Contribution à la Stratégie nationale en intelligence artificielle

Mars 2017

Anticiper les impacts économiques et sociaux de l'intelligence artificielle

Rapport du groupe de travail co-piloté par

Rand Hindi, Conseil national du numérique
Lionel Janin, France Stratégie

CNNum
Conseil National du Numérique

FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

STRATÉGIE NATIONALE EN INTELLIGENCE ARTIFICIELLE

RAPPORT DU GROUPE DE TRAVAIL 3.2

ANTICIPER LES IMPACTS ÉCONOMIQUES ET SOCIAUX DE L'INTELLIGENCE ARTIFICIELLE

Co-pilotes

Rand Hindi (Conseil national du numérique)

Lionel Janin (France Stratégie)

Rapporteurs

Charly Berthet (CNNum), Julia Charrié (France Stratégie)

Anne-Charlotte Cornut (CNNum), Francois Levin (CNNum)

FRANCE STRATÉGIE

CNNum Conseil National du Numérique

MARS 2017

AVANT-PROPOS

En mars 2016, le champion coréen de jeu de go Lee Sedol subissait une cuisante défaite face au logiciel AlphaGo. Cette victoire symbolique de la machine apprenante sur l'homme résonna pour beaucoup comme le coup de cymbale signalant qu'une nouvelle étape venait d'être franchie dans l'élargissement des frontières techniques.

Depuis quelques années, l'intelligence artificielle est devenue un sujet de débat bien au-delà du seul champ du numérique, sans doute parce qu'on pressent toute l'ampleur de ses conséquences économiques et sociétales. Cette technologie de laboratoire a connu de fait un saut qualitatif – avec notamment l'apprentissage machine – qui permet son intégration dans nombre de processus qui nécessitaient hier encore une intervention humaine. La traduction, la reconnaissance de la voix ou le classement d'images et de vidéos font figure de premiers terrains d'expérimentation et d'avancée. D'autres suivront, même si la technologie n'est pas infaillible et si la vitesse de progression reste incertaine.

Pour l'heure, ces prouesses nouvelles suscitent des réactions contradictoires. Elles font resurgir le spectre d'une disparition du travail par l'automatisation. Jusqu'où la machine peut-elle remplacer l'homme ? Mais cette révolution peut-elle être synonyme de bien-être accru pour la population, pour les travailleurs ? Laisse-t-elle augurer des gains de productivité dans les entreprises et une relance de la croissance économique en berne depuis plusieurs années ?

Inquiètes ou optimistes, les interrogations sont nombreuses. D'où l'utilité de commencer sans tarder à poser les jalons d'une stratégie nationale à moyen et long terme. Il faut prendre la vague en cours, alors que les grands acteurs mondiaux, américains ou asiatiques, ont déjà déployé des moyens considérables sur le sujet.

L'intelligence artificielle a d'abord besoin d'intelligence collective. Sollicités par la secrétaire d'État chargée de l'Innovation et du Numérique et par le secrétaire d'État chargé de l'Enseignement supérieur et de la Recherche, des acteurs institutionnels, des chercheurs, des entreprises établies et des start-up, réunis au sein de différents groupes de travail, ont consacré deux mois à définir les grandes orientations de la

France en matière d'intelligence artificielle. Les travaux ont porté sur les thématiques de la recherche et de la formation, du transfert de la technologie vers les secteurs économiques ou des impacts de l'intelligence artificielle. Le présent rapport constitue ainsi une contribution à la stratégie nationale présentée dans son intégralité par le gouvernement le 21 mars 2017.

Pour croiser les points de vue et les objectifs, mais avec un souci partagé d'éclairer la décision publique, chaque groupe de travail était co-piloté par un membre d'une administration et un représentant du secteur privé. Les travaux ont pris des formes variées : auditions d'experts, ateliers participatifs avec des étudiants sur les métiers de demain, séminaires de travail sur les impacts macroéconomiques, sous-groupes indépendants pour examiner les questions d'acceptabilité sociale et les enjeux juridiques.

Trois points du rapport peuvent être mis en avant.

Le premier, c'est l'importance de la donnée. Ce point est au fondement de la transition numérique à l'œuvre depuis quelques décennies. L'intelligence artificielle ouvre cependant une nouvelle étape, dans la mesure où les données sont nécessaires à l'apprentissage des algorithmes. Du contrôle de la donnée dépend non seulement la protection de la vie privée mais aussi la puissance économique, pour ceux qui parviennent à en tirer profit.

Le deuxième est de veiller à ce que la machine demeure au service de l'humain : les dispositifs techniques doivent améliorer les capacités ou les conditions de travail, et non augmenter la dépendance ou la mécanisation du travail. Si l'humain intervient dans un processus, il doit conserver un pouvoir de décision ; et si la machine prend une décision, les éléments sur lesquels elle s'est fondée doivent être explicables. Mais parce que certaines tâches répétitives sont ou seront automatisées, il convient d'adapter les formations pour que les travailleurs sachent tirer parti de la machine au lieu d'entrer en concurrence avec elle.

Enfin, troisième point, la méthode. La nouvelle technologie suppose un diagnostic partagé et des échanges entre les parties prenantes, à tous niveaux, depuis l'individu jusqu'aux structures nationales, dans les entreprises, les branches, avec les organisations syndicales. C'est pourquoi le rapport appelle à une large concertation sur l'intelligence artificielle et sur les transformations de l'emploi qui en découlent. L'objectif est de permettre aux organisations de s'approprier le potentiel comme les limites de la technologie. Rien ne serait pire que de considérer ces transformations comme inéluctables, alors qu'elles relèvent de choix collectifs soumis à débat.

Au cœur de ces trois questions figurent les efforts à fournir en matière de formation, aussi bien dans notre système éducatif que dans les entreprises. Parce que les tâches vont continuer à se transformer, nous avons plus que jamais besoin d'une formation individuelle tout au long de la vie. Faute d'anticipation, nous pourrions nous trouver confrontés aux suppressions d'emploi ; faute de mobilisation, nous courons le risque d'un décrochage économique et stratégique.

On le voit, ce rapport n'est que la première pierre d'un édifice qui reste à construire collectivement. Avec pour objectif global de réussir la transition numérique.

Je tiens à remercier tout particulièrement Lionel Janin (France Stratégie) et Rand Hindi (fondateur et PDG de SNIPS, membre du Conseil national du numérique) qui ont piloté le groupe de travail chargé d'anticiper les impacts économiques et sociaux de l'intelligence artificielle, ainsi que Charly Berthet (CNNum), Julia Charrié (France Stratégie), Anne-Charlotte Cornut (CNNum) et Francois Levin (CNNum), qui en ont été les rapporteurs.

Michel Yahiel
Commissaire général de France Stratégie

SOMMAIRE

SYNTHÈSE **9**

INTRODUCTION **15**

Continuités et discontinuités 16

Transformations profondes, incertitudes importantes 18

Points de vigilance 21

Penser la complémentarité humain-machine 21

IMPACTS ÉCONOMIQUES ET DISTRIBUTION DE VALEURS **23**

Enjeu de compétitivité internationale 24

Des réseaux de neurones en *open source* 25

Risques de dépendance et de plateformisation 26

L'accès aux données devient un avantage comparatif 28

La France et l'Europe doivent prendre conscience de leurs atouts 29

Inégalités territoriales au niveau national 30

PENSER LA COMPLÉMENTARITÉ HUMAIN-MACHINE **33**

Organiser une gouvernance en vue d'anticiper ces transformations 34

Quelles compétences substituables ? 35

Penser une bonne complémentarité humain-machine 36

Se former dans un monde d'intelligence artificielle 37

COMPOSITION DU GROUPE DE TRAVAIL **41**

PERSONNES AUDITIONNÉES **43**

SYNTHÈSE

L'intelligence artificielle (IA) est un sujet « chaud » : elle bénéficie d'une dynamique, d'un momentum, qui cristallisent les attentions et les énergies. Il faut donc agir maintenant pour en tirer parti.

Cette attention récente est fondée sur les progrès spectaculaires obtenus par la recherche en la matière. Il en résulte un accroissement brusque et imprévisible des tâches potentiellement automatisables qui amène à s'interroger sur les activités humaines. On passe d'un scénario de polarisation, dans lequel les emplois très manuels, d'une part, et les emplois basés sur les talents, d'autre part, semblaient préservés de l'automatisation, à un scénario de transformation qui touche potentiellement beaucoup plus de monde. C'est pourquoi la thématique de la disparition d'un grand nombre d'emplois tend à monopoliser le débat et à occulter tout raisonnement.

Le but de ce rapport est de sortir d'une vision simpliste (compréhension « magique » de l'IA, focalisation sur la disparition massive d'emplois, idée que l'IA nous tombe dessus quoi que l'on fasse induisant une passivité devant le choc annoncé). Une telle analyse a jusqu'à présent empêché de mettre en avant les perspectives positives réelles (gains de productivité dans les entreprises, amélioration de la qualité des services) comme les alertes essentielles : risque de perte d'indépendance économique, absence de gouvernance des transformations, sous-estimation des impacts sur la vie quotidienne, en particulier dans le monde du travail, et nécessaire inventivité dans les accompagnements, en particulier par l'appareil de formation.

Ce document a une triple ambition.

1. Clarifier l'acception de l'intelligence artificielle et son intégration dans la production. Il s'agit principalement d'apprentissage machine – la capacité d'un logiciel informatique à reproduire avec une très bonne qualité certaines aptitudes humaines dès lors que des données d'entraînement en quantité suffisante sont disponibles. Ces dispositifs sont utilisés pour la reconnaissance de la voix, le

classement d'images, de sons, de vidéos, le traitement de texte, par exemple la traduction automatique.

2. Attirer l'attention sur l'essentiel : les données. La valeur n'est pas là où l'on croit : elle provient des données nécessaires à l'apprentissage bien plus que de l'algorithme, dont les développements se font de manière ouverte (en *open source*). En conséquence, les entreprises ou les administrations (pour la production de services publics) doivent se concentrer sur l'identification des données susceptibles d'être utilisées et sur les modalités concrètes d'exploitation (mise à disposition des données, éventuellement mutualisation, en conservant le contrôle des exploitations qui en sont faites ou en bénéficiant d'un juste retour), Par ailleurs les réflexions sur la libre circulation des données, entre espaces géographiques mais également entre acteurs économiques, doivent être approfondies.

3. Souligner qu'il s'agira de transformations du travail beaucoup plus que de suppressions d'emplois : l'aménagement de ces transformations va demander une construction de la décision aussi anticipatrice que possible, incluant les acteurs pertinents, au premier chef les professionnels concernés. Le but à atteindre est de créer de la valeur dans le travail pour tous, de donner plus de pouvoir et d'intelligence grâce à l'IA et non de mécaniser les humains. Les politiques publiques, nationales et régionales, doivent être mobilisées pour construire une vision positive de l'IA, qui nécessite des plateformes plurielles d'évaluation et de co-conception des transformations.

Pour y arriver, un outil essentiel demeure : la formation, qui ne doit pas être perçue comme un recyclage et un pis-aller. L'IA est un chantier de formation professionnelle essentiel qui peut servir de cas d'école pour les nouvelles visions de l'éducation et de la formation qui sont depuis plusieurs années débattues dans la société : importance des compétences transversales (*soft skills*), littératie numérique, à laquelle il faut adjoindre une littératie de l'IA. Une telle formation passe par une éducation informatique générale et une mobilisation de la société sur les formations initiales (entreprises, secteur de l'éducation populaire) mais aussi une inventivité dans les formations sur poste, les formations en conception, ce qui implique une forte mobilisation des entreprises elles-mêmes, de la recherche et des partenaires sociaux pour ouvrir de nouvelles voies.

Si l'on ne fait rien, que se passera-t-il ?

- au plan économique et diplomatique, il existe un risque de décrochage économique et de perte d'indépendance, si le phénomène de concentration de l'activité numérique autour de quelques grandes plateformes se prolonge ;

- au plan social : l'absence d'anticipation, la pauvreté des solutions d'accompagnement, des approches en silo et un pilotage assis sur la seule réduction des coûts auraient les effets dramatiques de destruction d'emploi que nous annoncent les cassandres.

Il faut sortir des scénarios qui tétanisent toute prospective réaliste et plurielle, en écartant les scénarios irréalistes décrivant l'automatisation de toutes les fonctions productives. À cet égard, le rapport ne se positionne pas sur la question du revenu universel, mais se concentre sur celle de la transformation des emplois, en vue d'une reconstruction de la valeur sociale des activités humaines.

Quel peut être le chemin ? Il est crucial d'organiser dès à présent un dispositif de concertation permettant d'anticiper ces impacts. Une grande plateforme numérique, alimentée par des débats dans les territoires, pourrait être envisagée avec tous les acteurs de l'emploi : Pôle emploi, l'AFPA, les acteurs régionaux, nationaux, européens, les syndicats. Elle pourrait ensuite être déclinée à différentes échelles : par branche, dans des réseaux thématiques pertinents, au niveau régional.

Pour ce qui concerne une politique inventive de formation – mobilisation de la formation initiale, des entreprises, de l'économie sociale et solidaire –, c'est un objectif qui suppose d'examiner les secteurs d'application de l'IA, en particulier d'affiner les modèles économiques pour les secteurs non industriels, comme l'éducation, la santé ou les politiques de la ville (*smart city*, en relation avec des politiques écologiques).

Le CNNum et France Stratégie soutiennent le besoin d'une telle concertation : il est dans leurs attributions respectives d'aider les politiques publiques à anticiper les impacts des transformations numériques et à en analyser les aspects combinés, économiques, juridiques et sociaux. Ce rapport vise à prendre date. Le calendrier restreint dans lequel se sont inscrits ces travaux ne permet, à l'heure actuelle, que de donner quelques éclairages.

Au-delà de cette sensibilisation, le CNNum entreprend de développer ses analyses sur la transition des emplois et d'approfondir ses propositions. Pour le CNNum, le chemin qui s'ouvre doit avoir pour but de préserver et d'augmenter la valeur du travail humain, et de concilier de façon innovante une économie compétitive et une amélioration du bien-être dans une société inclusive.

France Stratégie, pour sa part, entend continuer ses travaux d'animation du débat public sur l'IA, thème central de la révolution numérique, en donnant la priorité à l'emploi et à la dimension territoriale. En méthode, ces travaux seront menés en

nouant toutes les collaborations utiles avec les administrations, les collectivités territoriales, les partenaires sociaux et les acteurs de la société civile.

Les recommandations

	Intitulé	Description
1	Organiser une concertation pour anticiper les impacts économiques et sociaux de l'IA	La réflexion sur les impacts économiques et sociaux de l'IA ne peut se faire de manière descendante au regard de l'ampleur des transformations qu'elle est susceptible d'engager. Il s'agit de penser, à tous les niveaux, la gouvernance permettant d'organiser un dialogue et des prises de décision à même de transformer nos sociétés et nos économies pour anticiper les impacts de l'IA. Une grande plateforme (numérique et sous forme de concertation nationale) avec tous les acteurs de l'emploi, Pôle emploi, AFPA, acteurs régionaux, nationaux, européens, syndicats serait mise en place à cet effet. La mise en place de cette plateforme pourrait servir à répondre aux objectifs des recommandations 3 (penser la complémentarité humain-machine), 4 (sensibiliser à la valeur des données) et 5 (intégrer l'IA dans les entreprises).
2	Transformer la formation tout au long de la vie	Viser une transformation d'envergure de la formation tout au long de la vie, en termes de ratio temps de travail/temps de formation, de contenus et de modes de formation ainsi que de structures porteuses.
3	Penser la complémentarité humain-machine	Mesurer la substituabilité des tâches au regard d'un faisceau de critères, techniques et sociaux, et prenant en compte un contexte économique allant au-delà de la seule organisation de l'entreprise ou du secteur d'activité : – la technologie est-elle suffisamment avancée pour que cette tâche soit automatisée ? – la tâche nécessite-t-elle des capacités cognitives verticales (orientées sur une tâche très spécifique) ou horizontales ? – l'automatisation de cette tâche est-elle acceptable socialement ? – cette tâche nécessite-t-elle le recours à une intelligence émotionnelle ? – cette tâche nécessite-t-elle une intervention manuelle complexe ?

4	Sensibiliser à la valeur des données pour entraîner l'IA	Sensibiliser les organisations à la valeur des données en tant que ressources essentielles à l'entraînement des algorithmes d'intelligence artificielle.
5	Intégrer l'IA dans les entreprises	Inciter les entreprises, grandes ou petites, à développer et intégrer des briques d'intelligence artificielle, notamment en fédérant les acteurs des différents écosystèmes autour d'un réseau thématique ; celui-ci mettrait à disposition, de façon contrôlée, les données nécessaires et faciliterait l'accès à l'expérimentation (comme pour le véhicule autonome).
6	Donner accès aux données publiques	Poursuivre et accentuer l'effort en matière d'accès aux données publiques et parapubliques.
7	Faire circuler les données	Approfondir la question de la libre circulation des données, entre espaces géographiques mais également entre acteurs économiques et en définir les conditions.

INTRODUCTION

- **Historique** de l'intelligence artificielle et des ruptures récentes (*deep learning*) ; malgré des applications nouvelles à fort impact, nous sommes loin d'une IA « générale » : les logiciels actuels ne sont capables de réaliser que des tâches très spécifiques.
- Le déploiement actuel de l'IA s'inscrit **dans la continuité des transformations numériques à l'œuvre depuis des années** ; nombre des questions que celles-ci posent font déjà l'objet d'études par ailleurs (vie privée et gestion des données, éthique, acceptabilité et responsabilité, impacts écologiques tant en termes de flux et de stockage des données qu'en termes d'épuisement des ressources).
- **Les transformations à venir sont profondes, les incertitudes aussi.** Il s'agit d'une technologie au potentiel très significatif, capable d'étendre fortement le champ des possibles.
- **Tout n'est pas rose.** Il faut avoir à l'esprit certains points de vigilance : risques de perte de compétitivité, de dépendance et de renforcement de la concentration de la valeur sur certains segments de la chaîne de production.

Un certain nombre d'études ont attiré l'attention sur la destruction massive d'emplois. Sans se prononcer sur la pertinence des réflexions générales sur la redistribution de valeur, ce rapport se concentre davantage sur la **nécessaire transformation des métiers, notamment la complémentarité homme-machine**.

Reconnaissance d'images, filtre anti-spam à l'efficacité remarquable, système de recommandation de film ou de musique, traduction automatique, jeux vidéo aux personnages de plus en plus riches, effets spéciaux psychédéliques, victoire de la machine sur l'homme au jeu de Go, au poker il y a quelques semaines, agent

conversationnel sur Internet ou à domicile et demain, logiciel de conduite automatique ouvrant la voie à la voiture sans chauffeur. Telles sont les réalisations permises par les progrès récents et spectaculaires de l'intelligence artificielle (IA), notamment les dispositifs d'apprentissage machine ou d'apprentissage statistique.

Si les principes de l'intelligence artificielle remontent loin, aux années 1940 par exemple pour les réseaux de neurones, un palier significatif a été franchi il y a quelques années, sous la triple poussée de l'augmentation de la puissance de calcul (permise en particulier par les cartes graphiques développées au départ pour le jeu vidéo), de la constitution d'immenses bases de données d'images, de vidéos ou de sons qualifiés (grâce à la volonté de quelques grands acteurs et aussi aux travaux des chercheurs et aux réseaux sociaux), et des progrès en algorithmique.

L'IA a certes fait des progrès spectaculaires, comme en attestent les exemples précédemment mentionnés et popularisés régulièrement dans les médias. Pour autant, les logiciels actuels ne réalisent que des tâches très spécifiques et nécessitent le plus souvent d'immenses volumes de données d'apprentissage pour aboutir à un résultat satisfaisant. Nous sommes encore très loin de l'intelligence artificielle dite « générale » ou « forte », susceptible de prendre en compte le contexte et d'apprendre à partir de quelques exemples.

Continuités et discontinuités

Le déploiement de l'intelligence artificielle dans la société et l'économie en particulier s'inscrit plutôt dans la suite des transformations numériques à l'œuvre depuis des années. Il inclut aussi les progrès de la robotique, de la réalité virtuelle ou augmentée, et s'inscrit dans le prolongement de l'économie des données, dominée à l'échelle mondiale par les plateformes numériques américaines et asiatiques, avec un retard de la France et de l'Europe plus généralement.

L'IA soulève des questions classiques, qui ont déjà fait l'objet de nombreux travaux par ailleurs. À titre d'exemple, l'IA se nourrit de données pour faire des diagnostics personnalisés ; elle soulève donc des questions de données personnelles. L'enjeu est d'arriver à protéger la vie privée tout en tirant parti du potentiel de l'IA qui suppose d'avoir accès à ces données.

L'IA pose également des nouvelles questions d'éthique et d'acceptabilité. Il y a une responsabilité du concepteur de l'algorithme mais également un impact des données utilisées pour réaliser l'apprentissage. Enfin, l'IA pose la question de la transparence des interactions humain-machine et de l'explicabilité des décisions que prennent les

dispositifs automatiques. Refus d'un prêt bancaire, accident d'un véhicule autonome : il devra être possible de demander une explication sur la manière dont ont été prises les décisions, voire de déterminer des domaines où il est nécessaire que l'humain conserve la décision finale. C'est un sujet de recherche qui doit accompagner le déploiement de la technologie (voir encart ci-dessous, résumant l'annexe 1 « L'intelligence artificielle en quête d'acceptabilité et de confiance »[1]).

