexte de l'entreprise cliente et peut être étudiée et challengée par l'ensemble de ses salariés ?

Ensuite et de façon plus globale, la généralisation de l'intelligence collective comme outil de vie politique au sens premier du temps, *i.e.* en tant qu'outil de vie dans la Cité. Depuis plusieurs années, on assiste à une multiplication des initiatives citoyennes<sup>75</sup> au cours desquelles la population exprime clairement un besoin d'être écoutée, entendue et de pouvoir participer à la co-construction des politiques publiques de demain. Dans le cas de l'intelligence artificielle, plus que technologique, l'enjeu est avant tout politique et social<sup>76</sup>. En effet, réussir à promouvoir l'inclusion et la confiance dans les outils intelligents réside principalement dans le fait de parvenir à intégrer l'IA au débat public pour en faire l'objet d'une vraie discussion et lui rendre sa dimension collective. Comment initier un débat public et citoyen sur l'IA ? Quels dispositifs et quelles méthodologies doit-on donc imaginer pour co-construire des systèmes techniques complexes en intégrant la dimension citoyenne et collective ? Cette question s'ajoute à la liste des réflexions à mener pour repenser le dialogue social, mais cette fois-ci, à l'échelle de la société.

75. Jeffredo, A. (2021). L'intelligence collective au cœur des enjeux politiques et sociaux, The Conversation. Disponible sur : <https://theconversation.com>.

76. Usbek et Rica (2021). Humaine ou artificielle, il n'y a pas d'intelligence sans collectif. Disponible sur : <https://usbeketrica.com>.

L'ensemble des éléments précédents est fortement corrélé à la notion d'éducation et de formation. Sur la même lignée que les initiatives nationales et internationales présentées dans le [paragraphe 3.](#), François Taddei<sup>77</sup> milite pour la création d'un service public de l'apprendre qui parviendrait à articuler plusieurs dispositifs répondant aux besoins de formation initiale, formation tout au long de la vie (dont formation professionnelle, petite enfance, troisième et quatrième âges) et développement personnel qui soient centrés sur les individus ainsi que leur capacité à apprendre et proposent des alternatives aux services commerciaux actuels. Appliqué au développement et au déploiement massif des systèmes d'IA et des algorithmes, ce concept pourrait permettre à chaque individu d'être en capacité de comprendre les évolutions technologiques en cours au fil de leur apparition, de savoir ce qui se cache derrière l'algorithme de tel ou tel service et de l'utiliser en toute connaissance de cause, de s'approprier et d'utiliser ses données personnelles, de comprendre l'environnement et la société dans lesquels il évolue...

Parallèlement à cela, certains chercheurs estiment que la pensée informatique<sup>3</sup>, *i. e.* la capacité d'utiliser les processus informatiques pour résoudre des problèmes dans n'importe quel domaine, représente un levier permettant de s'ancrer dans une démarche critique et créative à l'aide des concepts et processus du monde numérique. Ce type d'éducation semble répondre aux besoins de développement d'un esprit critique et de capacités créatives face à l'émergence de l'IA.

Dans une approche plus systémique, d'autres chercheurs, encore, invitent à "cultiver nos IA [en nous donnant] peu à peu les moyens de ne plus subir des IA trop fantasmagoriques pour être intelligentes, et peut-être à terme de concevoir les nôtres"<sup>78</sup>. Dans ce but, les œuvres computationnelles<sup>79</sup> seraient un moyen d'observer la réalité de l'intelligence artificielle et de réaliser à quel point elle est différente de l'intelligence humaine sur bien des aspects.

Finalement, dans le domaine de la conception et de la production des systèmes intelligents, un certain nombre d'outils et de méthodes<sup>79</sup> existent pour permettre de se diriger de plus en plus vers une IA inclusive et de confiance. Le mouvement OpenAI, initié en France par le Grand Défi "IA de Confiance"<sup>80</sup> a notamment établi une liste des grands principes de design pour une IA plus éthique, tel un test de sécurité avant le lancement, et en continu durant sa durée de vie. Dans cette dynamique, certains grands acteurs du numérique, comme IBM<sup>81</sup> ou Google<sup>82</sup> ont d'ailleurs créé des outils de vérification des algorithmes à l'usage des entreprises. L'objectif est de fournir les clés de compréhension des décisions des algorithmes via l'analyse de la chaîne de décision ainsi que des faits et données sur lesquels l'IA s'est appuyée. Suite à cette analyse, des pistes

77. François Taddei est chercheur en génétique moléculaire évolutive et médicale et en sciences de l'éducation. Responsable du département « Frontières du vivant et de l'apprendre » à l'université Paris V et soutenu par la fondation Bettencourt-Schueller, il met en place des approches éducatives innovantes tels que les Savanturiers-École de la recherche. François Taddei et Ariel Lindner co-fondent ainsi le Centre de recherches interdisciplinaires (CRI) à Paris. Il milite activement pour l'innovation dans l'éducation, et co-rédige ou pilote divers documents consultatifs pour le ministère de l'Éducation nationale français. Source : Wikipédia.