Enjeux d'acceptabilité et de confiance dans l'intelligence artificielle

Sous-groupe « Acceptabilité et confiance »

Le développement de l'IA est de nature à transformer considérablement la sphère de l'emploi et du travail, avec notamment pour conséquence la multiplication des systèmes d'IA dans l'environnement de travail, l'intensification des collaborations entre humains et machines ainsi que l'automatisation de la fonction de prédiction à travers les secteurs industriels et la délégation croissante à des machines de tâches et d'activités autrefois réalisées par les humains. Ces collaborations entre humains et machines vont revêtir toutes sortes de formes. Certains métiers pourront être entièrement remplacés par des machines autonomes (exemple : conducteurs) tout en créant des nouvelles tâches de supervision. D'autres resteront entièrement le propre de l'homme car le remplacement de l'homme par l'IA sera considéré inacceptable, inefficace ou non bénéfique pour la société. Dans d'autres cas encore, sans doute plus nombreux, des formes de coopération homme-machine verront le jour dans des environnements de travail où les IA seront des partenaires, des assistants plutôt que des remplaçants. Des agents autonomes participent déjà à différents types de négociations, de transactions (exemple : transactions boursières), d'interactions avec les clients, les usagers, les entreprises et organisations. À court terme vont être affectés de manière encore plus large les métiers du calcul et ceux qui suivent une procédure formellement établie. De nouveaux métiers vont également apparaître comme ceux de concepteur, d'entraîneur ou de gestionnaire (exemple : cogniticien) de données et d'informations, de surveillant ou encore de réparateur d'IA, etc.

Les membres du sous-groupe « Acceptabilité et confiance » ont mené une réflexion sur les concepts clés dont la signification est susceptible d'évoluer

[1] Les annexes sont téléchargeables sur le site internet de France Stratégie : www.strategie.gouv.fr/publications/anticiper-impacts-economiques-sociaux-de-lintelligence-artificielle.

sous l'effet de l'IA (responsabilité, mérite et valeur, créativité) et les risques de l'utilisation de l'IA dans certaines dimensions (affective, corporelle, mimétique, téléologique, créative) des métiers. Les effets positifs étant réels, gain de temps, prise en charge de tâches très répétitives ou des routines procédurales sans intérêt, il est néanmoins nécessaire d'envisager les risques, notamment ceux de désengagement, de déresponsabilisation, d'accroissement des disparités sociales, etc. Le résultat est un ensemble de recommandations qui concernent en particulier la nécessité d'expérimentations sur le terrain et d'études pluridisciplinaires. Un effort d'éducation et de communication sera crucial.

Pourquoi l'IA suscite-t-elle autant d'attention auprès du grand public et des médias ? S'agit-il d'une simple bulle amenée à se dégonfler ? L'expression « intelligence artificielle » possède un pouvoir d'évocation qui explique sans doute en partie les fantasmes qui y sont associés.

Transformations profondes, incertitudes importantes

Deux points à mettre en avant pour justifier l'attention particulière que mérite ce sujet de la part des pouvoirs publics, des entreprises et plus généralement de la société civile :

- d'une part, il s'agit bien d'une technologie au potentiel très significatif, capable d'étendre fortement le champ des possibles (jusqu'au véhicule sans chauffeur ou à la création automatique de scénarios, voire de films) ;
- d'autre part, il subsiste de fortes incertitudes sur son déploiement effectif et la manière dont elle sera appropriée, notamment dans le monde du travail, ainsi que sur la vitesse des progrès à venir. Le risque de destruction massive d'emplois écrase trop souvent toute considération sur l'intégration pratique de cette technologie.

Les questions du revenu universel, de la personnalité juridique des robots ou d'une éventuelle taxe sur les robots sont posées dans le débat politique[1]. Ce rapport ne se positionne pas sur leur pertinence mais précise le diagnostic, identifie des points de vigilance et formule quelques recommandations en envisageant plutôt des transformations du travail que sa disparition irréversible.

[1] Rapport Delvaux adopté par le Parlement européen le 16 février 2017, Recommandations à la Commission concernant des règles de droit civil sur la robotique.

Enjeux juridiques de l'intelligence artificielle

L'IA et les bouleversements qu'elle est susceptible d'entraîner conduisent à une réflexion sur les enjeux juridiques s'agissant notamment de la protection de l'IA ou encore de la pertinence des mécanismes de responsabilité actuels. Fort de ce constat, le sous-groupe « Enjeux juridiques et réglementaires de l'IA » a mené des réflexions et organisé des auditions sur le droit de l'IA afin de recueillir les observations des acteurs de l'IA, de la PME au groupe international, quant aux difficultés potentielles rencontrées en la matière ou au contraire, les solutions juridiques d'ores et déjà pratiquées de manière opérationnelle.

Les membres de ce sous-groupe s'accordent sur le fait qu'à ce stade de l'évolution de l'IA, il n'existe pas de nécessité impérieuse de repenser le dispositif législatif et règlementaire en vigueur. Les mécanismes et régimes juridiques actuels couplés à la souplesse contractuelle permettent aujourd'hui de faire face aux évolutions technologiques en apportant, sauf cas particuliers, un niveau de sécurité juridique satisfaisant. En outre, il ressort des réflexions menées que, comme sur d'autres sujets d'innovation de pointe, le droit ne doit pas devenir un frein à l'innovation. Aussi, la France et surtout l'Union européenne, qui apparaît dans tous les cas le niveau d'intervention adéquat pour éviter tout désavantage compétitif, doivent laisser le temps aux technologies et aux modèles économiques liés à l'IA de parvenir à une certaine maturité avant d'entreprendre le cas échéant un travail de complétion législative.

Le sous-groupe a réfléchi spécifiquement aux enjeux liés aux algorithmes, aux données, à la responsabilité des robots et à la protection des créations issues d'une IA. Ses réflexions sur ces thématiques sont annexées au présent rapport (voir annexe 2).

L'IA permet des gains de productivité au sein des organisations (filtre anti-spam, traitement automatique de dossier, réponse à des questions simples par un agent conversationnel automatisé[1]), des améliorations en termes de qualité de traitement

[1] Logiciel capable d'interagir avec l'homme en langage naturel (par écrit ou par la voix) pour répondre à des questions simples.

(amélioration de la détection de tumeur[1]), mais elle n'échappe pas aux interrogations existantes sur les effets du numérique sur le bien-être ou sur la croissance économique (voir le débat entre les tenants d'un ralentissement durable des gains de productivité, la stagnation séculaire – *The Rise and Fall of American Growth*, Robert Gordon, 2016 – et les tenants d'une croissance à venir – *The Second Machine Age*, Erik Brynjolfsson et Andrew McAfee, 2014). Pourquoi observe-t-on un ralentissement de la croissance de la productivité ? S'agit-il d'un manque d'investissement dans les technologies de l'information et de la communication (TIC), d'une insuffisante appropriation par les organisations, d'un déficit de formation ? Le numérique se contenterait-il de remplacer des processus existants et manquerait-il de nouveaux produits, comparables à ceux qui ont assuré la croissance économique lors des révolutions industrielles antérieures (avion, automobile) ? Sommes-nous dans une situation de croissance lente, durable ou s'agit-il d'un effet temporaire ? Le débat reste ouvert, au-delà de la question de l'IA[2].

Ce qui est sûr, c'est que l'IA est une technologie générique (*general purpose technology* en anglais) : ses objets de travail (image, son, vidéo, texte, langage naturel) peuvent se retrouver dans de multiples activités, ouvrir des opportunités de création de nouveaux produits et accroître les performances des utilisateurs. Les techniques d'apprentissage peuvent être diffusées d'un secteur à un autre, moyennant des adaptations limitées. Investir dans l'IA est donc susceptible d'avoir des effets d'entraînement dans de nombreux secteurs, à condition de réaliser des investissements complémentaires à cette technologie pour en tirer tout le potentiel. Ces investissements complémentaires sont à la fois matériels, en équipement numérique par exemple, ou immatériels, notamment en données, en logiciel, en savoir-faire et compétence ou en organisation. Les modèles économiques reposant sur le concept de technologie générique montrent en outre que les effets de substitution de la technologie au travail (les destructions d'emplois) peuvent être plus que compensés par l'amélioration de la compétitivité de certains secteurs, qui augmente leur potentiel à l'exportation (voir l'exemple de l'Allemagne, très robotisée mais avec plus de salariés dans l'industrie que la France), par les gains de pouvoir d'achat, qui stimulent la consommation, et par l'innovation de produits, qui créent de nouveaux marchés, et donc de nouveaux emplois.

[1] Lors du *Camelyon Grand Challenge for Metastatic Cancer Detection* qui s'est tenu en 2016, les meilleurs résultats ont été obtenus par les équipes faisant coopérer l'humain et l'IA : 0,5 % de taux d'erreur pour l'homme assisté de l'IA, 7,5 % pour l'IA seule, 3,5 % pour l'homme seul.

[2] Voir la conférence organisée par la Banque de France et France Stratégie sur ce sujet le 1er février 2017, « Productivité : une énigme française ? ».

Points de vigilance

Pour autant, ce scénario rose doit bien entendu être nuancé. L'IA est l'une des pièces du débat sur le lien entre robotisation, automatisation et emplois, qui fait naître l'inquiétude d'une disparition massive et rapide de ces derniers. Sur ce sujet, les études, dont une revue a été faite notamment par le rapport publié en janvier 2017 par le Conseil d'orientation pour l'emploi (COE)[1], rappellent que les emplois qui sont automatisables sont en nombre limité (une dizaine de pourcents, que ce soit dans l'étude de l'OCDE[2], de France Stratégie[3] ou les travaux du COE). En outre, la possibilité technique d'automatiser des emplois ne veut pas dire qu'ils le seront, puisque ce résultat dépend de l'incitation économique à le faire et de choix d'organisation et sociaux. Les mêmes études établissent cependant que nombre de tâches sont susceptibles d'être transformées (près de la moitié). C'est sans doute la bonne lecture de l'étude d'Oxford (Frey et Osborne)[4] qui a contribué à l'attention portée au sujet.

Par ailleurs, l'IA est largement développée et commercialisée par les grandes plateformes du numérique (américaines et asiatiques), ce qui repose la question du retard français et européen dans le numérique et du risque de prolonger la dépendance actuelle. C'est parce qu'il existe un danger de se trouver à nouveau en retard dans la maîtrise et le déploiement de ces technologies que la stratégie nationale de l'intelligence artificielle est particulièrement bienvenue. Elle devra cependant *in fine* nourrir une stratégie européenne sur la question pour être totalement pertinente, comme pour nombre de questions numériques (fiscalité, données personnelles, etc.).

Face à une telle transformation, l'enjeu de l'IA est donc celui de la diffusion de la technologie dans les processus de production et de la transformation des emplois.

Penser la complémentarité humain-machine

Une bonne intégration de l'IA passe par l'identification des apports de cette technologie pour améliorer une organisation de travail, offrir de nouveaux services ou créer de nouveaux emplois, en assurant la mise en place d'une complémentarité

---

[1] Conseil d'orientation pour l'emploi (2017), *Automatisation, numérisation et emploi*, 10 janvier.

[2] Arntz M., Gregory T. et Zierahn U. (2016), « The risk of automation for jobs in OECD Countries: A comparative analysis », *OECD Social, Employment and Migration Working Papers*, n° 189, juin.

[3] Le Ru N. (2016), « L'effet de l'automatisation sur l'emploi : ce qu'on sait et ce qu'on ignore », *La Note d'analyse*, n° 49, France Stratégie, juillet.

[4] Frey C. B. et Osborne M. A. (2017), « The future of employment: How susceptible are jobs to computerisation? », *Technological Forecasting and Social Change*, vol. 114, janvier, p. 254-280.

entre l'homme et la machine. L'IA peut supprimer des tâches répétitives, pour permettre de se concentrer sur les plus intéressantes ; elle peut « augmenter » les travailleurs, en leur fournissant un outil de diagnostic qui enrichit leur expérience (réalité augmentée, pré-diagnostic en matière médical améliorant l'orientation des patients). Mais de tels outils comportent aussi le risque de transformer le travailleur en un simple exécutant déqualifié, suivant les prescriptions de la machine. Ces possibilités dépendent des conditions de déploiement et d'accompagnement de la technologie. Pour éviter l'aliénation et la prolétarisation, il ne faut pas mettre en place de dispositif automatique exerçant un contrôle complet sur l'humain (sinon, autant l'automatiser jusqu'au bout) : l'humain doit garder la capacité de reprendre la main.

IMPACTS ÉCONOMIQUES ET DISTRIBUTION DE VALEURS

- L'intégration de l'IA au sein de nos économies est porteuse de **gains de productivité importants** et permet de créer davantage de valeur.
- Dans des économies ouvertes comme les nôtres, il y a un **fort enjeu de compétitivité** et donc une urgence à intégrer rapidement ces innovations.
- Les ruptures récentes de l'intelligence artificielle résultent de la conjugaison de trois facteurs : un accès à des données massives, une puissance de calcul importante et des algorithmes sophistiqués. La recherche et le développement des algorithmes de *machine learning* et de *deep learning* se font très largement sur un mode *open source* : les algorithmes sont disponibles. **L'enjeu se situe dans l'accès aux données et à la puissance de calcul**.
- **Risque de dépendance sur le long terme et de perte de souveraineté** : les entreprises françaises et européennes pourraient à terme n'être que des consommatrices de logiciels et de solutions développées à l'extérieur.
- Dans ce contexte de recherche ouverte, **l'accès aux données nécessaires aux algorithmes devient donc un avantage compétitif**. Les plateformes disposent d'une position favorable. Risque de concentration de la valeur.

L'IA peut conduire à renforcer les **inégalités territoriales** – prime aux métropoles – qui reprennent des logiques plus générales propres à l'économie numérique : concentration de la production dans les centres urbains, etc.

L'intégration de l'intelligence artificielle au sein de nos économies est vectrice de gains de productivité importants. En tant qu'innovation technique, elle constitue un intrant pour les entreprises à la fois dans leurs processus internes (management, logistique, service client, assistant, etc.) et dans leurs productions, qu'il s'agisse de biens de consommation (objets intelligents, véhicules autonomes, etc.) ou de services (banque, assurance, juridique, médical, etc.). Ne pas intégrer ces technologies représenterait un risque de perte de compétitivité majeur.

Trois risques paraissent néanmoins devoir être pris en compte, en termes de dépendance économique, de souveraineté et de possible renforcement des inégalités. Dans la configuration actuelle de l'économie numérique, le risque que les entreprises françaises ne soient que des consommatrices de solutions développées à l'étranger est important. Le développement de l'intelligence artificielle pourrait alors renforcer largement les effets de plateformisation et de captation de la valeur qui sont déjà caractéristiques de l'économie numérique : en effet, ce développement nécessite des moyens, en termes de données et de capacités de calcul, dont peu d'acteurs disposent. De plus, ces technologies pouvant servir d'outils d'aide à la décision, voire de prise de décision, il est nécessaire de préserver la capacité à en déterminer le sens et le contenu. Enfin, la diffusion de ces technologies peut avoir un effet important sur l'augmentation des inégalités territoriales et sociales, du fait notamment de la concentration de la valeur dans certains lieux.

Enjeu de compétitivité internationale

L'intelligence artificielle pourra apporter des gains de productivité (voir annexe 3 « L'intelligence artificielle et l'enseignement des technologies génériques ») et un surcroît de valeur à la production. Dans une économie ouverte, comme celle de la France, la production nationale se devra d'intégrer rapidement cette innovation pour rester compétitive autant en termes de prix (gains de productivité) que de qualité.

Comme d'autres innovations de l'ère numérique, l'intelligence artificielle est porteuse de changements radicaux dans les secteurs économiques. La France risque de se trouver confrontée à un conflit de transition, avec de grandes entreprises qui n'ont pas les moyens (stratégiques, logistiques, organisationnels) d'intégrer efficacement les innovations et qui utilisent des verrous institutionnels pour ralentir le développement de nouvelles entreprises plus innovantes. *In fine,* ce sont les entreprises qui ont pu se développer dans un cadre propice qui, en entrant sur le marché français, apportent les innovations et la disruption. La transition peut alors être assez brutale avec une conquête massive de la part des clients grâce à des démarches dématérialisées peu contraignantes, comme ce fut le cas avec Uber.

Recommandation : inciter les entreprises, grandes ou petites, à développer et intégrer des briques d'intelligence artificielle, notamment en fédérant les acteurs des différents écosystèmes autour d'un réseau thématique ; celui-ci mettrait à disposition, de façon contrôlée, les données nécessaires et faciliterait l'accès à l'expérimentation (comme pour le véhicule autonome).

Des réseaux de neurones en *open source*

En juin 2016, le magazine Wired prophétisait « la fin du code » en expliquant qu'à l'avenir, nous ne programmerons plus des ordinateurs : nous les « dresserons » à la manière dont nous pouvons dresser un chien[1]. De ce point de vue, l'apprentissage machine (*machine learning*) – en particulier les techniques d'apprentissage profond, basées sur des réseaux de neurones – a donné des résultats remarquables, en rupture majeure par rapport à l'existant, au cours des dernières années. Pour y parvenir, les algorithmes d'apprentissage profond nécessitent d'allier des moyens de calcul puissants à un accès à des données massives (*big data*).

La recherche et le développement des algorithmes de *machine learning* et de *deep learning* se font très largement sur un mode *open source*. Sur GitHub, une plateforme de mise à disposition de programmes, les projets en apprentissage machine figurent parmi les plus actifs[2]. Parmi ces projets, *TensorFlow*, qui regroupe les projets logiciels en apprentissage machine des équipes de Google, est rapidement devenu le répertoire le plus populaire sur la plateforme[3]. La France n'est pas en reste puisque la bibliothèque *Scikit-learn*[4], regroupant des outils d'IA développés en *open source* et également très utilisés, y a été largement développée. De la même manière, Facebook, Microsoft, IBM et Amazon ont opté pour un modèle *open source*, mettant ainsi gratuitement leurs travaux à la disposition de tous. Même Apple, d'ordinaire réputé pour son extrême discrétion, a annoncé en décembre dernier l'ouverture de ses recherches en intelligence artificielle[5]. Ces entreprises partagent régulièrement l'avancée de leur recherche avec la communauté scientifique. Leurs chercheurs[6] s'expriment également lors de conférences. À l'image de leurs

[1] Wired, juin 2016 : www.wired.com/2016/05/the-end-of-code/.
[2] https://github.com/showcases/machine-learning.
[3] http://donnemartin.com/viz/pages/2015.
[4] http://scikit-learn.org.
[5] www.theverge.com/2016/12/6/13858354/apple-publishing-ai-research-siri-self-driving-cars.
[6] Cette large ouverture de la recherche mondiale mérite néanmoins d'être relativisée : il semble qu'à l'inverse de la recherche occidentale, la recherche chinoise – aujourd'hui la plus active au monde ! – forme une communauté plus distincte. Des barrières linguistiques découle en effet une forme

homologues américains, les géants chinois du numérique ont eux aussi entrepris d'ouvrir leurs algorithmes d'intelligence artificielle : c'est le cas notamment de Baidu[1] qui a mis à disposition ses outils d'apprentissage profond en *open source*.

Risques de dépendance et de plateformisation

L'ouverture des algorithmes permet à la communauté des développeurs de progresser sur une technologie complexe qui demande une mobilisation collective. Néanmoins, seuls certains acteurs (Google, Amazon, IBM, etc.), disposant d'une capacité de calcul suffisante, sont en mesure de proposer aux entreprises des services d'intelligence artificielle – notamment hébergées dans le cloud.

Watson, l'intelligence artificielle développée par IBM, est déployé dans certaines entreprises pour assister les collaborateurs qui peuvent lui poser des questions en langage naturelle. Dans le cabinet d'avocat Baker & Hostetler, Ross, la version spécialisée en droit de Watson, est utilisée par l'équipe chargée des dossiers de faillite. Dans un autre domaine, Alexa, l'assistant personnel d'Amazon, qui se place au cœur des foyers pourrait devenir une porte d'entrée vers les clients incontournables pour de nombreux services et commerces, qui deviendraient ainsi complètement dépendants de cet intermédiaire.

Ces entreprises ont donc la capacité de se placer au cœur d'une activité et en interface entre une entreprise et ses clients. Elles peuvent ainsi reproduire le modèle des plateformes numériques comme Booking, Airbnb ou Uber qui ont su capter une part significative de la valeur d'un marché en se positionnant comme intermédiaire incontournable. Ce phénomène de plateformisation, typique du développement de l'économie numérique, induit donc des risques majeurs en termes de dépendance des acteurs économiques. L'avènement de l'intelligence artificielle déboucherait ainsi sur une concentration encore plus importante de la valeur économique – qui s'accompagnerait de conséquences sociales négatives (par exemple, moindres recettes fiscales). La dépendance de l'économie à des solutions d'intelligence artificielle développées à l'étranger peut également conduire à une perte de souveraineté, dès lors que les règles régissant les avancées technologiques échappent largement au contrôle national. Une fois les innovations établies et

d'asymétrie : les chercheurs chinois, qui parlent généralement anglais, bénéficient d'un accès aux travaux rédigés dans cette langue. À l'inverse, la communauté anglophone accède plus difficilement aux travaux issus de la recherche sinophone.

[1] www.theverge.com/2016/9/1/12725804/baidu-machine-learning-open-source-paddle.

déployées à l'étranger, il devient très difficile d'en infléchir le fonctionnement ou de déterminer d'autres règles pour les régir.

Impact industriel de l'intelligence artificielle : exemple du véhicule autonome

La médiatisation à partir de 2010 de projets de véhicule autonome portés par des entreprises du numérique comme Google, Apple, Tesla ou encore Uber a mis en lumière les avancées technologiques liées à l'intelligence artificielle dans le secteur de la production automobile. L'horizon du déploiement de véhicules autonomes se rapproche avec des premiers usages non expérimentaux prévus dès 2020 et une diffusion à grande échelle envisagée à partir de 2030.

L'équipement des véhicules en intelligence artificielle de conduite autonome pose la question du partage de la valeur avec les constructeurs automobiles. De 40 % de la valeur ajoutée d'un véhicule, l'équipement logiciel pourrait passer à 70 % d'ici dix ans. Les constructeurs automobiles qui comptent sur leur savoir-faire en termes de design et de moteur pour maintenir leur image de marque sauront-ils développer des intelligences artificielles suffisamment performantes ou devront-ils intégrer des logiciels extérieurs qui pourraient se placer en nouvel intermédiaire avec leurs clients ? Au-delà de leur capacité à développer une intelligence artificielle, les consommateurs seront-ils plus enclins à acheter un véhicule équipé d'un logiciel du constructeur ou de celui d'une entreprise numérique ?

En fonction du rapport de force qui s'installera, c'est-à-dire de la marque qui s'imposera comme celle du véhicule, le partage de la valeur se fera en faveur soit du constructeur automobile soit du fournisseur d'intelligence artificielle.

Au-delà de cette bataille dans les véhicules, c'est à une remise en cause de toute l'organisation économique des transports routiers qu'il faut se préparer : impact sur les assurances, le permis de conduire, les garagistes et les sociétés de transport, transition potentielle d'un système où les ménages sont propriétaires de leur véhicule vers un système de « location de véhicule à la carte », complémentarité entre les transports en commun actuels et les véhicules autonomes[1]

[1] Janin L., Nemri M. et Raynard C. (2016), « La voiture sans chauffeur, bientôt une réalité », *La Note d'analyse*, n° 47, France stratégie, avril.

L'accès aux données devient un avantage comparatif

Dans le contexte d'une recherche et d'un développement de l'intelligence artificielle largement ouverts, la question posée est moins celle de l'accès aux algorithmes que celle de l'accès aux données pour les entraîner. Celles-ci sont alors un élément crucial du développement de ces machines auto-apprenantes : sans une masse suffisante de données, les algorithmes d'apprentissage profond ne pourraient pas atteindre des taux d'erreur satisfaisants. En matière de *deep learning*, plus une machine est entraînée, plus elle est pertinente. Un algorithme de médiocre qualité pourrait ainsi être supérieur à un bon algorithme s'il est mieux entraîné.

Au sein des organisations, entreprises et administrations, les enjeux et l'importance des mégadonnées (*big data*) semblent de plus en plus compris et assimilés. Ce qui l'est beaucoup moins, c'est l'importance cruciale des données brutes en matière d'intelligence artificielle. Prises en grand nombre, ces données, même non raffinées, peuvent servir à entraîner des machines à détecter des corrélations pour, à moyen et long termes, prendre des décisions sur la base de celles-ci. De fait, les organisations qui ont accès à de larges bases de données sont en possession d'un actif crucial.

Recommandation : sensibiliser les organisations à la valeur des données en tant que ressources essentielles à l'entraînement des algorithmes d'intelligence artificielle.

Il est indéniable que les géants du numérique – qu'ils soient chinois, russes ou américains – dont le succès s'explique largement par leur capacité à collecter les données, notamment grâce à une taille de marché importante, partent avec un certain avantage dans la course mondiale à l'intelligence artificielle. Cela peut accroître les risques de dépendance économique et les inégalités en termes de répartition du pouvoir, de manière générale, et de la valeur économique, en particulier. Les géants du numérique ont alors tout intérêt à ouvrir les algorithmes de *machine learning*, pour capitaliser sur les logiques de l'innovation ouverte, tout en gardant la main sur les actifs qui en constituent la clé. Si ces grandes entreprises nouent aujourd'hui de nombreux partenariats avec des organismes de recherche pour alimenter en données les algorithmes et les améliorer, elles se gardent bien d'ouvrir toutes leurs données. Elles conservent ainsi la main sur leur actif stratégique dont elles déterminent les conditions d'accès éventuel et les modalités d'utilisation.

Conséquences du développement de l'IA sur le marché de l'externalisation

L'intégration de l'IA dans les processus de production conduit à modifier les chaînes de valeur. Différents effets peuvent être relevés. À l'échelle internationale, il est possible que des processus qui avaient été externalisés vers des pays à bas coût de main-d'œuvre soient à nouveau internalisés, dès qu'ils peuvent être réalisés par une IA. Ce processus ne met pas en danger l'emploi dans les pays développés, puisqu'il était déjà externalisé. En revanche, il peut constituer une menace pour certains pays qui se fondent uniquement sur une main-d'œuvre peu qualifiée spécialisée dans des tâches cognitives répétitives. Ce processus est clairement à l'œuvre pour certaines activités bancaires ou dans les centres d'appel.

Un autre processus est également à l'œuvre avec le développement des plateformes de micro-tâches, telles qu'Amazon Mechanical Turk ou Foule Factory. Sur ces plateformes, les développeurs d'intelligence artificielle ont recours à des travailleurs pour entraîner les algorithmes, *via* des tâches de labellisation, de nettoyage de bases de données, de reconnaissance d'images, etc.

La France et l'Europe doivent prendre conscience de leurs atouts

La question de l'accès aux données risque de se faire encore plus pressante à mesure de l'émergence d'une intelligence artificielle de moins en moins cantonnée à des tâches spécifiques (ce que l'on désigne par intelligence artificielle faible), i.e. une intelligence artificielle capable d'apprendre de nouvelles tâches ou d'acquérir de nouvelles capacités par transfert, voire par « capillarité ». Par exemple, DQN, un algorithme développé par DeepMind (Google) parvient à comprendre par lui-même les règles de différents jeux auxquels il n'a jamais été entraîné à partir de son entraînement sur un jeu spécifique. Si l'émergence d'une véritable intelligence semi-forte[1] n'est pas attendue avant plusieurs décennies, les capacités de l'intelligence artificielle actuelle tendent donc bien à s'horizontaliser et, de ce point de vue, les

[1] On distingue l'intelligence artificielle faible, capable de reproduire une tâche spécifique (celle qui est déployée aujourd'hui), l'IA semi-forte, capable de réaliser des tâches plus générales en comprenant notamment le contexte (un sujet de recherche) et l'IA forte, qui égale, voire dépasse, l'intelligence humaine (et relève toujours, à ce jour, de la science-fiction).

grandes plateformes bénéficient d'une position favorable. La diversification de leur activité, le dynamisme des acquisitions et des partenariats qu'elles concluent conduisent les plateformes à amasser des informations dans de nombreux secteurs, ce qui est propice à l'horizontalité, c'est-à-dire à la capacité d'effectuer différentes tâches. La richesse d'une donnée tient en effet pour beaucoup à son potentiel de croisement avec d'autres.

Dans cette compétition mondiale, en France et en Europe les atouts ne manquent pas. En France, par exemple, les bases médico-administratives de santé sont alimentées chaque année par 1,2 milliard de feuilles de soin, 500 millions d'actes médicaux et 11 millions d'hospitalisations. D'autres données, comme les données juridiques, pourraient gagner à être ouvertes de manière plus large.

Recommandation : poursuivre et accentuer l'effort en matière d'accès aux données publiques et parapubliques.

Néanmoins, il ne s'agit pas de se focaliser uniquement sur les données publiques ou parapubliques, mais également, à plus long terme, d'engager une réflexion sur les modes de partage de la valeur entre acteurs privés. En ce sens, le CNNum s'est saisi de la question de la libre circulation des données aux niveaux international et européen, d'un point de vue de transfert entre les territoires mais également entre acteurs économiques.

Recommandation : approfondir la question de la libre circulation des données, entre espaces géographiques mais également entre acteurs économiques et en définir les conditions.

Inégalités territoriales au niveau national

À l'échelle du territoire national, l'IA pose les mêmes questions de prime aux métropoles et d'inégalités territoriales que celles posées par le développement économique des dernières années[1]. La numérisation fait courir le risque que les activités de production se concentrent dans des centres urbains de taille suffisante pour bénéficier des externalités de réseau, notamment pour la production d'IA car celle-ci suppose de réunir des facteurs de production localisés principalement dans

[1] Lainé F. (2017), « Dynamique de l'emploi et des métiers : quelle fracture territoriale ? », *La Note d'analyse* n° 53, février, France stratégie.

les métropoles (capital humain, R & D, capital physique, écosystème). En revanche, en ce qui concerne l'utilisation des services permis par une IA, les gains escomptés peuvent être répartis sur tout le territoire. Pour ce faire, ses services doivent être diffusés dans les territoires, en veillant à ce qu'ils soient accessibles au plus grand nombre.

PENSER LA COMPLÉMENTARITÉ HUMAIN-MACHINE

- Réfléchir en termes de tâches plutôt que de métiers permet de se pencher sur la transformation plutôt que sur la destruction d'emplois.
- Afin d'appréhender ces transformations, il est urgent de définir la gouvernance permettant d'anticiper ces impacts, notamment sur trois points clés :
  - mesurer la destruction comme la transformation des emplois, et donc évaluer les compétences substituables au regard d'un faisceau de critères :
    - *la technologie est-elle suffisamment avancée pour que cette tâche soit automatisée ?*
    - *la tâche nécessite-t-elle des capacités cognitives verticales (orientées sur une tâche très spécifique) ou horizontales ?*
    - *l'automatisation de cette tâche est-elle acceptable socialement ?*
    - *cette tâche nécessite-t-elle le recours à une intelligence émotionnelle ?*
    - *cette tâche nécessite-t-elle une intervention manuelle complexe ?*
  - déterminer les compétences qui demeurent à l'avantage de l'humain peut ne pas suffire, il est nécessaire de réfléchir à la complémentarité humain-machine au niveau de l'individu et de l'organisation.
  - À partir de la détermination des compétences non substituables, **une réforme en profondeur de la formation tout au long de la vie** doit être effectuée, en termes de temporalité, de contenu et de structures.

Le développement de l'IA aura également un impact sur les conditions de travail des individus et la structure des organisations. Les études macroéconomiques se penchent surtout sur la question de la substituabilité de la machine par rapport au travail humain. Il serait évidemment hors de propos de balayer d'un revers de main l'hypothèse d'une diminution irréversible du nombre d'emplois. Pour autant, et même si cela n'obère pas des réflexions parallèles sur le mode de redistribution de la valeur, il nous semble nécessaire de réfléchir à la transformation des métiers plutôt que d'acter cette diminution irréversible. En effet, cette hypothèse raisonne à tâches constantes. Or, comme nous l'apprend le rapport du COE qui prend une approche par tâches et non par métiers, un certain nombre de tâches vont être automatisées tandis que d'autres ne le seront pas, celles où réside l'avantage comparatif de l'être humain. En effet, selon le COE sur la base de cette approche par tâches, la moitié des emplois existants pourrait voir son contenu notablement ou profondément transformé et moins de 10 % des emplois cumulent des vulnérabilités qui pourraient en menacer l'existence. Cela signifie que l'on peut partir du constat que les métiers seront en grande partie transformés plutôt que détruits. Ce constat peut s'appuyer sur les enseignements issus des précédentes révolutions industrielles, qui ont vu – parfois non sans une transition difficile – de nouveaux emplois apparaître en remplacement de ceux qui étaient détruits. De plus, l'efficacité d'un algorithme associé à un humain est bien plus importante que celle d'un algorithme seul : l'IA prend en charge les tâches répétitives, et l'humain peut se concentrer sur les tâches les plus riches. Pour ces deux raisons, il semble nécessaire de réfléchir à la notion de complémentarité homme-machine. Comment s'organiser pour mener ces réflexions de manière collective ?

Organiser une gouvernance en vue d'anticiper ces transformations

La réflexion sur les impacts économiques et sociaux de l'IA ne peut se faire de manière descendante au regard de l'ampleur des transformations qu'elle est susceptible d'engager. Il s'agit, à tous les niveaux, de définir la gouvernance qui structure un dialogue et des prises de décision nécessaires pour anticiper les impacts de l'IA, et de réfléchir à la répartition des gains de productivité permis par le numérique.

À titre d'exemple, cinq échelons pourraient être dans un premier temps ciblés : le niveau national, le niveau régional, le bassin d'emploi, la branche, l'entreprise. À chacun de ces niveaux, un travail préalable d'identification du réseau des acteurs pertinents serait mené, pour organiser une concertation et la construction participative de projets adaptés.

Une plateforme (numérique et sous forme de concertation nationale) avec tous les acteurs de l'emploi, Pôle emploi, AFPA, acteurs régionaux, nationaux, européens, syndicats, serait mise en place à cet effet.

Trois sujets prioritaires feraient l'objet de cette concertation nationale :

- réfléchir à la complémentarité entre les compétences automatisables et celles non automatisables, afin de déterminer les compétences humaines substituables et de maîtriser la transition ;
- de manière corrélative, identifier des modes de complémentarité vertueux qui puissent être pérennes et qui ne soient pas créateurs d'aliénation supplémentaire au niveau individuel et collectif ;
- enfin, repenser de fond en comble les modes de formation tout au long de la vie, du fait de la complémentarité évoquée ci-dessus.

Quelles compétences substituables ?

De nombreuses études se sont attachées à définir les compétences humaines qui, à ce jour n'apparaissent pas substituables par l'IA (voir annexe 1 « L'intelligence artificielle en quête d'acceptabilité et de confiance »). Les compétences englobant une dimension affective et relationnelle, certaines compétences physiques de précision et surtout, les capacités d'analyse générale (flexibilité, capacité d'adaptation, capacité à résoudre les problèmes, créativité), semblent *a priori* demeurer accomplies de manière plus efficace par l'homme.

Plus précisément, il est urgent de mesurer la substituabilité des tâches au regard d'un faisceau de critères, techniques *et* sociaux, et prenant en compte un contexte économique allant au-delà de la seule organisation de l'entreprise ou du secteur d'activité :

- la technologie est-elle suffisamment avancée pour que cette tâche soit automatisée ?
- la tâche nécessite-t-elle des capacités cognitives verticales (orientées sur une tâche très spécifique) ou horizontales[1] ?
- l'automatisation de cette tâche est-elle acceptable socialement ?
- cette tâche s'appuie-t-elle recours à une intelligence émotionnelle ?
- cette tâche nécessite-t-elle une intervention manuelle complexe ?

[1] Cette distinction verticale/horizontale renvoie à la distinction entre IA faible et IA forte, développée page 9

Les compétences non substituables doivent être développées (voir *infra* la partie relative à la formation). Néanmoins, ce développement, s'il est nécessaire, n'est pas suffisant pour assurer une complémentarité vertueuse entre l'homme et l'intelligence artificielle. De nombreux facteurs influent sur les modalités de cette complémentarité, au premier titre desquels l'organisation du travail. L'introduction de l'intelligence artificielle peut avoir des impacts négatifs aux niveaux individuel et collectif. Certains risques sont décrits dans l'annexe 1 « L'intelligence artificielle en quête d'acceptabilité et de confiance ».

Penser une bonne complémentarité humain-machine

Définir des principes de complémentarité permettant de garantir l'acceptabilité et l'appropriation de l'IA, une diffusion pérenne de ses apports, une protection contre les risques systémiques qu'elle peut contenir en germe, s'impose aussi.

Par ailleurs, au niveau individuel, le travail quotidien avec l'IA au-delà de ses nombreux atouts peut également présenter des risques sociaux et psychiques. La concentration des capacités d'analyse et d'intelligence dans les nouveaux outils d'IA renferme le risque de mener à une « prolétarisation »[1] d'un nombre encore plus grand de travailleurs, au sens de la privation de leurs savoirs qui valorisent leurs compétences. La délégation d'un nombre toujours plus important de tâches à l'IA est également susceptible d'engendrer un désengagement, une déresponsabilisation des travailleurs, du fait de l'appauvrissement des relations interpersonnelles ainsi que d'une délégation accrue de la prise de décision.

Au niveau des organisations, la diffusion des technologies liées à l'IA devra être appréhendée en termes de risques cognitifs, notamment d'accroissement potentiel de la propension à la « stupidité fonctionnelle ». Une introduction massive de l'IA dans les organisations peut favoriser l'obéissance, la conformité et la prévisibilité des comportements. En effet, si la diffusion de l'IA entraîne l'automatisation d'un nombre encore plus élevé de tâches et d'activités répétitives, elle peut dans le même temps contribuer à la délégation de la prise de décision et favoriser des contextes organisationnels diminuant la mobilisation des capacités cognitives. Elle pourrait ainsi favoriser le manque de réflexivité des individus et de l'organisation ainsi que l'absence de justification des décisions. En optimisant toutes les tâches, elle réduirait, voire annihilerait les capacités à innover. De telles situations mèneraient alors au

[1] Pour reprendre la notion développée par Bernard Stiegler dans son ouvrage *La Société automatique*, Fayard, 2015.

développement de risques systémiques, qui menaceraient, au-delà de l'organisation, les structures économiques et sociales.

Cas pratique : l'exemple du centre d'appel

Le cas du métier d'opérateur de centre d'appel est intéressant car il montre l'ambivalence de l'automatisation au sein des organisations.

a. Scénario où le travailleur est « augmenté » par l'IA

- Automatisation des appels les plus simples : 80 % des appels sont automatisés.
- Concentration de l'opérateur sur les 20 % d'appels restants, les plus complexes et qui permettent de développer des compétences commerciales et relationnelles.
- Sur les 20 % d'appels restants, l'opérateur est aidé par l'IA : il peut utiliser des moteurs de recherche sémantiques permettant de retrouver très rapidement l'information et, à nouveau, de se concentrer sur la relation client.

b. Scénario où le travailleur est subordonné à l'IA

- Une intelligence artificielle analyse les appels clients et propose un script d'interactions à l'opérateur, qui se contente de lire à haute voix les réponses proposées par la machine. Ces réponses sont optimisées au regard de l'analyse de l'ensemble des autres interactions passées.
- L'opérateur devient donc un support vocal et émotionnel en réponse quasi automatique à l'intelligence artificielle.

Se former dans un monde d'intelligence artificielle

Au regard de l'objectif d'une complémentarité vertueuse entre l'IA et le travailleur, la formation joue un rôle incontournable : en effet, développer des compétences transversales de créativité, d'adaptabilité et de capacité à résoudre des problèmes rend nécessaire une réforme importante de la formation tout au long de la vie, même si celle-ci n'est pas uniquement liée aux modifications du marché du travail induites par l'intelligence artificielle. La formation professionnelle fait régulièrement l'objet de nombre de critiques concernant ses modes de fonctionnement et son efficacité. De plus, l'exigence d'un décloisonnement entre formation initiale et formation continue, au regard de la difficulté d'anticiper les activités et les métiers de demain, a déjà été

défendue dans un certain nombre de travaux, notamment ceux du CNNum sur la transformation numérique de l'enseignement supérieur et la recherche[1]. Néanmoins, les bouleversements massifs que va connaître le marché du travail du fait du développement de l'IA pose la question de la formation tout au long de la vie avec une acuité nouvelle. **En effet, en accélérant l'obsolescence des compétences acquises lors de la formation initiale, l'intelligence artificielle pose la question de l'accroissement du temps de formation continue et de la modification de ses contenus et de ses structures.** Le lien entre le fait d'avoir suivi une formation initiale et celui d'exercer le métier correspondant toute sa vie est clairement à remettre en cause : d'où l'objectif d'une formation tout au long de la vie qui puisse véritablement accompagner les transitions, à l'intérieur d'un même métier mais également vers les métiers et les secteurs d'avenir.

Or le ratio actuel entre temps de travail et temps de formation pourrait ne pas être à la hauteur de l'exigence de renouvellement constant des compétences. C'est pourquoi il faut réfléchir à l'extension des temps de formation continue, pour donner à chacun la possibilité d'acquérir les nouvelles capacités techniques et non techniques dont auront besoin les individus dans un monde d'intelligence artificielle. Cette extension des temps de formation continue devra aller de pair avec la mise en place d'une continuité plus importante entre formation initiale et formation continue. **Les professions potentiellement les plus touchées par l'automatisation pourraient être la cible prioritaire d'une politique de formation préventive.**

La modification des contenus de formation s'impose ensuite. Le défi de la formation initiale et continue consiste à trouver l'équilibre entre l'acquisition de savoirs et compétences en phase avec les besoins actuels de la société et la nécessité de préparer un avenir dont on sait au moins qu'il exige de tous un esprit critique, une capacité d'apprendre à apprendre. Le développement de l'intelligence artificielle renforce cet état de fait. Il nécessite à la fois de permettre l'acquisition des compétences techniques qui lui sont spécifiques afin que les besoins en main-d'œuvre du secteur puissent être satisfaits, mais également de développer les qualités générales d'adaptabilité, de créativité et d'esprit critique. À terme, ces compétences peuvent seules permettre aux individus de développer leur complémentarité avec la machine. Cela signifie qu'outre les formations spécifiques à l'intelligence artificielle, la question de la réorientation de la formation continue vers des formations généralistes, permettant de développer des compétences transversales d'adaptabilité et de

[1] Voir travaux du CNNum sur la transformation numérique de l'enseignement supérieur et de la recherche : Référentiel de transformation numérique de l'Enseignement supérieur et de la Recherche, 19 décembre 2016.

créativité, devra être posée. Le développement de ces compétences nécessite de réinventer les modes de formation (formation en contexte, sur poste, en conception) et le rapport au savoir lui-même, autour de pratiques de partage, de mise en réseau, d'expérimentation, de projets, etc.

La transformation des structures de la formation tout au long de la vie est enfin inévitable. Au regard de la nécessité accrue, pour chaque travailleur, de pouvoir se former selon ces exigences, les entités existantes ne sont ni suffisantes ni adaptées. Le rôle de l'Université dans la délivrance de la formation tout au long de la vie pourrait être ainsi davantage affirmé, pour assurer à tous une formation de qualité. Pôle emploi évoluerait pour s'inscrire dans une logique préventive. Enfin, les dispositifs de formation informels (tiers-lieux, *fablabs*, *makerspaces*, etc.) et d'expérimentation de pratiques pédagogiques nouvelles doivent être intégrés pleinement à l'évolution des structures de formation tout au long de la vie.

Recommandation : viser une transformation d'envergure de la formation tout au long de la vie, en termes de ratio temps de travail/temps de formation, de contenus et de modes de formation ainsi que de structures porteuses.