78. Celeux-Lanval, M. (2020). L'intelligence artificielle va-t-elle remplacer les artistes ?, Beaux-Arts. Disponible sur : <https://www.beauxarts.com>.

79. Perspectives IA (2021). L'intelligence artificielle : reproductrice massive de biais ? Disponible sur : <https://perspectives-ia.fr>.

80. Gouvernement français (2019). [Grand Défi] Sécuriser, certifier et fiabiliser les systèmes fondés sur l'intelligence artificielle. Disponible sur : <https://www.gouvernement.fr>.

81. <https://developer.ibm.com/technologies/artificial-intelligence/projects/ai-fairness-360/>

82. <https://pair-code.github.io/what-if-tool/>



de solutions pour atténuer les biais détectés sont proposées mais également un accès open source à une partie des recherches de ces grands acteurs pour que le plus grand nombre puisse travailler sur des solutions d'IA moins biaisées.

Le dispositif d'innovation collaborative "Transformateur numérique"<sup>83</sup>, porté par l'Agence nationale pour l'amélioration des conditions de travail (ANACT) et la Fing permet d'apporter un soutien aux projets combinant innovation technologique et innovation sociale dans le but d'intégrer la qualité de vie au travail dans les projets de transformation numérique et d'utiliser les outils et la culture numérique pour favoriser la qualité de vie au travail.

Le réseau Design justice regroupe une communauté d'individus et d'organisations mandatées pour repenser les méthodologies et processus design afin qu'ils soient désormais centrés sur les personnes le plus souvent marginalisées par ce domaine.

La méthodologie "éthique *by-design*"<sup>39</sup> est un sujet de recherche sur le cœur de l'IA qui considère, à la fois, une éthique de l'intention en amont de la conception et une éthique conséquentialiste, en aval du processus de conception. *A priori*, cela permet de penser, dans une démarche d'anticipation, les motivations, valeurs et visions du monde susceptibles d'être codées dans le dur des systèmes socio-techniques. *A posteriori*, cela permet d'aborder les biais inconscients ou les conséquences non-anticipées, afin d'établir des mécanismes de rétroaction par lesquels il est possible de les corriger. En tant que sujet de recherche, cette méthodologie est amenée à évoluer, tout comme les documents dont l'objectif est d'apporter un support à sa mise en œuvre. Néanmoins, de premiers référentiels pratiques<sup>84</sup>, qui déclinent les grands principes "éthique *by-design*" de manière opérationnelle restent tout de même disponibles.

Bien qu'incomplète, la liste des éléments présentés au sein de ce paragraphe permet d'envisager, dans le cas d'une diffusion et d'une appropriation par le plus grand nombre, que chacun ait la possibilité et la capacité de mettre les mains dans le processus de conception et de production de l'IA. Ainsi, en misant sur l'intelligence collective, chacun s'assurerait à son niveau de la production de solutions intelligentes dignes de confiance et qui répondraient aux besoins français et européens.

83. <https://www.anact.fr/services/le-dispositif-transformateur-numerique>

84. Syntec Numérique & Cigref (2018). Éthique et numérique : un référentiel pratique pour les acteurs du numérique. Disponible sur : <https://www.cigref.fr>.



005

# Conclusion

Le présent rapport s'intéresse au domaine de l'IA et apporte des éléments d'explication des notions d'intelligence (humaine et artificielle), de travail et d'emploi. Il présente les initiatives principales à l'échelle internationale et à l'échelle française pour développer une stratégie européenne et française reposant sur une approche humaniste. Le développement et le déploiement de cette stratégie s'appuie sur le domaine de recherche "Trustworthy AI" et s'inscrit, notamment, dans le cadre du Grand Défi français IA de Confiance. Enfin, il s'intéresse plus précisément aux enjeux d'inclusivité et réunit des outils, méthodes, réflexions qui pourraient venir en support à l'ensemble des parties prenantes pour concrétiser des futurs de l'IA souhaitables.