COMPOSITION DU GROUPE DE TRAVAIL

Co-pilotes

Rand Hindi, président de Snips, membre du CNNum

Lionel Janin, adjoint au directeur du Département développement durable et numérique, France Stratégie

Rapporteurs

Charly Berthet, Secrétariat général du CNNum

Julia Charrié, cheffe de projet, France Stratégie

Anne-Charlotte Cornut, Secrétariat général du CNNum

François Levin, Secrétariat général du CNNum

Membres

Gilles Babinet, Serial entrepreneur, EU Digital Champion

Alain Bensoussan, fondateur de Lexing Alain Bensoussan Avocats

Yann Bonnet, secrétaire général du CNNum

Antonio Casilli, sociologue spécialiste des réseaux sociaux, maître de conférences en humanités numériques à Télécom ParisTech

Jacques Crémer, professeur d'économie à la Toulouse School of Economics, membre du CNNum

Jean-Michel Dalle, directeur d'Agoranov

Laurence Devillers, professeure à l'université Paris-Sorbonne et chercheuse au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi) du CNRS *(présidente du sous-groupe « Acceptabilité et confiance »)*

Paul Duan, fondateur et président de Bayes Impact

Sylviane Gastaldo, directrice de programme Évaluation des investissements publics au CGI

Ghislaine Hierso, présidente de l'Association française des Petits Débrouillards, membre du CNNum

Boris Le Hir, chef de projet, France Stratégie

Françoise Mercadal-Delasalles, directrice des ressources et de l'innovation du groupe Société Générale, membre du CNNum

Laurent Michel, directeur adjoint du programme « Économie numérique », CGI

Pierre Nougué, fondateur d'Ecosys Groupe et Reporters d'espoirs

Marion Panfili, chargée de mission « Économie du numérique », DGE

Antoine Petit, PDG d'Inria, professeur d'informatique à l'ENS Cachan, membre du CNNum

Sophie Pène, membre du CNNum

Gregory Renard, Chief AI Officer, CTO, TEDx & Stanford Speaker, Deep Learning Faculty & AI Community Activist

Jean-Charles Samuelian, PDG d'Alan.eu

Olivier Savrimoutoo, France Stratégie

Eric Scherer, directeur de la prospective, France Télévision

Marie Soulez, cabinet Lexing - Alain Bensoussan

Bernard Stiegler, philosophe, directeur de l'Institut de recherche et d'innovation

Célia Zolynski, professeur de droit à l'université de Versailles-Saint-Quentin-en-Yvelines, membre du CNNum

i

PERSONNES AUDITIONNÉES

Laurent Alexandre, co-fondateur de Doctissimo

Gilles Babinet, Digital Champion

Daniel Benoilid, Foule Factory

Alain Bensoussan, cabinet Lexing Alain Bensoussan Avocats

Raphaëlle Bertholon, CFE-CGC

Pierre Blanc, Athling

Alexandre Cadain, Hyperloop

Marie-Claire Carrère-Gée, Conseil d'orientation pour l'emploi

Igor Carron, LightOn

Antonio Casilli, Telecom ParisTech

Corentin Chéron, Airinov

Mady Delvaux, Parlement européen

Romain Drosne, RefundMyTicket

Paul Duan, Bayes Impact

Diane Dufoix, IBM

Franck Escoubes, Bluenove

Bernard Georges, Société générale

Yannig Goude, EDF

Dominique Guellec, OCDE

Hubert Guillaud, InternetActu

Olivier Guillaume, O^2 Quant

Éric Humphrey, Spotify

Tim Hwang, Google

Frédéric Kaplan, École polytechnique fédérale de Lausanne

Eric Keller, FO

Gisèle Legionnet, School of Creative Leadership

Rose Luckin, UCL Knowledge Lab

Gérard Mardiné, CFE-CGC

Alexis Massé, CFDT

Jean-Luc Molins, UGICT-CGT

Louis Monier, Qwiki/Kiron

Dennis Mortensen, X.ai

Yann Moulier-Boutang, UTC Compiègne

Pierre Nougue, Ecosys

Grégory Renard, X Brain

Jean-Charles Samuelian, Alan

Éric Scherer, France TV

Bernard Stiegler, Institut de recherche et d'innovation

Daniel Susskind, université d'Oxford

Joseph Thouvenel, CFTC

Joëlle Toledano, CentraleSupélec

Dominique Turcq, Boostzone

Jérôme Vivenza, CGT

Michel Volle, Institut de l'iconomie

Clément Wolf, Google

Directeur de la publication
Michel Yahiel, commissaire général
Directeur de la rédaction
Fabrice Lenglart, commissaire général adjoint
Secrétaire de rédaction
Sylvie Chasseloup
Contact presse
Jean-Michel Roullé, directeur du service Édition/Communication/Événements
01 42 75 61 37, jean-michel.roulle@strategie.gouv.fr

RETROUVEZ
LES DERNIÈRES ACTUALITÉS
DE FRANCE STRATÉGIE SUR :

www.strategie.gouv.fr

francestrategie

@Strategie_Gouv

Ce rapport est issu des travaux communs de France Stratégie et du CNNum. Les opinions exprimées engagent leurs auteurs et n'ont pas vocation à refléter la position du gouvernement.

FRANCE STRATÉGIE

France Stratégie, laboratoire d'idées public, a pour mission d'éclairer les choix collectifs. Son action repose sur quatre métiers : évaluer les politiques publiques ; anticiper les mutations à venir dans les domaines économiques, sociétaux ou techniques ; débattre avec les experts et les acteurs français et internationaux ; proposer des recommandations aux pouvoirs publics nationaux, territoriaux et européens. Pour enrichir ses analyses et affiner ses propositions France Stratégie s'attache à dialoguer avec les partenaires sociaux et la société civile. France Stratégie mise sur la transversalité en animant un réseau de huit organismes aux compétences spécialisées.

FRANCE STRATÉGIE - 18, RUE DE MARTIGNAC - 75700 PARIS SP 07 - TÉL. 01 42 75 60 00

STRATÉGIE NATIONALE EN INTELLIGENCE ARTIFICIELLE

L'INTELLIGENCE ARTIFICIELLE EN QUÊTE D'ACCEPTABILITÉ ET DE CONFIANCE

Contribution au groupe de travail 3.2

« Anticiper les impacts économiques et sociaux de l'intelligence artificielle »

Auteurs :

Laurence Devillers, Sorbonne/CNRS, CERNA

Danièle Bourcier, CNRS, CERNA

Stéphane Buffat, médecin, IRBA

Alexei Grinbaum, CEA, CERNA

Arnaud Lavit d'Hautefort, cabinet Ytae

Christophe Lazaro, université de Namur (Belgique), CERNA

Mars 2017

Sommaire

I. Champ d’application 4
II. Impacts éthiques et sociaux 6
III. Recommandations 9

Les opinions exprimées dans ce document engagent leurs auteurs et n’ont pas vocation à refléter la position du gouvernement.

De manière générale, il faut souligner le caractère ambivalent de toute technologie qui offre de nouvelles opportunités en même temps qu'elle crée des risques. Il ne faut pas négliger les aspects éthiques du développement de ces technologies. Les systèmes d'intelligence artificielle deviennent aujourd'hui capables d'accomplir certaines activités qui auparavant étaient l'apanage exclusif des humains. Le développement de ces systèmes conduit au renforcement progressif de leurs facultés d'autonomie propre et de cognition – la capacité à apprendre par l'expérience et à prendre des décisions de manière indépendante –, lesquelles sont susceptibles de faire de ces systèmes des agents à part entière pouvant interagir avec leurs opérateurs et leur environnement et les influencer de manière significative. Ce développement technologique est de nature à transformer considérablement la sphère de l'emploi et du travail, au point d'engendrer une nouvelle phase de révolution informatique dans l'industrie et dans les services. Celle-ci aura notamment pour conséquence la multiplication des systèmes apprenants dans l'environnement de travail, l'intensification des collaborations entre humains et machines ainsi que la délégation croissante de tâches et d'activités autrefois réalisées par les humains à des machines. Ces collaborations entre humains et machines vont revêtir toutes sortes de formes et seront différentes des collaborations classiques.

Cette conjoncture exige qu'on étudie les effets potentiels de l'IA en termes d'éthique et d'acceptabilité sociale. À ce titre, il est nécessaire de soulever toute une série de questions et de s'interroger sur la formation à l'IA, la transformation du travail grâce à l'IA, l'acceptabilité des mutations et la régulation de l'IA. Voici une liste de questions que la société doit se poser avant de décider du pourcentage d'IA susceptible d'accroître la rentabilité. En effet chaque personne doit rester le plus possible utile à la société par son travail ou ses compétences :

- **Comment l'IA va-t-elle modifier la conception de l'emploi et du travail ?**
- **Que faut-il préserver des valeurs du travail humain ?**
- **Les humains vont-ils accepter de travailler avec des IA ou sous les ordres de systèmes d'IA ?**
- **Serons-nous toujours ceux qui supervisent/surveillent ?**
- **Comment évaluer la responsabilité d'un agent humain ou numérique dans un monde où l'humain n'aura désormais plus le monopole de la décision et de l'action?**
- **Comment créer les conditions d'une coordination des activités et d'une collaboration saines entre l'homme et les systèmes d'IA ?**

Cette annexe au rapport *Anticiper les impacts économiques et sociaux de l'intelligence artificielle* entend proposer un certain nombre de pistes de réflexion destinées à construire à court et moyen terme une relation de confiance entre les travailleurs, les organisations et les systèmes d'IA. Elle a également pour objectif de formuler diverses recommandations visant à garantir le respect des valeurs humaines aux stades de la conception, de l'acceptation, de l'appropriation et de l'utilisation de ces systèmes.

I. Champ d'application

a) Les métiers bouleversés par l'IA

L'IA va bouleverser de nombreux métiers, comme l'ont fait les machines-outils lors des révolutions industrielles précédentes. D'innombrables tâches, actions et fonctions pourront être réalisées et supervisées par une IA. Des agents autonomes participent déjà à différents types de négociations, de transactions (exemple : transactions boursières), d'interactions avec les clients, les usagers, les entreprises et organisations. À court terme vont être affectés de manière encore plus large les métiers du calcul et ceux qui suivent une procédure formellement établie. Certains métiers pourront être entièrement remplacés par des machines autonomes (exemple : conducteurs) tout en créant des nouvelles tâches de supervision. D'autres resteront entièrement le propre de l'homme car l'IA sera considérée comme inacceptable, inefficace ou non bénéfique pour la société. Dans d'autres cas encore, des formes inédites de coopération homme-machine verront le jour dans des environnements de travail où les IA seront des partenaires, des assistants plutôt que des remplaçants. De nouveaux métiers vont également apparaître comme ceux de concepteur, d'entraîneur ou de gestionnaire (cogniticien) de données et d'informations, surveillant ou encore réparateur d'IA, etc. **La valeur du travail ne sera plus le *faire* mais le contrôle, la vérification, l'approbation et la validation.**

Exemple : *on peut imaginer des phénomènes de retour à des métiers/tâches avec une réalisation partiellement manuelle et des résultats apparents. Vincent de Gaulejac : « Cette tendance peut s'expliquer par deux éléments. C'est un double changement, à la fois vers le manuel : de l'abstrait vers le concret. Et il y a cette volonté de posséder à nouveau ce que l'on fait. Les cadres ont de plus en plus l'impression d'être pris dans un univers où leur travail ne fait plus sens. »* Paru sur le site internet Atlantico.

La rémunération reviendra aux concepteurs d'un procédé ou aux entraîneurs, aux vérificateurs et non pas à celui qui l'exécute car c'est la machine qui exécutera. La signature et la responsabilité finales pourraient en revanche rester humaines, c'est un choix. **Les systèmes d'IA ne prendront pas la place des humains dans toutes leurs tâches et fonctions. Il sera également nécessaire de créer de nouveaux métiers d'entraides sociales où l'humain sera au centre.**

Cette annexe met l'accent sur les métiers de service et vise, en particulier, à appréhender l'impact des IA sur certains secteurs de l'économie qui semblent plus « protégés » que les autres, mais qui ne sont pas pour autant à l'abri de changements significatifs.

Ces secteurs concernent les métiers revêtant une dimension :
(1) *affective* (exemple : soins infirmiers, assistance sociale, psychothérapie) ;
(2) *corporelle* (exemple : médecine, physionomie, analyse kinésique) ;
(3) *mimétique* (exemple : activités sportives, éducation) ;
(4) *téléologique* (exemple : management, justice, prêtrise) ;
(5) *créative* (exemple : activités artistiques et de recherche).

Ces différentes dimensions peuvent bien sûr se retrouver au sein d’une même activité ou profession.

b) Les phases de transformation du travail

Il existe deux échelles temporelles de transformation du travail :

1. Une lente évolution des métiers dans le contexte et l’environnement existants : les choix des utilisateurs portent sur « l’ancien » ou le « nouveau », mais l’ancien reste une norme parfaitement acceptée et disponible.

2. Une intégration totale des fonctions par le nouvel environnement de la vie humaine (ville intelligente, maison intelligente, objets connectés). Les contextes « anciens », sans disparaître complètement, deviennent archaïques ou exceptionnels (exemple : la maison haussmannienne non adaptée aux exigences technologiques modernes, le cheval remplacé par l’automobile). L’environnement « nouveau » s’impose en vertu des choix par défaut faits par la majorité, des décisions du secteur marchand ou des gouvernements. Ces options seront de toute façon stimulées par l’offre du marché. Dans la deuxième phase, les métiers subissent un changement de rupture, tandis que, dans la première, les compétences évoluent graduellement. Le passage de la première étape à la seconde n’est pas automatique, mais dépend de la réaction de la société, des changements générationnels et du niveau d’éducation.

L’acceptabilité des technologies apparaît dans ce processus comme un va-et-vient permanent entre les innovateurs et les divers types d’utilisateurs des nouvelles technologies.

- **Notion d’acceptabilité :** acceptabilité sociale (avant la mise sur le marché, tient compte de la pluralité des usages et perceptions) : essayer à travers des méthodes qualitatives (par exemple des scénarios) de cerner les questions d’usage que se pose le public : focus group, feedback à l’ingénieur.

Remarque : *quel choix va réellement s’offrir à l’individu : un choix social, d’éducation et de formation ?*

- **Notion d’appropriation :** sur le plan individuel, on peut définir l’appropriation comme un processus reposant sur l’acquisition progressive d’une certaine maîtrise technique et cognitive des objets techniques ainsi que sur l’intégration significative de ces objets dans la pratique quotidienne.

Remarque : *il y a deux niveaux d’appropriation celui de l’interface de médiation humain-IA et celui de la compréhension fine des mécanismes sous-jacents. Cela peut être intéressant en cas de scénario « d’accident » pour décider de la responsabilité.*

II. Impacts éthiques et sociaux

Nous discuterons les concepts et les idées dont la signification est susceptible d'évoluer sous l'effet de l'IA (responsabilité, mérite et valeur, créativité), puis les effets positifs (moins d'oubli, moins de perte de temps, prise en charge de tâches très répétitives ou des routines procédurales sans intérêt, etc.) ; enfin, les effets susceptibles d'être négatifs (manque d'exercice de la mémoire, etc.) si la société ne met pas de garde-fous et ne régule pas l'IA.

a) Changements de concepts

1. **Responsabilité**

L'entraîneur est celui qui réalise l'apprentissage du système d'IA en lui fournissant des données. L'entraîneur peut être le concepteur de la machine mais aussi celui qui a construit le corpus de données ou les utilisateurs conscients ou non conscients du fait que la machine apprend au cours de son utilisation. **La responsabilité sera partagée entre le concepteur, l'entraîneur et l'utilisateur selon des modalités qui restent à établir. Cela nécessite de modifier le droit pour prendre en compte ce type de partage.**

2. **Mérite et valeur**

La valeur du travail ne sera plus le *faire* mais le contrôle, la vérification, l'approbation et la signature. La rémunération reviendra au **responsable** d'un procédé et pas à celui qui l'exécute (car la machine exécutera).

3. **Créativité**

Un système d'IA convenablement programmé peut créer de l'innovation par expérimentation ou par sérendipité, mais seul l'humain donne un sens à la réalité nouvelle. Il est important de distinguer les créations des machines et les créations humaines.

b) Risques

Les aspects positifs de l'IA sont nombreux à condition de respecter un certain nombre de valeurs : une mémoire élargie, moins de perte de concentration ou d'oubli d'information pertinente, un conseil expert plus efficace, moins subjectif ou arbitraire, mieux justifié, etc., **à condition que la machine soit neutre, loyale et non discriminante**. Nous ne développons pas ici tous ces points déjà suffisamment soulignés et discutés par les acteurs de l'IA, mais nous analysons les risques éthiques. À chaque fois, il faut réfléchir aux aspects positifs et plus négatifs des situations et trouver les conditions pour transformer ces risques en atouts.

1. Désengagement : l'utilisation de l'IA et de machines autonomes peut conduire à un *désengagement* de l'humain. Remplacer le facteur humain par de l'IA pourrait conduire à une déshumanisation des pratiques et un appauvrissement des interactions sociales, lesquels constituent très souvent la raison d'être de certains métiers. C'est pour cela qu'il s'agit non pas de remplacer l'humain pour toutes les tâches sociales (comme par exemple dans la santé) mais de trouver les garde-fous pour accompagner au mieux la collaboration humain-machine. N'oublions pas que nombre de « décisions » sont actuellement prises

automatiquement sans que le destinataire en soit informé. L'IA peut restituer au moins les traces et les raisonnements activés.

Exemple : le secteur de la santé et, en particulier celui des soins apportés aux patients, aux personnes en situation de handicap et aux personnes âgées va, dans les prochaines années, être profondément transformé par le développement de technologies offrant une grande diversité d'applications allant de l'assistance au contrôle, jusqu'à l'accompagnement quotidien. La forte dimension affective au cœur de la relation de soin va évoluer et, dès lors, va modifier à la fois le travail des aides-soignants et thérapeutes ainsi que la vie des patients. Sans contester l'utilité indéniable des IA et des agents autonomes (robot portant, robot soignant ou robot coach) pour faciliter et soulager le travail des prestataires de soins, il est important de souligner que le contact humain est l'un des aspects essentiels de la relation de soin.

2. **Déresponsabilisation :** dans la continuité du point précédent, l'utilisation de l'IA et de machines autonomes peut entraîner une *déresponsabilisation* du travailleur, susceptible de ne plus devenir qu'un simple relais de la décision de la machine. Dans les cas de plus en plus nombreux de décisions prises sur la base d'un traitement automatisé de données, l'être humain pourrait avoir tendance à s'en remettre exclusivement à la proposition de la machine, en évitant d'engager sa responsabilité. Dévier de la solution préconisée par la machine entraînerait une prise de risque trop grande, susceptible de lui faire encourir d'éventuelles sanctions en cas de problème.

Exemple : le banquier qui s'en remettrait uniquement à la décision suggérée par la machine sur la base d'un profil pour accorder ou refuser un prêt à un client. Il y a également un problème de représentation des échelles, dans le trading par exemple, et la question de la numérosité à travers l'IA semble intéressante à creuser.

3. Transformation du mérite du travail : l'utilisation de l'IA et de machines autonomes peut bouleverser radicalement l'évaluation des performances des travailleurs et l'*idée de mérite*. En transformant les modalités de réalisation des tâches, les technologies feront évoluer les évaluations du mérite individuel et donc de l'équité des bénéfices induits par l'effort et des récompenses dues en vertu des talents acquis par l'individu.

Exemple : comment évaluer les performances d'un joueur d'échecs, d'un chirurgien ou d'un trader assisté d'une IA ? Qui est à la source de la performance ? Est-ce la technologie, le talent propre de l'humain ou l'association des deux ? Leur récompense (que ce soit en termes de salaire, de réputation) est-elle juste ? Est-elle méritée ?

4. « Prolétarisation » des savoirs et des savoir-faire : l'utilisation de l'IA et de machines autonomes peut contribuer à une *« prolétarisation » des savoirs et des savoir-faire*. À force de ne plus être entièrement partie prenante dans l'élaboration et l'expression d'une décision dans l'environnement de travail, les humains risquent de se voir déposséder de leur expertise en terme de *know-how,* et de perdre un ensemble de capacités et de compétences, qui non seulement peuvent être utiles à la société, mais qui contribuent aussi à alimenter le respect de soi. Ce risque pourra survenir si l'humain n'est pas associé à l'entraînement et à l'évaluation des modèles et au choix des données pour améliorer leur

performance. **Se pose la question d'un conservatoire des métiers, avant l'émergence massive des intelligences artificielles.**

Exemple : le douanier dans un aéroport auquel on ne demandera plus d'exercer ses compétences et son sens de la physionomie acquis au fil des années à travers une expérience sur le terrain et qui devra s'en remettre aux prédictions de la machine en matière de profils de dangerosité.

5. Autonomie et créativité des travailleurs : l'utilisation de l'IA et de machines autonomes peut altérer *l'autonomie et la créativité des travailleurs***.** Derrière les vertus facilitatrices de certains dispositifs peut se déployer de manière sous-jacente une normativité, prenant la forme d'un « paternalisme technologique » aux multiples visages : ainsi, les IA alertent, recommandent, rappellent à l'ordre, bloquent, interdisent, voire influencent.

Exemple : pour des raisons légitimes de sécurité ou d'organisation, certains dispositifs pourraient empêcher les êtres humains d'adopter des comportements considérés comme « sub-optimaux », de tenter certaines expériences, voire de commettre des erreurs qui souvent sont à la source de nouveautés et de découvertes.

6. **Vie privée et intimité au travail :** l'utilisation de l'IA et de machines autonomes peut nuire à la *vie privée et à l'intimité* des travailleurs en imposant un contrôle et une surveillance disproportionnés de leurs activités. Sans contester les bénéfices que peuvent procurer les IA dans la réalisation de multiples tâches, leur utilisation risque d'être mise en œuvre de manière abusive afin d'évaluer la productivité et le rendement des travailleurs et de les soumettre à un contrôle strict.

Exemple : pour les avocats dans un cabinet d'avocats ou les journalistes dans un groupe de presse et pour de nombreux autres métiers.

7. Épanouissement dans le travail : l'utilisation d'IA et de machines autonomes peut nuire à l'*épanouissement* des travailleurs sur le lieu de travail en perturbant le rythme de travail, les routines et en battant en brèche les conditions qui permettent de se réaliser dans le travail.

Exemple : le degré d'appréciation de la dangerosité d'un conducteur par un policier n'aura plus de valeur face aux contrôles automatiques. L'interruption du processus de pensée par une sollicitation non hiérarchisée de stimulation venant de plusieurs agents peut contribuer à désorganiser, voir à annihiler la capacité à prendre des décisions (*task switching*).

8. Disparités sociales affaiblies ou aggravées : l'utilisation de l'IA et de machines autonomes peut conduire à *affaiblir certaines disparités sociales (*exemple : valeur du « capital culturel »*) ou au contraire les aggraver* (les personnes non initiées perçoivent encore moins l'utilité de l'acquisition de ces connaissances à l'apparence « superflue »).

Exemple : la machine mémorise pour le travailleur, du coup celui-ci fait de moins en moins d'efforts pour mémoriser les informations nécessaires pour résoudre un problème, comprendre et surveiller ce que fait la machine.

9. Pratique des managers et des représentants du personnel : dans une société qui se numérise, les pratiques du digital tendent à redéfinir les modèles d'interaction entre les individus. En entreprise, le numérique contribue *à redéfinir la pratique des managers et des représentants du personnel***,** qui jouent un rôle majeur d'interface entre la direction et les salariés.

Exemple : DIGILABSOCIAL - Digital Management & Relations sociales
Bénédicte Tilloy, secrétaire générale et DRH de SNCF Réseau, et Jean-Paul Bouchet, secrétaire général de l'Union confédérale des cadres de la CFDT, ont pris l'initiative de lancer avec Sciences Po et le cabinet Ytae un laboratoire d'idées permettant de mieux dessiner les contours de ce que seront demain les fonctions de manager et de représentant du personnel dans une entreprise digitalisée, et de proposer des pratiques à expérimenter. Le laboratoire associe, à parts égales, une cinquantaine de managers et de représentants du personnel issus d'entreprises appartenant à des secteurs d'activité différents.

10. La « normalisation » d'une nouvelle technologique (au sens de Kuhn, où elle cesse d'être vue comme révolutionnaire) passe toujours par la normalisation (= standardisation). Or avec l'IA (comme avec les nanotechnologies), nous avons une technologie transversale qui ne se limite pas à un secteur industriel donné. D'où le dilemme suivant : soit on instaure des normes par secteur, soit on crée des « normes d'IA » pour tous les secteurs.

Exemple : la France a expérimenté la dernière approche avec les nanotechnologies et cela a très mal marché. En revanche, l'Europe fait de la normalisation par secteur avec les nanotechnologies et cela se présente plutôt comme un succès.

11. Transformation des capacités humaines : les facilités qu'apporte la machine, par exemple sur la mémorisation, vont ***transformer certaines des capacités humaines.*** Devons-nous anticiper ce besoin dans l'enseignement ? (en soulignant l'utilité d'exercices sollicitant la mémoire court terme et long terme (exemple : calcul mental).

Remarque : *« Aujourd'hui, j'oublie sans conséquence car la machine (téléphone, ordinateur) me rappelle ce que j'ai oublié. De ce fait, je n'exerce pas ma mémoire, mais mon attention aux indices de rappel… la machine aide à se souvenir mais rend dépendant » (https://lejournal.cnrs.fr/articles/le-numerique-nous-fait-il-perdre-la-memoire).*

III. Recommandations

Afin de répondre aux impacts éthiques et sociaux des systèmes issus de l'IA et de faciliter leur adéquation et leur appropriation par tous les acteurs du monde du travail, il est important de réfléchir dès maintenant aux différentes étapes et conditions pour instaurer un climat de confiance.

La confiance se décline à un triple niveau :

(1) ***la conception*** : certaines valeurs doivent être prises en compte dès la phase de conception technologique et intégrées dans celle-ci (« value by design ») ;

(2) ***l'appropriation*** : il est nécessaire de faciliter l'apprentissage par l'humain et son acquisition progressive de la maîtrise technique et cognitive des dispositifs techniques ainsi que l'intégration significative de ces dispositifs dans la pratique quotidienne ;

(3) ***l'utilisation*** : il faut informer les utilisateurs de l'origine automatique ou non de la décision, anticiper les conséquences possibles de l'adoption de ces systèmes de l'IA par les êtres humains et cerner en quoi ces systèmes vont redéfinir les modalités concrètes du travail.

Afin de garantir une transition plus aisée de la recherche à la commercialisation, nous proposons au gouvernement de promouvoir des programmes de recherche qui intègrent des mécanismes de vérification des résultats à court et moyen terme afin de comprendre les opportunités et les risques réels qui sont associés au déploiement et à la dissémination de ces technologies. La formation à l'IA et aux mutations est essentielle. La conduite du changement doit être menée en bonne intelligence avec tous les acteurs : dirigeants de société et syndicats.

Des ressources suffisantes doivent être consacrées à la résolution des défis éthiques et sociaux soulevés par le développement et les applications de ces technologies.

De manière générale, l'approche développée doit être graduelle, pragmatique et prudente. Dans cette perspective, nous formulons diverses recommandations.

1. Code d'éthique : l'éducation des ingénieurs à la réflexion éthique, par toutes les méthodes pédagogiques y compris le développement d'un code d'éthique, est nécessaire pour encadrer le développement et la production des systèmes d'IA. À moyen et long terme, cela complétera de façon réfléchie et informée le cadre légal existant à l'échelle nationale et européenne. Cette réflexion éthique des ingénieurs doit être fondée sur les principes de bienfaisance ou non-malfaisance, d'autonomie et de justice, ainsi que sur les principes et valeurs consacrés dans la Charte européenne des droits fondamentaux. **Pourquoi pas une déontologie des ingénieurs en IA ?**

Objectif 1 : « la formation » afin d'augmenter la compréhension de l'éthique par les ingénieurs.
Moyen : intégrer l'éthique dans le maquettage des formations.
Coûts : ECTS.
Évaluations : applicabilité de la formation à l'utilisation des ressources de l'éthique.

Objectif 2 : « les ressources de l'éthique » afin d'accompagner la conception de l'IA, et la réévaluation nécessaire.
Moyen : rédaction de chartes éthiques, méthodologie d'évaluation, mise en œuvre d'outils pour valider le suivi des chartes dans différents secteurs, par exemple un robot d'aide à la personne, aide au diagnostic, benchmarks sur les capacités des IA.
Coûts : ETP + plateformes/réseaux.
Évaluations : degré d'utilisation, appropriation (degré), mise à jour.

2. Éducation : avec l'arrivée de l'IA, une révolution sociétale est en marche qui nécessite de la part des dirigeants de définir un nouveau projet de société clair et apparent : faire évoluer les formations initiales et l'enseignement dans les écoles et lycées ainsi que les formations aux métiers qui vont être bouleversés.

3. Surveillance et formations : il faut surveiller de près le marché du travail, spécialement la création et la perte d'emplois et de savoir-faire dans différents champs professionnels. Il est fondamental d'identifier le plus tôt possible divers tendances et de chiffrer le taux de remplacement par l'IA dans différents secteurs de l'automatisation. **Les comités d'entreprise et les organismes de représentation des travailleurs (syndicats) devront être associés à ces études et formés à ces nouveaux enjeux.**

4. Recherches et études de terrain : il est nécessaire de **tester les dispositifs de l'IA dans des situations réelles (« real-life scenarios ») sur le long terme** (ou au moins à plusieurs moments d'utilisation dans le temps) afin d'identifier et d'évaluer les risques qu'ils peuvent engendrer au-delà de la phase d'études à court terme.

5. Implication et information : afin de promouvoir l'acceptabilité sociale des technologies et de faciliter leur appropriation, Il faut **informer les personnes sur les développements technologiques en cours dans la société et les alerter des conséquences potentielles sur leur travail**. Les usagers doivent être adéquatement représentés dans la définition des politiques publiques en matière d'IA et pris en compte dans les processus d'innovation. Les métiers et opportunités de cette révolution ne seront accessibles dans un premier temps qu'aux personnes averties et sensibilisées à ces outils. L'information doit aussi se faire auprès des élèves et étudiants afin de les attirer en amont vers ces formations.

6. Engagement : **dans certains secteurs d'activité, il est important de préserver le contact humain et la dimension affective et corporelle de l'activité**. À ce titre, il faut prévoir d'aménager des « espace-temps » où les interactions sociales sont fortes et où l'implication subjective dans l'activité demeure importante.

7. Prise de décision : il est nécessaire d'amener le travailleur à **coopérer en synergie avec la machine, de le faire participer au processus décisionnel** et de lui conférer un rôle susceptible d'engager sa responsabilité en sollicitant l'engagement personnel et la mobilisation subjective dans le processus de travail. Il faut que le travailleur puisse reprendre le contrôle de la machine (reprise en main) si nécessaire. À cette fin, il est utile de créer de nouveaux métiers tels que « compreneur » ou « interprète » des processus d'apprentissage non supervisé afin de **créer de nouveaux outils que le travailleur pourra utiliser pour connaître l'état de la machine, comprendre son comportement et la surveiller. Il doit aussi pouvoir apporter des suggestions sur des améliorations ou des corrections éventuelles.**

8. Autonomie : préserver l'autonomie et la créativité des individus, notamment en envisageant des possibilités d'improvisation et de spontanéité est une autre priorité. **Dans certains cas, le travailleur doit pouvoir désactiver les dispositifs issus de l'IA** afin d'exercer son activité en toute indépendance et liberté.

9. Intimité et vie privée : il est nécessaire de **garantir la vie privée et une sphère d’intimité dans l’environnement de travail** afin que chacun ait, quand il le souhaite, des moments de répit, propices à la concentration et à la réflexion. Le travailleur et l’utilisateur doivent être informés des données privées collectées ou traitées durant l’utilisation de la machine.

10. Co-travail avec l’IA : dans le co-travail avec une IA, la machine doit respecter les valeurs de l’humain à travers le travail. À cette fin, les travailleurs et les usagers seront formés pour leur permettre de se familiariser avec les exigences techniques de leurs secteurs et de **développer de nouvelles compétences.** Les standards d’excellence ou de productivité devront probablement être revus pour éviter une compétition perdue d’avance sur des domaines où l’humain n’est pas le plus performant.

11. Normalisation : les autorités doivent lancer une concertation en vue du développement des normes techniques et industrielles pour le déploiement de l’IA. Cela doit comporter des normes spécifiques par secteur industriel ainsi que des standards généraux portant sur les aspects transversaux de l’IA.

12. Expérimentation : création de laboratoires ou plateformes d’expérimentation des nouveaux métiers, sur l’exemple qui est cité **(**DIGILABSOCIAL).

STRATÉGIE NATIONALE EN INTELLIGENCE ARTIFICIELLE

INTELLIGENCE ARTIFICIELLE
ENJEUX JURIDIQUES

Contribution du sous-groupe 3.2.B au groupe de travail 3.2

« Anticiper les impacts économiques et sociaux de l'intelligence artificielle »

Membres du sous-groupe et auteurs :

Georgie Courtois, avocat associé, De Gaulle Fleurance & Associés / Responsable du groupe

Alexandra Bensamoun, professeur, université Rennes 1 / CERDI Paris XI / Co-responsable du groupe

Danièle Bourcier, directrice de recherche CNRS, CERSA (CNRS)

Sébastien Duplan, délégué aux relations institutionnelles, Syntec Numérique

Stéphane Larrière, directeur juridique, ATOS

Grégoire Loiseau, professeur, université Paris-I-Panthéon-Sorbonne

Jean-Sébastien Mariez, Senior counsel, De Gaulle Fleurance & Associés

Marc Mossé, vice-président en charge de la transition numérique, AFJE

Camille Para, responsable affaires publiques, TECH IN France (ex-AFDEL)

Marie Soulez, avocat, Alain Bensoussan Avocats Lexing

Danièle Bourcier, directrice de recherche, CNRS CERSA

Olivier Guilhem, directeur juridique, SoftBank Robotics

Nina Gosse, élève-avocate stagiaire, De Gaulle Fleurance & Associés

Mars 2017

Sommaire

I. Propos préliminaires 2

II. L'application limitée des régimes juridiques actuels pour protéger les algorithmes utilisés en matière d'IA 2

III. Transparence et loyauté : un préalable nécessaire au bon développement de l'IA 5

IV. Propriété des données : préserver les investissements sans bloquer l'accès aux données nécessaires au développement de l'IA 8

V. La responsabilité du fait des objets intelligents.....11

VI. Propriété intellectuelle et IA.....13

Annexe n° 1 – Synthèse des recommandations 16

Annexe n° 2 – Auditions et contributions.....18

Les opinions exprimées dans ce document engagent leurs auteurs et n'ont pas vocation à refléter la position du gouvernement.

I. Propos préliminaires

En l'absence de cadre juridique spécifique à l'intelligence artificielle (IA), l'objectif de notre groupe de travail fut d'analyser les enjeux principaux en la matière, et d'identifier d'éventuelles solutions, relevant tant du droit « dur » que du droit « souple ». Cet objectif est capital puisqu'il s'agit d'assurer un meilleur contrôle du fonctionnement de l'IA sans pour autant freiner l'innovation.

Plus précisément, notre groupe a décidé de travailler autour de quatre sujets majeurs, liés au développement de l'IA, et porteurs d'enjeux juridiques : **les algorithmes, les données, la responsabilité et la création assistée**. Pour mener à bien nos travaux, nous avons tenu plusieurs réunions afin d'échanger sur ces sujets, et avons par ailleurs mené des auditions afin de cerner au mieux les attentes des parties prenantes.

Dans le cadre de l'établissement d'une stratégie optimale en matière d'IA, nous pensons qu'il est important de se livrer à un travail préalable de définitions.

En effet, il ressort de nos travaux et des auditions menées que le terme « intelligence artificielle » est souvent utilisé de manière trop englobante, et qu'il convient de rappeler, pour une approche juridique et règlementaire rigoureuse, que cette notion couvre, au sens strict, **les technologies capables de traiter des sources hybrides et notamment des données non structurées**. Elle constitue à ce titre un troisième niveau d'innovation, qui se situe après l'*analytics* et l'*advanced analytics* (cette dernière notion couvrant notamment le *machine learning*).

Si les solutions d'intelligence artificielle ne sont pas nouvelles, elles ont connu un réel essor ces dernières années, et ce, en raison de trois facteurs principaux : le volume de données (*big data*), des algorithmes de plus en plus sophistiqués (grâce aux progrès informatiques et mathématiques), et l'expansion du *cloud*.

Les algorithmes et les données sont ainsi au cœur de l'IA : ils en permettent le fonctionnement, et leur relation va mener à son perfectionnement. De manière très simplifiée, on peut considérer qu'un algorithme initial est nourri par des données afin d'aboutir à un algorithme final enrichi. Cette représentation permet de mieux appréhender les questions de propriété et d'accessibilité aux algorithmes et données détenus par une entreprise.

II. L'application limitée des régimes juridiques actuels pour protéger les algorithmes utilisés en matière d'IA

Selon les modèles d'exploitation pratiqués, deux cas de figure se présentent. Dans un premier cas, la technologie d'IA repose, selon le choix de leur exploitant, sur des algorithmes dans le domaine public ou soumis, à titre alternatif, à un régime dit « open source ». C'est ici, la valeur ajoutée résultant de leur combinaison ou de leur éducation/apprentissage, que l'exploitant va chercher à protéger. Dans un second cas, c'est la protection des algorithmes en tant que tels qui peut être recherchée. Dans les deux cas, le recours aux droits de propriété intellectuelle (droit d'auteur, logiciel ou

brevet) n'est pas totalement satisfaisant et les exploitants semblent privilégier à l'heure actuelle la protection par le secret des affaires. Il ressort des travaux du sous-groupe que la flexibilité (les garanties de confidentialité et conditions d'accès peuvent être définies par contrat) et la sécurité juridique apportées par cette solution sont accueillies favorablement par les exploitants dans un environnement technologique en évolution constante.

Pour les entreprises qui placent l'IA au centre de leurs activités se pose la question de la définition d'une stratégie de protection et de valorisation de l'innovation. Il ne s'agit pas ici de la protection du robot ou de la machine en général, mais de la solution innovante placée en son centre, qui lui confère son « intelligence » (logiciel et/ou algorithme). À cet égard se pose la question de la protection de l'algorithme intégré à la machine. Sur ce sujet, des moyens de protection existent et sont utilisés, même si chacun présente des limites et s'ils ne constituent pas, aujourd'hui, des outils de protection totalement satisfaisants.

L'octroi d'un droit privatif au titre de la propriété intellectuelle apparaît ainsi comme une protection nécessairement indirecte.

Conformément à la théorie de l'unité de l'art, les œuvres de l'esprit sont protégeables par le **droit d'auteur**, quels qu'en soient le genre, la forme d'expression, le mérite ou la destination. En conséquence, le droit d'auteur protège les logiciels ou programmes d'ordinateur, sous réserve du respect des conditions classiques applicables en la matière. En revanche, il est établi que les idées et principes à la base d'une œuvre de l'esprit ne sont pas protégeables par le droit d'auteur. Cette exclusion couvre notamment les méthodes ainsi que les principes mathématiques.

Seule sera donc protégée la mise en forme d'un algorithme, par des documents ou un logiciel (algorithme intégré au code source), à condition toutefois de comporter un certain degré d'originalité. En matière de logiciel, cette originalité s'apprécie selon « la marque de l'apport intellectuel » de son auteur. Contrairement à l'appréciation classique de l'originalité, construite autour de la notion d'empreinte de la personnalité de l'auteur, l'originalité du logiciel est davantage fondée sur un « effort intellectuel personnalisé ». Dès lors, le logiciel intégrant un algorithme devra aller au-delà de la simple exécution. Par ailleurs, la dissociation entre protection du logiciel et de ses fonctionnalités a pour conséquence la possible réutilisation de l'algorithme par un tiers qui serait parvenu à l'extraire légalement.

De son côté, le **droit des brevets** vise précisément à protéger l'innovation, sous réserve de remplir les critères de nouveauté, d'inventivité et d'application industrielle.

Néanmoins, le code de la propriété intellectuelle exclut expressément les programmes d'ordinateurs et les méthodes mathématiques du champ de la brevetabilité, même si cette interdiction n'a pas vocation à s'appliquer à une invention dont la seule mise en œuvre requiert un programme d'ordinateur. Autrement dit, comme pour le droit d'auteur, la protection indirecte reste envisageable. Une attention particulière doit donc être portée à la rédaction du brevet, celle-ci devant revendiquer le procédé et non le programme en tant que tel. Ainsi, l'Office européen des brevets (OEB) affirme que « un brevet peut être délivré pour une invention mise en œuvre par un ordinateur qui résout un problème technique de

manière non évidente »[1]. Les décisions IBM I et II[2] sont communément citées en la matière afin d'illustrer cette frontière tracée par le droit des brevets, en affirmant qu'un programme d'ordinateur n'est pas exclu de la brevetabilité « si sa mise en œuvre sur un ordinateur produit un effet technique supplémentaire, allant au-delà des interactions physiques "normales" entre programme et ordinateur ». Ainsi, un algorithme ou un logiciel d'IA intégré dans une machine, et dont la mise en œuvre produit un effet technique supplémentaire semblent pouvoir être brevetés. Un avantage est que, contrairement au droit d'auteur, l'algorithme inclus dans la demande de brevet sera protégé au même titre que l'invention elle-même[3]. Toutefois, ce moyen de protection implique de divulguer l'algorithme au public (donc à ses concurrents), et de déposer toute modification de l'algorithme.

Les entreprises auront souvent recours au secret comme palliatif, ou complément, pour une protection efficace de leurs innovations.

La **contractualisation** reste un moyen classique et privilégié par l'entreprise pour protéger ses secrets industriels, que cela passe par un accord ou une clause de confidentialité. Toute divulgation sera ainsi sanctionnée par le droit de la responsabilité contractuelle, et ce même en l'absence de faute. L'intérêt de cette solution réside dans le principe de liberté contractuelle et la créativité offerte aux parties (définition du secret, interdiction d'exploitation, maintien de l'obligation après la cessation du contrat…).

En l'absence de contrat, la voie de la **responsabilité délictuelle** reste envisageable. Ainsi, l'**action en concurrence déloyale** permet de sanctionner des faits de **parasitisme** économique. Toutefois, il faudra être en mesure de démontrer une faute dans l'appropriation et dans la réutilisation du secret[4].

Concernant la voie pénale, il n'existe pas, aujourd'hui, d'infraction permettant de sanctionner directement la violation du secret des affaires. Toutefois, l'algorithme pourra utilement être protégé par la **Directive 2016/943 du 8 juin 2016 sur le secret des affaires** (qui doit être transposée au plus tard le 9 juin 2018). En effet, cette directive prévoit une définition large du secret des affaires, les informations devant seulement présenter les trois caractéristiques cumulatives suivantes :

- être secrètes « en ce sens que, dans leur globalité ou dans la configuration et l'assemblage exacts de leurs éléments, elles ne sont pas généralement connues des personnes appartenant aux milieux qui s'occupent normalement du genre d'informations en question, ou ne leur sont pas aisément accessibles » ;
- avoir une valeur commerciale du fait qu'elles sont secrètes ;

[1] EPO (1999), *Des brevets pour les logiciels ? Droit et pratique en Europe*, février.
[2] OEB Ch. des recours techniques IBM I T 1173/97 du 1-7-1998 ; IBM II T 935/97 du 4-2-1999.
[3] Marc Schuler et Benjamin Znaty (2015), « Quelle protection juridique pour l'algorithme ? », in *La propriété intellectuelle & la transformation numérique de l'économie*, INPI, p. 41-50.
[4] La Cour d'appel de Versailles a ainsi reconnu comme agissement parasitaire fautif le fait que « la société ayant bénéficié du travail d'analyse, des algorithmes et des codes sources du programme […] même s'il n'est pas contestable qu'elle a procédé à une réécriture complète de son propre logiciel » (Cour d'appel de Versailles, 12e chambre, section 2, Arrêt du 9 octobre 2003, Microsoft France / Synx Relief et autres).