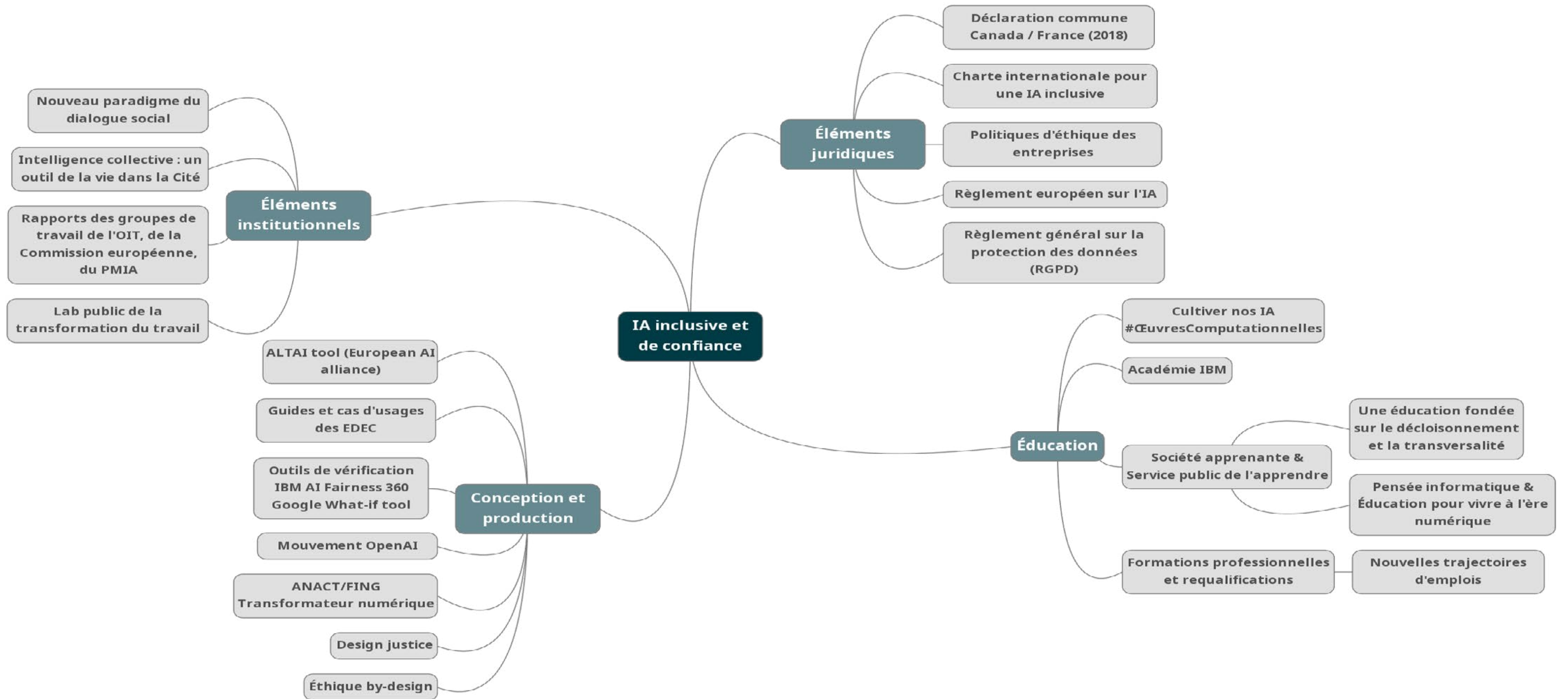


Figure 1 : Mindmap récapitulatif des principaux éléments opérationnels pouvant promouvoir une IA inclusive et de confiance.





# MATRICE

Fondé en 2016, à la fois organisme de formation, incubateur, centre de recherche, laboratoire d'innovation et lieu de création artistique, Matrice s'affranchit des lignes de partage classiques pour assembler dans une même communauté étudiant-e-s, entrepreneur-e-s, chercheur-se-s et artistes, par-delà les ancrages disciplinaires.

Au quotidien nous construisons des dispositifs qui permettent de former ou d'innover en articulant l'analyse sociale des contextes à grand renfort de sciences humaines, la réalisation de projets avec des méthodes issues de l'entrepreneuriat et une capacité de production de solutions technologiques.

Au centre de notre action est le numérique conçu comme un environnement social qui détermine nos manières de voir le monde, lire, écrire, penser et créer. En ce sens, il est l'enjeu à investiguer comme entrepreneur-se-s, chercheur-se-s, innovateur-ric-e-s ou artistes, afin de dessiner des futurs souhaitables.

## **MATRICE opère aujourd'hui cinq types de programmes :**

→ **Des programmes de formation** à l'entrepreneuriat et aux métiers du numérique à destination d'étudiant-e-s, d'entrepreneur-e-s, de professionnel-le-s, de demandeur-se-s d'emploi et de personnes en reconversion.

→ **Des programmes d'innovation** portés par MATRICE Lab, pour accompagner des entreprises et institutions publiques dans leur transformation numérique, organisationnelle, stratégique ou sociale des projets d'innovation sur-mesure.

→ **Des programmes d'incubation de startups** destinés aux entrepreneur-e-s souhaitant déployer un projet innovant, ambitieux mais aussi responsable, dans un cadre collectif.

→ **Des programmes artistiques**, sous la bannière d'Atelier B : le programme artistique et culturel de Matrice consacré aux mutations contemporaines et aux cultures numériques.

→ **Des programmes de recherche et de transfert de technologie** pour permettre à la science de sortir des laboratoires, ainsi qu'une recherche interne qui nourrit chacun de nos dispositifs.

## **Vos contacts chez Matrice :**

### **Emmanuel Ea**

Directeur général adjoint en charge des programmes

06.58.95.81.30 - eea@matrice.io

### **Lucie Poulet**

Responsable projet

06.62.71.79.57 - lpoulet@matrice.io

**www.matrice.io**

15 square Vergennes, 75015 PARIS