- avoir fait l'objet, de la part de la personne qui en a le contrôle de façon licite (le « détenteur de secrets d'affaires »), de dispositions raisonnables, compte tenu des circonstances, destinées à les garder secrètes.

L'apport de la directive réside, en outre, dans les mesures de protection et de réparation : mesures conservatoires et provisoires (comme en matière de contrefaçon), procédure spéciale pour préserver la confidentialité des débats lors de la procédure, méthodes particulières de calcul des dommages-intérêts.

Ce nouveau texte doit permettre à l'Union européenne d'assurer une protection efficace du patrimoine immatériel des entreprises. La Directive ne prévoit pas une infraction pénale spécifique, la prérogative étant laissée aux États membres. La France pourrait utilement saisir cette occasion pour réprimer l'atteinte au capital immatériel de l'entreprise, dont la valeur n'est plus à démontrer aujourd'hui[5].

III. Transparence et loyauté : un préalable nécessaire au bon développement de l'IA

La confiance dans l'IA est directement liée à la transparence et à la loyauté des traitements algorithmiques. Afin que le besoin légitime de transparence ne devienne pas un frein pour l'innovation et sa valorisation, un équilibre doit être recherché. Aussi, afin de satisfaire cette exigence tant vis-à-vis du consommateur que dans une perspective concurrentielle, des initiatives de co-régulation tendent à l'adoption généralisée de bonnes pratiques. À l'issue de ses travaux, le sous-groupe préconise que la France encourage le développement d'une initiative auprès des acteurs de l'IA pour la mise en place de grands principes de référence en la matière (guides et bonnes pratiques, norme relative aux critères de transparence…), permettant d'assurer une transparence objective.

À rebours de l'idée de protection et de secret, les questions liées au mouvement de transparence relatif aux traitements opérés et à la preuve de ces traitements constituent également un enjeu juridique pour le développement de l'IA. Le sujet de la transparence est au cœur des débats actuels, en réponse, notamment, aux craintes exprimées envers l'objectivité des algorithmes susceptibles de biais, de discriminations, ou encore de risque pour la diversité de l'information sélectionnée. **Si ces craintes sont, dans une certaine mesure, légitimes, un équilibre doit être préservé entre exigences de transparence des algorithmes et innovation.**

Deux obstacles s'opposent à ces exigences. D'une part, la nature fluctuante et évolutive des algorithmes ; d'autre part, leur valeur d'un point de vue concurrentiel. Comme l'a fort

[5] La proposition de loi de MM. Bruno le Roux et Jean-Jacques Urvoas, relative à la protection du secret des affaires, déposée à l'Assemblée nationale le 16 juillet 2014, proposait notamment une telle infraction.

justement souligné le Conseil supérieur de l'audiovisuel (CSA), « *Toute tentative de réguler les algorithmes serait probablement vouée à l'échec en raison de l'évolutivité de la technologie et du caractère confidentiel et concurrentiel des développements* »[6]. En outre, il est important de noter que la transparence et l'objectivité des résultats seront fonction du volume et de la nature des données traitées par l'algorithme. Enfin, il serait contre-productif de « perdre » l'individu dans une masse d'information.

À ce stade, il convient donc surtout de définir ce qu'on entend par le terme de « transparence ». À ce sujet, il existe un relatif consensus sur l'idée de rendre transparentes pour l'utilisateur les informations relatives à l'origine des données utilisées par la machine et la finalité poursuivie. Certains principes communs, souples et équilibrés pourraient utilement être adoptés par les concepteurs d'IA. À titre d'exemple, Nicholas Diakopoulos et Sorelle Friedler, de l'Institut de recherche Data & Society, ont proposé cinq principes pour rendre les algorithmes fiables : responsabilité, explicabilité, exactitude, auditabilité et justiciabilité[7]. Tout comme en matière de données personnelles, la confiance des utilisateurs dans l'IA (aujourd'hui faible en France[8]) représente un avantage compétitif pour les entreprises. Certains groupes l'ont d'ailleurs bien compris, puisqu'ils commencent à travailler sur ces sujets. À titre d'exemple, Google, Facebook, IBM, Microsoft et Amazon viennent de créer un « Partnership on Artificial Intelligence to Benefit People and Society », en vue de définir ensemble des bonnes pratiques[9]. De son côté, le fondateur de Tesla et SpaceX, Elon Musk, a créé OpenAI[10].

Bien que n'étant pas spécifiques à l'IA, il faut noter que deux outils législatifs trouvent déjà à s'appliquer, et constituent plus largement une potentielle source d'inspiration pour le régulateur. En effet, la **loi pour une République numérique** impose une exigence de transparence aux administrations quant aux traitements algorithmiques servant à prendre des décisions individuelles (art. 4 et 6), et a renforcé l'actuel article L. 111-7 du code de la consommation pour élargir l'exigence de loyauté des plateformes et l'information des consommateurs, au-delà des seules plateformes de mise en relation (art. 49 & suiv.). En application de ce texte, tout opérateur de plateforme en ligne est tenu de délivrer au consommateur une information loyale, claire et transparente sur les conditions générales d'utilisation du service d'intermédiation qu'il propose, et « *sur les modalités de référencement, de classement et de déréférencement des contenus, des biens ou des services auxquels ce service permet d'accéder* ». Par ailleurs, les opérateurs de plateformes en ligne dont l'activité dépasse un seuil de nombre de connexions devront élaborer et

---

[6] Pour Dominique Cardon, deux choses sont essentielles et devraient être rendues publiques : savoir ce qui rentre dans la machine (quelles données ?) et quel est l'objectif d'apprentissage inculqué (cf. Compte-rendu de l'événement, organisé par la Cnil, de lancement du cycle de débats publics sur les enjeux éthiques des algorithmes, 23 janvier 2017).

[7] Nicholas Diakopoulos et Sorelle Friedler (2016), « How to Hold Algorithms Accountable », *MIT Technology Review*, 17 novembre. En matière audiovisuelle, le CSA utilise la notion de « loyauté », et propose la publication d'indices mesurant différents aspects qualitatifs des algorithmes (cf. *Le rôle des données et des algorithmes dans l'accès aux contenus*, CSA Lab, janvier 2017).

[8] Sondage CNIL/IFOP, Notoriété et attentes vis-à-vis des algorithmes, janvier 2017.

[9] www.partnershiponai.org.

[10] https://openai.com/blog/.

diffuser aux consommateurs des bonnes pratiques visant à renforcer les obligations de clarté, de transparence et de loyauté.

De son côté, le **règlement général de l'UE sur la protection des données personnelles (« RGPD »)** prévoit que le responsable de traitement doit informer la personne concernée de « *l'existence d'une prise de décision automatisée, y compris un profilage*[11] *(…) et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée* » (art. 13-2-f). À partir de cette information est offert à la personne concernée un droit d'opposition[12]. Deux exceptions sont néanmoins prévues à cette faculté : lorsque cela « *est nécessaire à la conclusion ou à l'exécution d'un contrat entre la personne concernée et un responsable du traitement* » ; ou si le consentement explicite de la personne a été recueilli[13].

On voit que, quel que soit le vocable utilisé, l'idée sous-jacente est toujours de permettre à l'individu de comprendre le fonctionnement de l'IA. On retrouve cette idée dans la récente **résolution du Parlement européen sur la robotique**[14], qui « *insiste sur le principe de transparence, à savoir qu'il devrait toujours être possible de fournir la justification rationnelle de toute décision prise avec l'aide de l'intelligence artificielle* »[15]. À cet effet, le Parlement européen propose que chaque robot soit doté d'une « boîte noire » contenant les données sur chaque opération réalisée par la machine, y compris les logiques ayant contribué à la prise de décisions.

Il est important de préciser que la question de l'éthique ne doit pas se limiter aux algorithmes mais doit aussi couvrir les conditions d'accès aux données, tant personnelles que non personnelles ou anonymisées. En effet, en matière d'IA, algorithmes et données forment un tout indissociable, à l'image d'un moteur et de son carburant. Comme l'a expliqué l'entrepreneur Rand Hindi, « *La donnée est à l'Intelligence Artificielle ce que l'uranium est au nucléaire, avec la même problématique des usages qui en sont faits (civils ou militaires)* »[16]. En ce qui concerne les données personnelles, il ne fait aucun doute que la conception et la mise en œuvre de l'IA se doivent d'être **respectueuses de la vie privée et des données personnelles**, sachant que cette dernière notion est très large, et l'anonymisation rarement

---

[11] L'article 4 du RGPD définit le profilage comme « toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ».

[12] L'article 22 dispose ainsi que « La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire ».

[13] On relèvera également l'article 39 de l'actuelle loi Informatique et Libertés selon laquelle « Toute personne physique justifiant de son identité a le droit d'interroger le responsable d'un traitement de données à caractère personnel en vue d'obtenir (…) 5° Les informations permettant de connaître et de contester la logique qui sous-tend le traitement automatisé en cas de décision prise sur le fondement de celui-ci et produisant des effets juridiques à l'égard de l'intéressé (…) ».

[14] Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique.

[15] *Ibidem*, Considérant n° 12.

[16] Compte-rendu de l'événement de lancement du cycle de débats publics sur les enjeux éthiques des algorithmes, CNIL, 23 janvier 2017.

irréversible. L'IA ne bénéficiant pas encore d'une entière autonomie, il revient aux concepteurs de garantir un niveau élevé de protection.

À cet égard, le RGPD fournit aux acteurs un cadre modernisé pour développer leurs produits et services. Les nouveaux principes qu'il consacre, notamment ceux de *privacy* et de *security by design*, ou encore d'*accountabilty*, trouvent ici toute leur pertinence, même si on ne peut ignorer qu'ils peuvent techniquement constituer un frein ou une barrière à l'entrée de certains acteurs émergents. D'autres règles, plus classiques, doivent également être respectées, comme la minimisation des données et la limitation des finalités. S'il ne semble donc pas nécessaire, à ce stade, de s'orienter vers une régulation spécifique relative à la protection des données en matière d'IA, certains concepts mériteraient davantage d'éclaircissement afin d'éviter toute source d'insécurité juridique, tant pour les entreprises que pour les individus. À ce titre, il est primordial que soit rapidement adopté un code de conduite ou un mécanisme de certification, pour démontrer le respect des obligations incombant au responsable du traitement ainsi qu'aux sous-traitants. Sur le sujet de la sécurité, la Commission européenne a indiqué qu'elle présenterait, d'ici fin 2017, une proposition de mécanismes de certification et de labélisation communs en matière d'Internet des objets (IoT) afin d'assurer la cybersécurité de ces dispositifs. Une telle intervention ne doit cependant pas freiner l'innovation. Un niveau élevé de sécurité des systèmes reste, néanmoins, indispensable à une utilisation appropriée de l'IA. Les acteurs doivent donc concevoir des produits sûrs, sécurisés et adaptés à l'utilisation à laquelle ils sont destinés. Toutefois, on peut supposer que les acteurs appliqueront un principe de proportionnalité afin d'assurer un certain équilibre sécurité/coûts/risques. À cet égard, l'harmonisation des normes techniques doit être poursuivie au niveau européen et international, en gardant à l'esprit que leur traduction technique peut s'avérer délicate dans la mise en œuvre (temps de traitement par exemple) ou tout simplement coûteuse, créant ainsi une possible barrière à l'entrée pour certains acteurs.

IV. Propriété des données : préserver les investissements sans bloquer l'accès aux données nécessaires au développement de l'IA

S'agissant des modèles économiques et techniques que le sous-groupe a eu l'occasion d'examiner dans le cadre de ses travaux, il apparaît que l'exploitant d'IA est positionné contractuellement vis-à-vis de son client en tant que prestataire de solution technique. Tant les données brutes, objets du traitement algorithmique, que les données résultant de ce traitement demeurent contractuellement la propriété du client. Dans un tel schéma, la problématique de propriété des données ne semble pas poser de difficulté particulière, pour autant que l'équilibre des intérêts soit de mise. Pour assurer cet équilibre, une veille sur les clauses contractuelles utilisées par les opérateurs économiques pourrait être mise en place afin d'identifier et de développer des bonnes pratiques en la matière.

Dans son étude sur le numérique et les droits fondamentaux, le Conseil d'État s'opposait à la reconnaissance d'un droit de propriété des individus sur leurs données, écartant par là l'introduction d'une logique patrimoniale dans la protection des droits personnels. En effet, la

protection des données personnelles, en France et en Europe, repose sur une logique de droits attachés à la personne. Pour cette raison, le Conseil soutenait davantage la consécration du droit à l'autodétermination, notion qui empreint indirectement la loi pour une République numérique et le RGPD. Ainsi, si les individus ne disposent pas d'un titre de propriété sur leurs données personnelles, celles-ci restent, en théorie, entre leurs mains et ils doivent pouvoir en disposer dans le cadre de leur droit à la portabilité[17].

Au-delà des données à caractère personnel se pose aujourd'hui avec acuité **la question de la propriété des données non personnelles ou anonymisées**. L'enjeu ici est de trouver un équilibre entre la valorisation économique de ce « nouvel or noir » et l'accessibilité des données (*open data)*, nécessaire à l'innovation.

En France, la loi pour une République numérique a ouvert les données publiques, en ce compris les données d'intérêt général (délégations de service public ou subventions, énergie… Elle n'est, en revanche, pas intervenue sur la question de l'accessibilité à d'autres types de données.

La Commission européenne s'interroge actuellement sur le cas des entreprises qui détiennent des données « brutes » (i.e. données qui n'ont pas été traitées ou modifiées depuis leur enregistrement) et anonymes, et les enjeux concurrentiel associés. Il est important de noter que la position française au niveau européen est très en soutien de cette l'initiative de la Commission en matière de *Free Flow of Data*[18]. La Commission porte, à cet égard, une attention particulière aux données collectées par des capteurs intégrés dans des machines, des outils et/ou des dispositifs (IoT par exemple), et qui sont en mesure de les partager[19]. Afin de favoriser l'innovation et d'éviter les exploitations en silos, la Commission soumet à la réflexion plusieurs options pour améliorer l'accès à ces données, dont entre autres : l'élaboration de lignes directrices, l'adoption de clauses contractuelles par défaut, la généralisation d'un contrôle des clauses abusives B-to-B, ou encore le développement de formats ouverts et d'APIs[20]. Le Parlement européen rejoint ces objectifs d'ouverture et d'accessibilité dans sa résolution sur la robotique[21].

Au-delà des lignes directrices et des clauses encouragées par les institutions, il ressort des auditions et contributions qui ont alimenté les présents travaux, que le mode de développement et de fonctionnement de l'IA, organisé en écosystèmes très évolutifs, peut avantageusement être aménagé autour de contrats. Conclus entre les différents acteurs de l'IA (plateformes, réseaux, fournisseurs d'algorithmes ou de données), les contrats permettent, au plus près de la réalité technique, de définir leurs rôles et leurs responsabilités dans le développement et dans le rendu des services créés autour de l'IA, et d'aménager les

---

[17] Voir sur ce point les recommandations du G29 sur la portabilité des données dans le cadre du RGPD : http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf.
[18] Voir sur ce point www.economie.gouv.fr/files/files/PDF/NAFlibrecirculationdesdonnees_20161110.pdf.
[19] Communication « pour créer une économie européenne fondée sur les données », 10 janvier 2017.
[20] API : *Application Programming Interface* (interface de programmation applicative).
[21] Le Parlement européen demande notamment à la Commission de « promouvoir un environnement ouvert qui passe notamment par des normes et des plateformes ouvertes, des modèles d'attribution des licences innovants et de la transparence, afin d'éviter les effets de verrouillage dans les systèmes exclusifs qui limitent l'interopérabilité » (considérant n° 9).

conséquences qui en résultent pour les utilisateurs ou les bénéficiaires finaux. Protecteur des intérêts en balance (créateurs, bénéficiaires, utilisateurs), le contrat, par la liberté et la souplesse qu'il offre, permet de s'accommoder, à juste temps, des situations d'agilité nécessaire aux démarches permanentes de recherche et développement, d'innovation et d'amélioration dans lesquelles s'inscrit l'IA enrichie en continue par le *machine learning*. Le contrat se présente donc comme une réponse pour l'appréhension de cette technologie « fugitive » que constitue l'IA[22] : s'il demeure un lieu de concordance des volontés des acteurs, il permet aussi de refléter, au-delà de leur commune intention, les intérêts des parties. À cet égard, il s'agit conformément à notre droit commun des contrats, de protéger la partie affaiblie dans le processus de négociation et de rendu des services de l'IA, afin de la prémunir contre tout abus. Mais il s'agit aussi, à la recherche d'un équilibre contractuel toujours fragile, de préserver la viabilité des modèles techniques et économiques émergents avec l'IA. Pour d'assurer cet équilibre, une veille sur les clauses contractuelles utilisées par les opérateurs économiques pourrait être mise en place afin d'identifier et de développer des bonnes pratiques en la matière.

C'est dans cet esprit que la Commission européenne compte aussi évaluer prochainement la **Directive 96/9/CE** du 11 mars 1996 concernant la protection juridique des bases de données, et interroge sur la question de la propriété des données brutes, avec en vue l'objectif d'accroître l'octroi de licences d'utilisation. À cet égard, si elle envisage un partage de la propriété (entre le propriétaire, l'utilisateur et le fabricant), elle propose aussi la reconnaissance d'un droit *sui generis* de propriété intellectuelle au fabricant ou au producteur.

En effet, aujourd'hui, seul le producteur d'une base de données bénéficie d'un droit de propriété, lorsque la constitution de la base « atteste d'un investissement financier, matériel ou humain substantiel » (art. L.341-1 du code de la propriété intellectuelle). Ce droit *sui generis* vient en supplément du droit d'auteur qu'il complète. *A contrario*, les investissements pour la création de données ne sont pas pris en compte, ni la gestion de données et les offres de services sans constitution de bases permanentes. Une grande partie des traitements de données récents et innovants (*data mining* par exemple) ne sont donc pas couverts par le régime de la loi du 1er juillet 1998. Des auteurs proposent ainsi de l'étendre en un « droit des producteurs et des exploitants de données et de bases de données »[23]. Cette extension impliquant au demeurant d'identifier « *quand et pourquoi le lien entre le producteur de bases de données et ses données pourra se rompre, au profit du curateur ou de celui qui l'emploie* », ce qui pourrait résider dans la condition première d'un effort économique (au sens large) propre. Comme le soulignait Nicolas Courtier, avocat, il s'agit de permettre à des activités de traitement et d'analyse de cohabiter[24].

---

[22] V. N. Colin et H. Verdier (2015), *L'Âge de la multitude*, Paris, Armand Colin, p. 25.

[23] Nicolas Courtier (2015), « La nécessaire évolution du droit des producteurs de bases de données pour permettre son adaptation à l'émergence du Big Data », in *La propriété intellectuelle & la transformation numérique de l'économie*, INPI, p. 23-39.

[24] Ce dernier proposait, en complément d'une évolution de la législation, que l'INPI développe un outil électronique de déclaration des bases de données afin d'éviter toute insécurité juridique.

Enfin, concernant l'exception de *text and data mining*, introduite en France par la Loi pour une République numérique, à la suite notamment du Royaume-Uni et figurant dans le projet de réforme européenne du droit d'auteur, certains acteurs craignent qu'en recentrant l'accès sur la seule recherche, la fouille numérique de textes et de données soit *de facto* exclue de leurs licences en l'absence de disposition expresse. Il semble donc nécessaire de prévoir expressément cette hypothèse contractuellement.

V. La responsabilité du fait des objets intelligents

Il résulte des travaux du sous-groupe qu'il n'y pas à l'heure actuelle de vide juridique, étant donné que les instruments du droit positif de la responsabilité peuvent être mobilisés, afin de trancher les litiges susceptibles d'émerger. Seul le recul de l'expérience pratique, au double point de vue juridique et économique, permettra, à terme, de faire des choix. Ce recul permettra de proposer, le cas échéant, une adaptation pertinente en termes de droit de la responsabilité, si les limites pressenties des régimes actuels devaient être confirmées. Des règles spéciales doivent cependant être d'ores et déjà envisagées s'agissant de certaines machines intelligentes, en particulier les voitures autonomes.

La réception par le droit du développement de la robotique et de l'IA pose au premier rang la question de la responsabilité du fait des objets intelligents. La résolution adoptée par le Parlement européen le 16 février 2017 « *contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique* » l'accompagne d'une mise en perspective que serait une responsabilité des robots eux-mêmes, ce qui oriente la réflexion vers la reconnaissance d'une **personnalité juridique** de ces machines intelligentes. Cette conception doit cependant être repoussée car elle est aussi inutile que dangereuse. En effet, la personnalité attribuée à certains objets intelligents – lesquels ?, selon quel critère ? – ne réglerait rien des questions qui se posent en termes de sécurité et de responsabilité. La sécurité peut plus sûrement compter sur des normes imposées aux concepteurs de logiciels, aux fabricants, voire aux utilisateurs de robots qu'en programmant des devoirs dont ces derniers seraient personnellement débiteurs. Quant à la responsabilité, elle se conçoit rationnellement comme une responsabilité du fait de la chose robot plutôt que comme une responsabilité du robot lui-même qui impliquerait d'abonder un patrimoine qui lui serait attaché pour l'exécution de la dette de réparation. La personnification des robots troublerait surtout gravement les catégories juridiques en donnant vie à une chimère, mi- personne mi-chose, à la fois sujet de droit et objet de droit.

A s'en tenir à une responsabilité du fait du robot ou de l'IA lorsqu'elle n'est pas intégrée dans une enveloppe physique, la détermination d'un responsable, ou du moins d'un débiteur d'indemnités par le truchement de l'assurance, peut se faire au sein d'un **groupe de responsabilités particulièrement large :** propriétaire, utilisateur, développeur de l'IA (ou sans doute plutôt éditeur), fabricant, etc. En pratique, la victime pourra attraire l'ensemble de ces intervenants, lesquels pourront le cas échéant opposer les uns aux autres des clauses limitatives ou exonératoires de responsabilité stipulées dans les contrats qui les lient. **En l'état actuel des choses, il nous semble prématuré de vouloir profiler un régime plutôt**

qu'un autre. Les choix ne pourront être faits, au double point de vue juridique et économique, qu'avec **le recul de l'expérience pratique**. Il faut, en particulier, compter sur les orientations que prendront les systèmes assurantiels, lesquelles détermineront le point d'ancrage de la couverture du risque. La résolution adoptée par le Parlement européen est sur cette ligne de temporalité, considérant que de nouveaux instruments ne pourront être mis en place qu'à un horizon de dix à quinze ans (point 51). Durant la période intermédiaire, il n'y a pas de vide juridique car **les instruments du droit positif peuvent être mobilisés**. Dans cette optique d'un traitement de la réparation par le droit actuel, la nature des risques devrait être déterminante dans le choix du mécanisme de responsabilité : si les risques de structure liés à la conception de l'objet intelligent peuvent justifier la responsabilité spéciale du fabricant du fait des produits défectueux, les risques de comportement tenant à l'utilisation de la chose entraîneraient la responsabilité de droit commun du fait des choses du propriétaire ou de l'utilisateur à l'égard de tiers. Tout au plus, la pratique révélera au besoin la nécessité d'adapter certains éléments de régime comme l'exonération pour risque de développement attachée à la responsabilité du fait des produits défectueux.

Sans attendre, des règles spéciales pourraient d'ores et déjà être adoptées en ce qui concerne certaines machines intelligentes. C'est le cas, en particulier, des véhicules autonomes, car il fait peu de doute que leur mise sur le marché suppose la confection d'une règlementation qui sécurise l'utilisation de ces nouveaux objets, et l'indemnisation des dommages qu'ils sont susceptibles de causer. Ne serait-ce que d'un point de vue d'acceptabilité sociale, la norme doit anticiper la mise en pratique. Un travail préparatoire a d'ailleurs déjà commencé à propos de leur utilisation. Par ordonnance du 3 août 2016 prise en application de la loi n° 2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte, le gouvernement a autorisé la circulation sur la voie publique de véhicules autonomes pour l'expérimentation de ces voitures sans conducteur ou disposant de systèmes d'aide à la conduite. Le texte constitue un préalable à l'élaboration d'un cadre règlementaire par décret en Conseil d'État.

Dans une vision plus générale de la responsabilité, qui ne doit pas se limiter à une question de réparation de dommages, **l'attention doit être attirée sur deux points**. Le premier est l'asymétrie des responsabilités qui se profile entre celle du fabricant de robots incarnés intégrant une IA et celle du développeur/éditeur d'IA. La responsabilité du fabricant du robot physique (la partie *hard*) intégrant une intelligence artificielle (la partie *soft* dont le fabricant n'est pas le plus souvent propriétaire) pourra être assez facilement engagée en cas de dommage, le *hard* et le *soft* formant un tout, un produit complexe. En revanche, la responsabilité de l'éditeur sera plus difficile à établir. Il ressort des auditions que les d'IA mettent contractuellement à disposition du client un simple outil, le système d'apprentissage dépendant de l'utilisateur-client qui ainsi contribue au processus décisionnel. L'autre point est la possibilité de risques systémiques. Le développement de l'intelligence artificielle entraînera des situations de « captivité » dans lesquelles certaines activités dépendront pour leur fonctionnement d'applications de l'IA. De quelles ressources dispose-t-on si l'IA ne fonctionne plus ou ne fonctionne plus correctement ? L'impact économique peut dépasser l'approche sectorielle et entamer la souveraineté (numérique) nationale.

Ces deux considérations doivent conduire à réfléchir à la place et au rôle du développeur dans la chaîne. Il ne s'agit pas d'ajouter un régime de responsabilité et de prendre le risque de brider le développement économique du secteur, mais des obligations pourraient être

imposées sous forme de principes au développeur d'IA afin de rééquilibrer l'asymétrie des responsabilités et de prévenir les risques systémiques. On peut penser, en premier lieu, au principe de transparence précité, notamment par l'explication de protocoles de test et des résultats obtenus sur chaque itération d'IA ou encore l'obligation d'indiquer l'objectif poursuivi par l'algorithme prédictif – ce qui peut être rapproché du principe, précédemment évoqué, de finalité du traitement en matière de données personnelles ou de la transparence exigée dans les algorithmes décisionnels publics par la loi pour une République numérique (art. L. 311-3-1 du code des relations entre le public et l'administration). Ce principe de transparence pourrait s'accompagner d'un principe de continuité des services décidée par une autorité administrative indépendante du type de l'agence européenne chargée de la robotique et de l'intelligence artificielle dont la création est préconisée dans la résolution adoptée par le Parlement européen. D'autres principes peuvent être envisagés : principe de neutralité de la base de référence, des tests et des traitements algorithmiques pour éviter d'aboutir à des résultats discriminatoires notamment ; principe de séparation des IA pour limiter le risque de propagation de biais viciant une intelligence artificielle ; droit de s'opposer à un traitement algorithmique ou de contester une décision issue d'un traitement algorithmique…

VI. Propriété intellectuelle et IA

L'IA peut intervenir dans le processus de création littéraire et artistique à deux titres principaux. Dans un premier cas de figure, l'IA peut être présentée comme un outil d'aide à la création, dont l'objet est de fournir une assistance à une personne physique. Les solutions juridiques traditionnelles sont applicables à cette situation, qui aboutit à l'attribution de la titularité des droits à la personne physique. Un second cas de figure correspond à la situation où l'IA, de manière autonome, développerait un résultat susceptible d'être qualifié d'œuvre, sous réserve du critère de l'originalité. Toutefois, en l'absence d'intervention d'un auteur personne physique, le critère d'originalité ne semble pas pouvoir être rempli, sauf à procéder à une refonte totale du régime du droit d'auteur. Selon le sous-groupe, la recherche des solutions juridiques les plus adaptées à ce dernier cas de figure devrait faire l'objet d'une réflexion approfondie dans le cadre de la stratégie à venir.

Les créations qui seraient générées par une intelligence artificielle, sans intervention directe d'une personne physique dans le processus créatif du résultat considéré risquent d'échapper à l'appropriation par le droit d'auteur, qui nécessite de qualifier l'empreinte de la personnalité ou encore l'effort technique personnalisé. Le droit d'auteur est intimement lié à la personne humaine dont la subjectivité et la capacité créative déterminent l'accession à la protection.

Les œuvres assistées par ordinateur sont protégeables par le droit d'auteur, l'ordinateur appréhendé comme un outil n'étant pas exclusif de créativité humaine. Ainsi, lorsque le processus créatif est laissé à la personne physique qui apporte l'empreinte de sa personnalité, cette personne sera auteur, au sens du code de la propriété intellectuelle, de l'œuvre considérée sous réserve de son originalité. À ce titre, la Cour d'appel de Paris a

déjà jugé que l'œuvre assistée par ordinateur « *peut être protégée par le droit d'auteur à conditions qu'apparaisse l'originalité voulue par le concepteur* » (Cour d'appel de Paris, 3 mai 2006, RG 05/03736). Tel sera le cas de la création générée par une intelligence artificielle au moyen des données sélectionnées par l'utilisateur en considération de ses choix et de sa personnalité. Sur un même modèle, la Nouvelle-Zélande, dans son Copyright Act du 15 décembre 1994, envisage la protection des œuvres créées par une intelligence artificielle dès lors que le processus créatif relève des choix d'une personne physique (article 5 : « *the person who creates a work shall be taken to be, (a) in the case of a literary, dramatic, musical, or artistic work that is computer-generated, the person by whom the arrangements necessary for the creation of the work are undertaken* »).

Tel n'est pas le cas de l'œuvre réalisée de façon autonome par une intelligence artificielle bénéficiant d'une capacité d'analyse de l'environnement ainsi que de facultés d'apprentissage et de subjectivité. La raison de cette absence de protection réside dans la définition de l'originalité qui est intimement liée à la personnalité de l'auteur et à l'intervention humaine sans laquelle aucune création ne peut prétendre à la protection par le droit d'auteur. L'importance de ce lien entre l'humain et la création n'est pas spécifique à la France : la Cour suprême australienne et la Cour de cassation hollandaise ont rappelé leur attachement à l'intervention humaine[25] et le Copyright Office américain en a fait une condition de l'enregistrement depuis l'affaire qui avait opposé un photographe et Wikimedia, le premier réclamant au second des droits sur un autoportrait réalisé avec son appareil par un singe[26].

Sur ce sujet, si certains auteurs estiment qu'il ne peut y avoir de droit d'auteur en l'absence de création réalisée par une personne physique, d'autres au contraire prônent une objectivisation du droit d'auteur (tant quant aux conditions d'accès à la protection qu'au régime d'exercice de ce droit). Le modèle du Royaume-Uni est alors cité en exemple en ce qu'il prévoit dans le Copyright, Design and Patent Act de 1988 la dévolution des droits sur les créations générés par ordinateur (article 178 (b) : « *is computer-generated a work which is generated by computer in circumstances where there is no human author, in connection with a work* ») à la personne ayant procédé aux paramétrages nécessaires à la création (article 9.3 « *In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person who undertakes the necessary arrangements for the creation of the work* »). Par ailleurs, le projet de rapport contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique du 31 mai 2016 (2015/2103 (INL)), dit Rapport Delvaux, « *relève qu'il n'existe aucune disposition juridique qui s'applique spécifiquement à la robotique, mais que les régimes et doctrines juridiques existants peuvent s'appliquer en l'état à ce domaine, certains aspects nécessitant néanmoins un examen approfondi* ». Le rapport « *demande de définir des critères de "création intellectuelle propre" applicables aux œuvres protégeables par droit*

[25] *Australia Sup. Court, Telstra Corporation Ltd c/ Phone Directories Compagny Pty Ltd* (2010), FCAFC 149 § 335 ; *Hoge Raad*, 30 mai 2008, *Zonen Endstra c/ Nieuw Amsterdam* (cité in A. Lucas, H.-J. Lucas et A. Lucas-Schloetter, *Traité de propriété littéraire et artistique*, LexisNexis, 2012, n° 57).

[26] Alexandra Bensamoun et Grégoire Loiseau, *L'intégration de l'intelligence artificielle dans l'ordre juridique : questions de temps*, à paraître.

d'auteur créées par des ordinateurs ou des robots ». Cette recommandation n'a pas été retenue dans le rapport final adopté par le Parlement européen le 16 février 2017.

Il résulte des échanges du sous-groupe que l'exclusion des créations générées par une intelligence artificielle autonome pourrait avoir un impact sur la valorisation des actifs immatériels au sein de l'entreprise notamment et nécessiterait de s'interroger sur une stratégie de valorisation adéquate. Toutefois, une telle stratégie de valorisation ne passerait pas obligatoirement par une refonte du droit d'auteur en niant le lien entre l'œuvre et l'humain, mais pourrait le cas échéant donner lieu à un régime de protection distinct.

À titre de constat général, le sous-groupe observe que les initiatives de régulation de l'IA, pour autant qu'elles se révéleraient nécessaires, ne devraient intervenir qu'une fois une certaine expérience acquise et ce, afin d'éviter de brider l'innovation. Dans cette perspective, l'innovation réglementaire aurait toute sa place et la France pourrait privilégier, au côté de la co-régulation, une initiative de règlementation expérimentale telle que prévue à l'article 37-1 de la Constitution[27]. Ce droit à l'expérimentation paraît un outil intéressant pour appréhender un domaine d'innovation forte telle que l'IA. Il pourrait permettre, par exemple, de tester la mise en place de réglementations dans un secteur donné.

[27] « La loi et le règlement peuvent comporter, pour un objet et une durée limitée, des dispositions à caractère expérimental ».

Annexe n° 1 – Synthèse des recommandations

Recommandation n° 1 : Selon les modèles d'exploitation pratiqués, deux cas de figure se présentent concernant la protection juridique de l'IA. Dans un premier cas, la technologie d'IA repose, selon le choix de leur exploitant, sur des algorithmes dans le domaine public ou, à titre alternatif, soumis à un régime dit « open source ». C'est ici, la valeur ajoutée résultant de leur combinaison ou de leur éducation/apprentissage que l'exploitant va chercher à protéger. Dans un second cas, c'est la protection des algorithmes en tant que tels qui peut être recherchée. Dans les deux cas, le recours aux droits de propriété intellectuelle (droit d'auteur, logiciel ou brevet) n'est pas totalement satisfaisant et les exploitants semblent privilégier à l'heure actuelle **la protection par le secret des affaires**. Il ressort des travaux du sous-groupe que la flexibilité (les garanties de confidentialité et conditions d'accès peuvent être définies par contrat) et la sécurité juridique apportées par cette solution sont accueillies favorablement par les exploitants dans un environnement technologique en évolution constante. *(Voir p. 2)*

Recommandation n° 2 : La confiance dans l'IA est directement liée à la transparence et à la loyauté des traitements algorithmiques. Afin que le besoin légitime de transparence ne devienne pas un frein pour l'innovation et sa valorisation, un équilibre doit être recherché. Aussi, afin de satisfaire cette exigence tant vis-à-vis du consommateur que dans une perspective concurrentielle, des initiatives de co-régulation tendent à l'adoption généralisée de bonnes pratiques. **À l'issue de ses travaux, le sous-groupe préconise que la France encourage le développement d'une initiative auprès des acteurs de l'IA pour la mise en place de grands principes de référence en la matière (guides et bonnes pratiques, norme relative aux critères de transparence…), permettant d'assurer une transparence objective.** *(Voir p. 5)*

Recommandation n° 3 : S'agissant des modèles économiques et techniques que le sous-groupe a eu l'occasion d'examiner dans le cadre de ses travaux, il apparaît que l'exploitant d'IA est positionné contractuellement vis-à-vis de son client en tant que prestataire de solution technique. Tant les données brutes, objets du traitement algorithmique, que les données résultant de ce traitement, demeurent contractuellement la propriété du client. Dans un tel schéma, **la problématique de propriété des données ne semble pas poser de difficulté particulière, pour autant que l'équilibre des intérêts soit de mise. Afin d'assurer cet équilibre, une veille sur les clauses contractuelles utilisées par les opérateurs économiques pourrait être mise en place afin d'identifier et de développer des bonnes pratiques en la matière.** *(Voir p. 8)*

Recommandation n° 4 : Concernant la responsabilité juridique des objets intelligents, il résulte des travaux du sous-groupe qu'il n'y pas à l'heure actuelle de vide juridique, étant donné que les instruments du droit positif de la responsabilité peuvent être mobilisés, afin de trancher les litiges susceptibles d'émerger. **Seul le recul de l'expérience pratique, au double point de vue juridique et économique, permettra, à terme, de faire des choix**. Ce recul permettra de proposer, le cas échéant, une adaptation pertinente en termes de droit de la responsabilité, si les limites pressenties des régimes actuels devaient être confirmées. **Des règles spéciales doivent cependant être d'ores et déjà envisagées**

s'agissant de certaines machines intelligentes, en particulier les voitures autonomes. *(Voir p. 11)*

Recommandation n° 5 : L'IA peut intervenir dans le processus de création littéraire et artistique à deux titres principaux. Dans un premier cas de figure, l'IA peut être présentée comme un outil d'aide à la création, dont l'objet est de fournir une assistance à une personne physique. Les solutions juridiques traditionnelles sont applicables à cette situation, qui aboutit à l'attribution de la titularité des droits à la personne physique. Un second cas de figure correspond à la situation où l'IA, de manière autonome, développerait un résultat susceptible d'être qualifié d'œuvre, sous réserve du critère de l'originalité. Toutefois, en l'absence d'intervention d'un auteur personne physique, le critère d'originalité ne semble pas pouvoir être rempli, sauf à procéder à une refonte totale du régime du droit d'auteur. **Selon le sous-groupe, la recherche des solutions juridiques les plus adaptées à ce dernier cas de figure devrait faire l'objet d'une réflexion approfondie dans le cadre de la stratégie à venir.** *(Voir p. 13)*

Recommandation n° 6 : À titre de constat général, le sous-groupe observe que les initiatives de régulation de l'IA, pour autant qu'elles se révéleraient nécessaires, ne devraient intervenir qu'une fois une certaine expérience acquise et ce, afin d'éviter de brider l'innovation. Dans cette perspective, l'innovation réglementaire aurait toute sa place et la France pourrait privilégier, au côté de la co-régulation, une initiative de règlementation expérimentale telle que prévue à l'article 37-1 de la Constitution[28]. Ce droit à l'expérimentation paraît un outil intéressant pour appréhender un domaine d'innovation forte telle que l'IA. Il pourrait permettre, par exemple, de tester la mise en place de réglementations dans un secteur donné. *(Voir p. 15)*

[28] « La loi et le règlement peuvent comporter, pour un objet et une durée limitée, des dispositions à caractère expérimental ».

Annexe n° 2 – Auditions et contributions

Dans le cadre de ses travaux, notre sous-groupe a organisé la tenue des auditions suivantes :

- **Mercredi 1er mars 2017 : auditions des représentants de Microsoft et d'Yseop**

Bernard Ourghanlian, directeur Technique et sécurité, Microsoft

Corinne Caillaud, directrice Affaires extérieures, affaires publiques et juridiques, Microsoft

Jean-Renaud Roy, directeur des Affaires publiques, Microsoft

Mathieu Coulaud, Head of Legal, Microsoft

John Rauscher, Directeur géneral, Yseop

- **Mercredi 8 mars 2017 : audition des représentants d'IBM et de SAP**

Vincent Martino, Direction juridique, IBM

Diane Dufoix-Garnier, directrice Affaires publiques, IBM

Amal Taleb, directrice adjointe Affaires publiques, SAP

- **Lundi 13 mars 2017 : audition des représentants de Google**

Nathan Klein, juriste

Olivier Esper, Senior Manager Public Policy

Hugo Ruggieri, stagiaire au département contentieux

- **Contributions écrites**

Benoît Gourdon, CEO, TELLMEPLUS (membre de Tech In France)

Pascal Gauvrit, Chief Technology Officer, EPTICA (membre de Tech In France)

Riccardo Masucci, Senior Privacy and Security Policy Manager, INTEL

Mario Romao, Senior Manager Policy, INTEL

Lionel Thoumyre, Director, Public Policy in IP, Content & Media EMEA, INTEL

STRATÉGIE NATIONALE EN INTELLIGENCE ARTIFICIELLE

L'INTELLIGENCE ARTIFICIELLE ET L'ENSEIGNEMENT DES TECHNOLOGIES GÉNÉRIQUES

Contribution au groupe de travail 3.2

« Anticiper les impacts économiques et sociaux de l'intelligence artificielle »

Auteurs :

Boris Le Hir, France Stratégie

Olivier Savrimoutoo, France Stratégie

Mars 2017

Sommaire

Le débat sur l'IA et l'emploi : une impression de déjà-vu 2

Les technologies génériques : un concept pour mieux penser l'impact de l'IA 3

Le déploiement des technologies génériques : un temps pour semer, un temps pour récolter 4

L'intelligence artificielle, une technologie générique ? 5

Penser le changement de paradigme techno-économique 6

Peut-on envisager des évaluations quantitatives *ex ante* ? 7

Conclusion 9

Les opinions exprimées dans ce document engagent leurs auteurs et n'ont pas vocation à refléter la position du gouvernement.

Le débat sur l'IA et l'emploi : une impression de déjà-vu

Les experts s'accordent à dire que l'intelligence artificielle représente un ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes qui requièrent une certaine forme d'intelligence. Le Centre national de ressources textuelles et lexicales (CNRTL) définit l'intelligence artificielle comme « la recherche de moyens susceptibles de doter les systèmes informatiques de capacités intellectuelles comparables à celles des êtres humains ».

La baisse du coût de la puissance de calcul parallèle, l'accès à de grandes bases de données, ainsi que de nouveaux algorithmes sont autant de transitions qui ont permis aux technologies issues de l'IA d'accomplir des tâches et de résoudre des problèmes d'une complexité considérable. De fait, cette technologie représente un levier de compétitivité majeur pour les entreprises, qui y voient un moyen d'accroître leur productivité en automatisant certains éléments de leur processus de production.

Toutefois, les impacts économiques d'une adoption massive de l'intelligence artificielle dans plusieurs secteurs d'activité suscitent de nombreux débats.

D'un côté, la description de ces technologies suggère qu'elles ont pour vocation de remplacer l'humain dans certaines tâches, d'où les craintes sur l'avenir de l'emploi. Des travaux publiés par diverses institutions qui mentionnent qu'une part élevée de l'emploi serait affectée par ces technologies alimentent cette anxiété : Frey et Osborne[1] estimaient en 2013 que 47 % des emplois sont susceptibles d'être en partie automatisés aux États-Unis ; plus récemment, l'OCDE[2], le Conseil d'orientation pour l'emploi[3] (COE) ou France Stratégie[4] ont évalué qu'une dizaine de pourcents des emplois, en moyenne, présentent un risque élevé d'automatisation, et que 50 % des tâches pourraient être automatisées pour 25 autres pourcents d'emplois ; enfin, le rapport de la Maison blanche sur l'automatisation affirme que ce sont les emplois les moins qualifiés qui encourent le plus grand risque d'automatisation[5].

D'un autre côté, les gains réalisés grâce à l'automatisation de certaines tâches pourraient être réinjectés dans l'économie par une baisse des prix, une hausse des salaires pour les employés (qualifiés) et/ou des profits plus importants pour les entreprises, ce qui aurait pour résultat une augmentation du PIB par habitant sur le long terme et une progression du niveau de vie.

---

[1] Carl Benedict Frey et Michael A. Osborne (2017), « The Future of employment: How susceptible are jobs to computerization », *Technological Forecasting and Social Change.*

[2] Melanie Arntz, Terry Gregory et Ulrich Zierahn (2016), « The risk of automation for jobs in OECD countries: A comparative analysis », *OECD Social, Employment and Migration Working Papers*, n° 189, OECD Publishing, Paris.

[3] Conseil d'orientation pour l'emploi (2017), *Automatisation, numérisation et emploi, Tome 1 : Les impacts sur le volume, la structure et la localisation de l'emploi*, janvier.

[4] Nicolas Le Ru (2016), « L'effet de l'automatisation sur l'emploi : ce qu'on sait et ce qu'on ignore » *La Note d'analyse*, n° 49, France Stratégie, juillet.

[5] Executive office of the President of the United States (2016), « Artificial intelligence, automation and the economy ».

Si la question d'un impact négatif sur l'emploi peut se poser, il reste que se limiter à une analyse d'effet direct est trop restrictif. Par le passé, des anticipations alarmantes sur l'emploi ont régulièrement été associées à des chocs technologiques, notamment dans l'agriculture et dans l'industrie, et se sont révélées en partie fausses. Elles étaient justement fondées sur une analyse limitée à l'impact direct sans prise en compte des autres effets induits potentiels et des effets de bouclage macroéconomique. Certes, à court terme, ces technologies sont censées accroître la productivité du travail et accélérer la substitution du capital au facteur travail. Certes, dans un monde où l'utilité marginale de la consommation est décroissante, l'accroissement de la demande issue de la baisse des coûts de production ne permettra sans doute pas de compenser cette moindre utilisation du travail pour une même quantité produite. Mais deux raisons principales font que cela ne suffit pas à affirmer que l'impact final sur l'emploi sera négatif. D'une part, il est faux de considérer que la composition de la production reste inchangée et que l'emploi perdu d'un côté ne peut pas être réalloué vers d'autres activités, éventuellement nouvelles. D'autre part, même si ces nouvelles technologies conduisent à rendre les processus de production plus efficaces et économes en travail, elles peuvent aussi contribuer à l'innovation de produit (apparition de nouveaux produits) permettant là encore de générer de la croissance et de l'emploi.

Les technologies génériques : un concept pour mieux penser l'impact de l'IA

Ces mécanismes sont mis en évidence par les théories économiques développées autour de la notion de technologie générique (TG). Celles-ci cherchent à déterminer les conditions sous lesquelles certaines technologies peuvent être des moteurs de croissance[6]. La littérature économique associe trois caractéristiques principales à ces technologies génériques[7] :

- ce sont des technologies qui affichent une dynamique de développement soutenue ;
- ce sont des technologies ubiquitaires, dans le sens où elles peuvent avoir de multiples utilisations dans de nombreux secteurs ;
- ce sont des technologies dites habilitantes[8] qui offrent de nombreuses opportunités d'innovations additionnelles.

Ces propriétés sont alors capables de générer une spirale vertueuse entre les producteurs et les utilisateurs de la technologie. Bresnahan et Trajtenberg[9] montrent en effet que les technologies vérifiant ces propriétés sont à même de devenir des moteurs de croissance par l'existence d'interactions entre les producteurs et les utilisateurs. En particulier, ils décrivent deux types d'externalités qui engendrent une complémentarité entre les innovations des

[6] Timothy Bresnahan et Manuel Trajtenberg (1995), « General purpose technologies, 'Engine of growth'? », *Journal of Econometrics*, vol. 65 (1), p. 83-108.
[7] Ces propriétés sont souvent décomposées en quatre comme dans Lipsey, Carlaw et Békar (2006), « Economic transformations ».
[8] Ce terme vient de l'expression « enabling technologies ».
[9] *Op. cit*.

producteurs (les inventions) et celles liées à l'adoption de la technologie par les utilisateurs (les co-inventions), et entre les co-inventions elles-mêmes[10].

D'un côté, l'innovation du producteur de technologie générique génère des opportunités d'innovation chez les utilisateurs. Par exemple, le développement de l'IA pour le véhicule autonome peut conduire le secteur des transports à la création de nouveaux services de transport « intelligents à la carte ». En retour, les innovations chez les utilisateurs, qui augmentent les rendements d'utilisation de la technologie générique, vont accroître la demande pour ces technologies, donc les incitations à innover chez le producteur, les autres utilisateurs bénéficiant alors de ces nouvelles innovations induites. Si l'on reprend notre exemple du véhicule autonome, plus on aura inventé de façons d'utiliser le véhicule autonome, plus le marché sera important et plus il sera rentable pour les concepteurs de véhicules autonomes d'investir en recherche afin d'accroître leur performance, ce qui bénéficiera en retour aux utilisateurs.

Ces externalités spécifiques au concept de technologie générique impliquent que les rendements sociaux des investissements dans ces technologies sont supérieurs aux rendements privés, et cela de façon plus importante que dans le cas des innovations traditionnelles. En contrepartie, elles révèlent que des investissements complémentaires à ceux dans la technologie génériques sont nécessaires. Les utilisateurs doivent en effet investir (se former, se réorganiser, voire faire de la recherche) pour être capables d'utiliser au mieux le potentiel de la technologie (en inventant de nouveaux usages).

Le déploiement des technologies génériques : un temps pour semer, un temps pour récolter[11]

Ces théories ont largement été développées et utilisées pour analyser l'impact économique de chocs technologiques historiques et en particulier pour expliquer les délais de leurs impacts. Paul David[12] par exemple rappelle que dans un premier temps la dynamo, inventée en 1900, n'a fait que remplacer la machine à vapeur sur les sites de production. Ce n'est que dans les années 1920, après l'invention du moteur unitaire (placé individuellement sur chaque machine), que les effets sur la productivité sont apparus. Grâce à cette « co-invention », les chaînes de production ont pu être réorganisées de façon beaucoup plus efficiente puisqu'il n'était plus nécessaire de disposer les machines le long d'un arbre de transmission alimenté par un moteur central (anciennement la machine à vapeur). La formalisation du concept de technologie générique a été développée au cours d'une période de doute sur le potentiel économique associé au progrès et au déploiement des technologies de l'information et de la communication (TIC) alors que ces technologies avaient suscité beaucoup d'attentes. Ce doute s'était notamment exprimé par le désormais célèbre paradoxe de Solow[13] : « Nous voyons les ordinateurs partout autour de nous, sauf

[10] Ces complémentarités caractérisent les secteurs producteurs et utilisateurs comme des compléments stratégiques (Bulow *et al*., 1985).
[11] Elhanan Helpman et Manuel Trajtenberg (1994), « A time to sow and a time to reap: Growth based on general purpose technologies », *NBER Working Paper*, n° 4854, septembre.
[12] Paul David (1991), « The computer and the dynamo: An historical perspective on the modern productivity paradox », *The American Economic Review*, vol. 80, p. 355-361.
[13] Robert Solow dans une déclaration pour le *New York Times Book Review*, 12 juillet 1987.

dans les statistiques de productivité ». Face à ce constat, la formalisation du concept de technologie générique a alors mis en évidence que si l'ordinateur est l'équivalent d'un vélo pour l'esprit[14], encore faut-il apprendre à pédaler. En d'autres termes, la diffusion des TIC ne suffit pas, il faut que des co-inventions par les utilisateurs se développent pour exploiter le potentiel de ces technologies. Ce n'est en effet qu'après 1995 que la croissance du PIB et de la productivité s'est accélérée aux États-Unis[15] et s'est propagée dans les secteurs utilisateurs intensifs où des investissements en formation et en capital organisationnel ont été nécessaires pour justement apprendre à utiliser ces nouveaux outils[16]. Cette approche permet donc d'expliquer en partie certaines dynamiques économiques mais elle éclaire aussi certaines divergences entre pays. Des études ont en effet montré que l'Europe a moins profité du développement des TIC que les États-Unis du fait à la fois d'un poids du secteur TIC plus faible et d'une moins bonne performance des secteurs de services intensifs en TIC[17]. Plusieurs facteurs (comme la taille du marché, la fragmentation linguistique, la régulation du marché du travail ou des biens et services par exemple) ont été avancés pour expliquer cette moindre capacité à développer et utiliser les TIC en Europe et ces analyses peuvent certainement apporter des enseignements pour le développement de l'IA en Europe.

L'intelligence artificielle, une technologie générique ?

Un certain nombre d'éléments laissent penser qu'il serait pertinent de considérer l'intelligence artificielle sous l'angle d'une technologie générique pour étudier son impact sur l'économie et l'emploi. En premier lieu, il semble que ces dernières années le développement de l'IA bénéficie d'une forte accélération faisant suite : i) à la baisse du coût de la puissance de calcul parallèle ; ii) à l'accès à de grandes bases de données ; iii) à de nouveaux algorithmes tirant parti des deux développements précédents. Différentes études ont aussi mis en évidence le potentiel d'utilisation de l'IA dans de multiples secteurs, que ce soit dans la santé, le transport, la distribution ou dans l'industrie. Enfin, il paraît assez intuitif que le déploiement de l'IA ne se fera pas uniquement par des investissements dans la technologie mais nécessitera de lourds investissements complémentaires en formation, en réorganisation et très certainement en développement de nouveaux produits. Il est aussi très probable que cela conduise à repenser certains systèmes économiques dans leur ensemble, i.e. des changements de paradigme techno-économiques. Le cas de la voiture autonome fournit une illustration de tels changements potentiels : il impliquera sans aucun doute des changements majeurs dans les secteurs de l'assurance et des services de

[14] Steve Jobs : « What a computer is to me is the most remarkable tool that we have ever come up with. It's the equivalent of a bicycle for our minds », *Memory and Imagination: New Pathways to the Library of Congress*, TV movie, 1990.

[15] Dale W. Jorgenson, Mun S. Ho et Kevin J. Stiroh (2008), « A retrospective look at the US productivity growth resurgence », *Journal of Economic Perspectives*, vol. 22 (1), p. 3-24 ; Gilbert Cette, Jacques Mairesse et Yusuf Kocoglu (2009), « La productivité en France, au Japon, aux Etats-unis et au Royaume-Uni au cours du XXe siècle », *Revue de l'OFCE*, n° 111, octobre.

[16] Erik Brynjolfsson et Adam Saunders (2009), *Wired for information: How information technology is reshaping the economy*, The MIT Press.

[17] Bart Van Ark, Mary O'Mahoy et Marchel P. Thimmer (2008), « The productivity gap between Europe and the United-States: Trends and causes », *Journal of economic perspectives*, vol. 22 (1), p. 25-44 ; Sophia P. Dimelis et Sotiris K. Papaioanou (2010), « ICT growth effects at the industry level: A comparison between the US and the EU », *Information Economics and Policy*, vol. 23 (1), p. 37-50.

transports, y compris hors transports routiers comme le train qui pourrait perdre ou gagner en utilité selon que la voiture autonome deviendrait un complément ou un substitut à ce moyen de transport ; il remettra en question l'existence des écoles de conduite, modifiera le métier des garagistes et des assurances ; il fera émerger de nouveaux « business models » où les individus ne posséderont plus leur véhicule mais loueront un service de transport autonome à la carte avec un véhicule adapté à leur besoin à chaque déplacement, le tout avec un risque d'accident fortement réduit et sans problème de congestion. Il est donc tout à fait possible que l'IA entraîne l'innovation mais dans quelle proportion, la réponse reste bien entendu très incertaine.

Penser le changement de paradigme techno-économique

Sous cet angle, la question de l'impact sur l'emploi du déploiement de l'IA devient donc multiple :

- Quels emplois seront en effet directement touchés par le choc technologique ? Avec quelle ampleur et avec quelle dynamique ? En d'autres termes, quels sont les « métiers du passé » et quelle est la population la plus susceptible d'être impactée ?
- À l'inverse, quels sont les secteurs et les métiers vers lesquels les travailleurs déclassés par la technologie peuvent être redirigés ? En d'autres termes, quels sont les « métiers d'avenir » ?
- Quelles sont les possibilités de reconversion des « métiers du passé » vers les « métiers d'avenir » ? Sur ce point, la notion de compétence est bien sûr primordiale mais la dimension territoriale doit aussi avoir un rôle important dans l'analyse.
- Quels sont les secteurs faisant partie de la chaîne de production de l'intelligence artificielle et est-il important de les développer sur le territoire ? Le fait d'importer ou de produire la technologie est-il déterminant sur son effet ?
- Quels sont les secteurs potentiellement utilisateurs de l'IA et quel impact l'IA peut-elle avoir sur ces secteurs ? Va-t-elle uniquement conduire à accroître leur productivité ou va-t-elle aussi leur permettre d'innover et de développer de nouveaux biens et services ?
- Quels impacts potentiels globaux la technologie peut-elle avoir sur l'ensemble du système productif ? Quelle est l'ampleur du changement de paradigme techno-économique ?
- Quels sont les efforts à mettre en œuvre pour libérer le potentiel de croissance que ces technologies peuvent générer ?

Ce sont autant de questions auxquelles il est nécessaire de répondre pour déterminer si le développement et la diffusion de l'IA auront un impact négatif ou positif sur l'emploi et pour définir les meilleures stratégies à adopter en termes d'investissement dans la technologie, tant au niveau de sa production et de son utilisation, qu'en matière de formation des travailleurs pour anticiper le plus tôt possible les difficultés de transition.

Peut-on envisager des évaluations quantitatives *ex ante* ?

Une approche par analogie

Afin de prendre en compte les effets indirects potentiels du développement de l'IA, les auteurs d'une étude d'Analysis Group[18] ont tenté des approches par analogies.

Dans une première approche, Chen *et al.* considèrent que les principaux effets indirects du développement de l'IA pourraient être comparables, en proportion, au développement d'autres technologies. En particulier, ils s'intéressent aux externalités associées aux investissements en R & D privée et publique et en capital-risque. Ils reprennent pour cela les valeurs des rendements associés à ces investissements disponibles dans la littérature empirique[19] et déduisent que, selon ces valeurs, 1 $ investi en R & D privée, en R & D publique et en capital-risque dans l'IA rapporterait respectivement 1,99 $, 2,69 $ et 3,33 $ de PIB supplémentaire.

Cette approche est intéressante pour mettre en évidence les rendements sociaux supérieurs à l'unité des investissements dans la recherche et le développement d'une nouvelle technologie mais elle reste très approximative et, surtout, elle ne prend pas en compte les mécanismes économiques propres à ce type de technologie qui ont été décrits ci-dessus.

Dans une seconde approche, ces auteurs envisagent la possibilité que le déploiement de l'IA puisse avoir des effets similaires à ceux qu'ont pu avoir d'autres technologies par le passé. Dans cette approche, ils se fondent notamment sur les évaluations empiriques des impacts des technologies de l'information, de l'internet haut débit, de la téléphonie mobile ou encore de la robotique industrielle. En utilisant le développement et l'adoption de ces technologies comme *benchmark*, Chen *et al.* estiment des ordres de grandeur de l'impact que pourrait avoir l'IA sur l'économie. Si l'on admet que les technologies considérées ici sont génériques, cette approche est sans doute plus à même de prendre en compte les effets globaux de la technologie, à condition que les évaluations historiques intègrent bien ces mécanismes. Les résultats obtenus suivant cette approche sont nettement supérieurs à ceux issus de la première démarche. Néanmoins la transposition directe des impacts de ces technologiques à l'IA est difficilement envisageable ; il est certainement plus pertinent de chercher à décomposer les effets pour mieux appréhender les principaux canaux de transmission dans l'économie et de tester l'importance relative de ces canaux.

Le recours à la modélisation macro-sectorielle

Outre ces analogies directes qui peuvent être trop approximatives, il est possible d'approfondir le raisonnement par l'utilisation de modèles macro-sectoriels. Parmi ces modèles, Némésis[20] incorpore des mécanismes de croissance endogène et attribue aux TIC

---

[18] Nicholas Chen, Lau Christensen, Kevin Gallagher, Rosamond Mate et Greg Rafert (2015), *Global Economic Impacts Associated with Artificial Intelligence*, Analysis Group Inc., étude financée par Facebook Inc.

[19] Ils se basent en particulier sur l'article d'Astrid Romain et Bruno Van Pottelsberghe (2004), « The economic impact of venture capital » qui évalue l'élasticité de l'output aux investissements en R & D privée, en R & D publique et en capital-risque à respectivement : 19,9 %, 13,6 % et 0,9 %.

[20] Le modèle Némésis est un modèle macro-sectoriel représentant 28 pays européens avec une décomposition sectorielle en 30 secteurs. Ce modèle a été développé par un consortium européen comprenant le laboratoire Erasme de l'École centrale de Paris – relayé depuis par l'entreprise Seureco – le Bureau fédéral du Plan belge et

des propriétés de technologie générique. Il permet en ce sens d'apporter un certain éclairage à diverses questions sur l'impact de l'intelligence artificielle dans l'économie française.

Ces mécanismes sont introduits dans le modèle par une fonction de production d'innovation multidimensionnelle dépendant, dans chaque secteur, des investissements dans trois actifs : la R & D, les TIC et d'autres intangibles (logiciels et formation)[21]. En outre, le modèle distingue, dans les effets d'innovation, ceux d'augmentation de la productivité et ceux de création de nouveaux produits. Cette distinction est d'autant plus importante dans notre analyse que l'impact final sur l'emploi tient au fait que l'effet dominant du progrès technique est soit l'accroissement de la productivité et la moindre utilisation des facteurs, soit l'amélioration des produits et l'accroissement de la demande.

Toutes les précautions d'usage dans l'interprétation des résultats doivent bien entendu être prises puisque ce modèle ne représente pas explicitement le développement et la diffusion de l'IA. Néanmoins, le fait qu'il intègre des propriétés de technologies génériques associées aux TIC met en évidence les canaux de transmission des impacts de ce type de technologie.

La nécessité de tester la sensibilité des résultats aux différentes hypothèses

Comme nous l'avons décrit plus haut, l'analogie entre la diffusion de l'IA et celle des TIC de façon générale semble pertinente mais la transposition ne peut pas se faire directement. Sur la base de cette approche, il reste à déterminer quelles pourraient être les spécificités de l'IA par rapport aux logiciels classiques ou aux TIC en général. Pour cela, plusieurs questions se posent :

- Quels secteurs et quels pays sont les principaux producteurs d'IA ?
- Quels secteurs sont potentiellement les principaux utilisateurs d'IA ?
- L'IA a-t-elle des effets de complémentarité et d'externalité similaires à ceux des TIC et des logiciels ?
- Les innovations induites par le déploiement de l'IA sont-elles davantage des innovations de produit ou des innovations de procédé comparées aux innovations traditionnelles moyennes? En d'autres termes, les paramètres de pondération entre les effets produits et les effets process dans le modèle doivent-ils être calibrés différemment ?

---

l'université polytechnique nationale d'Athènes (ICCS/NTUA). Il a par ailleurs été en grande partie construit dans le cadre de projets de recherche européens et a bénéficié de l'apport de chercheurs de diverses institutions telles que le laboratoire MERIT de l'université des Nations-Unies, le Centre for European Economic Research (ZEW) ou l'université libre de Bruxelles. Ce modèle a été jusqu'à présent principalement utilisé pour l'évaluation de politiques publiques aux niveaux national et européen dans les domaines de la fiscalité et du coût du travail, de l'énergie et de l'environnement et enfin dans le domaine de l'innovation.

[21] Cette extension des fonctions d'innovation du modèle a été réalisée grâce au développement des bases de données EU-KLEMS, qui fournissent des données sectorielles détaillées sur les investissements en TIC et en logiciels, et INTAN-Invest, fournissant des données pour divers investissements intangibles à des niveaux sectoriels. Ces bases de données, en particulier celles sur les actifs intangibles, sont encore en développement et pourront très certainement être étendues et améliorées.

Répondre à ces questions requiert un travail de futurologue dont les prospectives ne pourront être qu'incertaines.

Outre des analyses de scénarios simples, il est donc nécessaire, pour évaluer un champ des possibles, de réaliser des tests de sensibilité aux différentes hypothèses et aux différents calibrages du modèle. Des tests sur le poids des innovations de produit par rapport aux innovations de productivité, sur les complémentarités entre les facteurs d'innovation, sur les externalités générées par les efforts d'innovation ou encore sur la part des importations dans les investissements en IA permettraient d'identifier l'influence de ces variables.

Conclusion

Évaluer *ex ante* l'impact économique que l'IA peut avoir dans les prochaines décennies (parce qu'il semble bien que ce soit l'horizon temporel nécessaire pour qu'une telle technologie déploie son potentiel) est pour ainsi dire mission impossible. Néanmoins, il est nécessaire d'anticiper suffisamment tôt les mutations pour en prévenir les effets négatifs et en saisir les opportunités. Pour cette raison, il est tout à fait utile de réaliser des exercices de prospective mais il est indispensable de considérer les changements dans leur complexité et de tenter de déterminer leurs principaux canaux de diffusion. Anticiper l'impact de l'IA au moins dans sa nature et, autant que possible, dans son ampleur ne peut donc se limiter à recenser les emplois que cette technologie pourra substituer mais doit prendre en compte des effets complexes caractéristiques de technologies dites génériques et des effets de bouclages macroéconomiques. Il est en effet utile d'identifier les métiers qui risquent de subir un déclassement ainsi que les risques sociaux associés au déploiement de l'IA mais il tout aussi primordial de définir dans quels domaines et par quels moyens ces nouvelles technologies peuvent être un levier puissant. Les mutations que peuvent engendrer l'IA sont certainement importantes et il est très probable que celles-ci puissent être utiles dans la quête de nouveaux objectifs sociaux comme la transition écologique par exemple. Bien entendu, il ne s'agit pas de considérer que tout ce qui est nouveau est bon à prendre mais le fatalisme ou la crispation dans le changement n'aboutit en général qu'à en subir les effets négatifs sans les bénéfices. Il faut être acteur dans le changement pour être capable de prendre les bonnes décisions et de l'orienter de façon cohérente avec les préférences collectives.