L'intelligence artificielle dans la société

OCDE

L'intelligence artificielle dans la société

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Merci de citer cet ouvrage comme suit :
OCDE (2019), *L'intelligence artificielle dans la société*, Éditions OCDE, Paris,
https://doi.org/10.1787/b7f8cd16-fr.

ISBN 978-92-64-60811-5 (imprimé)
ISBN 978-92-64-94312-4 (pdf)

Les données statistiques concernant Israël sont fournies par et sous la responsabilité des autorités israéliennes compétentes. L'utilisation de ces données par l'OCDE est sans préjudice du statut des hauteurs du Golan, de Jérusalem-Est et des colonies de peuplement israéliennes en Cisjordanie aux termes du droit international.

Crédits photo : Couverture © Adobe Stock.

Les corrigenda des publications de l'OCDE sont disponibles sur : *www.oecd.org/about/publishing/corrigenda.htm*.

Préface

L'intelligence artificielle (IA) remodèle les économies et promet de générer des gains de productivité, d'améliorer l'efficience et de réduire les coûts. Elle contribue à une vie meilleure et aide les individus à affiner les prévisions et prendre des décisions plus éclairées. Pour autant, l'IA, qui n'en est qu'à ses débuts, est loin d'avoir révélé tout son potentiel en tant que pourvoyeuse de solutions face aux défis mondiaux et levier d'innovation et de croissance. Alors que ses impacts se font sentir au sein de la société, son pouvoir de transformation doit être placé au service des individus et de la planète.

Dans le même temps, l'IA préoccupe et soulève des questions éthiques. On s'interroge sur la fiabilité des systèmes, notamment sur les dangers de codifier et renforcer les biais existants, en particulier ceux liés au genre et à l'origine raciale, ou les risques de violation des droits de l'homme et des valeurs fondamentales comme le respect de la vie privée. Et l'on s'inquiète de voir les systèmes d'IA aggraver les inégalités, le changement climatique, la concentration des marchés et la fracture numérique. Aucun pays ou acteur ne dispose à lui seul de toutes les solutions à ces défis. D'où l'importance de miser sur la coopération internationale et les approches multipartites pour guider le développement et l'utilisation de l'IA afin qu'elle serve le plus grand nombre.

La présente étude, intitulée *L'intelligence artificielle dans la société*, examine le paysage de l'IA et met en évidence les grandes problématiques intéressant l'action des pouvoirs publics. Elle vise à favoriser une compréhension commune de l'IA actuelle et à court terme, et susciter un dialogue aussi large que possible sur les grands enjeux stratégiques que sont l'évolution des marchés du travail et la valorisation des compétences à l'ère du numérique ; la protection de la vie privée ; la redevabilité quant aux décisions prises sur fond d'IA ; ou les questions de responsabilité, de sûreté et de sécurité qu'elle soulève.

L'étude s'appuie sur les travaux du Groupe d'experts sur l'intelligence artificielle à l'OCDE, créé en 2018 pour rédiger des principes à même de faciliter l'innovation en matière d'IA, son adoption et la confiance dans les systèmes et technologies connexes. Les débats qu'il a menés ont nourri l'élaboration de la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle* – première norme intergouvernementale sur l'IA – adoptée par l'ensemble des membres de l'OCDE et plusieurs pays partenaires le 22 mai 2019. Ces travaux mettent en évidence la nécessité d'une coopération internationale pour façonner un cadre d'action qui favorise la confiance dans l'IA et son adoption.

Des efforts restent à déployer pour approfondir ensemble les questions techniques, éthiques et juridiques liées à l'IA, en vue d'harmoniser les normes et les codes de conduite tout en veillant à l'interopérabilité des législations et des règlementations. Nous devons nous y atteler d'urgence, compte tenu du rythme d'évolution et du champ infini des applications. Sans surprise, l'IA figure d'ailleurs en bonne place dans les priorités nationales et internationales, y compris celles du G7 et du G20.

L'adoption de la Recommandation et l'instauration d'un dialogue mondial marquent des premiers jalons essentiels. Mais le chemin est encore long. Avec le lancement, cette année, de l'Observatoire OCDE des politiques relatives à l'IA, nous mobilisons notre expertise dans les domaines de l'analyse, de la mesure et des politiques pour explorer des territoires

encore largement inconnus. L'Observatoire – plateforme inclusive dédiée aux politiques publiques sur l'IA – aidera les pays à encourager, accompagner et suivre le développement responsable de systèmes d'IA dignes de confiance, dans l'intérêt de la société.

Bientôt, le moment viendra de traduire les principes en actions. L'OCDE entend aider les pays à mettre en œuvre la Recommandation, afin de faire en sorte que nos sociétés et nos économies tirent le meilleur parti de l'IA et en partagent largement les bienfaits, tout en apportant les garanties nécessaires pour ne laisser personne de côté – aujourd'hui et pour les générations à venir.

Angel Gurría

Secrétaire général

OCDE

Avant-propos

Cette étude a pour objectif de favoriser une compréhension commune de l'intelligence artificielle (IA), actuelle et à court terme. Elle recense les incidences économiques et sociales des technologies et applications d'IA et passe en revue les répercussions sur l'action des pouvoirs publics, en présentant des données probantes et des solutions stratégiques. Elle a en outre vocation à favoriser la coordination et la cohérence avec les débats menés au sein d'autres instances internationales, dont le G7, le G20, l'Union européenne et les Nations Unies.

L'étude s'appuie sur les conclusions de la conférence organisée par l'OCDE en octobre 2017 sur le thème *AI: Intelligent Machines, Smart Policies* (http://oe.cd/ai2017) ; sur les activités et débats menés par le Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO) entre septembre 2018 et février 2019 ; et sur la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle*. Elle a, à son tour, nourri le projet de l'OCDE « Vers le numérique » et la publication de l'OCDE *Vers le numérique : forger des politiques au service de vies meilleures*.

Le chapitre 1, qui dresse un tour d'horizon du « Paysage technique de l'IA », retrace la genèse de l'intelligence artificielle, de la naissance de l'IA symbolique dans les années 50 aux dernières avancées dans le domaine de l'apprentissage automatique. Il présente les travaux menés par le Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO) pour décrire les systèmes d'IA – qui établissent des prévisions, formulent des recommandations ou prennent des décisions influant sur l'environnement – et leur cycle de vie. Il propose également une taxinomie de recherche destinée à aider les décideurs à décrypter les tendances en matière d'IA et identifier les enjeux stratégiques qui en découlent.

Le chapitre 2, consacré au « Paysage économique de l'IA », présente le rôle de l'IA en tant que technologie émergente à visée générique ouvrant la voie à une diminution du coût des prévisions et une optimisation de la prise de décisions. L'exploitation de l'IA exige de réaliser des investissements complémentaires dans les données, les compétences et la transformation numérique des flux de travail, et d'être à même d'adapter les processus organisationnels. Ce chapitre passe également en revue les tendances en termes de capital-investissement dans les startups spécialisées dans l'IA.

Le chapitre 3, qui aborde les « Applications de l'intelligence artificielle », recense dix domaines dans lesquels les technologies de l'IA connaissent une percée rapide – les transports, l'agriculture, la finance, le marketing et la publicité, la science, la santé, la justice pénale, la sécurité, le secteur public et les applications de réalité augmentée et de réalité virtuelle. Dans tous ces domaines, l'utilisation de l'IA est synonyme d'amélioration de l'efficience de la prise de décisions, de réduction des coûts et d'optimisation des ressources.

Le chapitre 4, qui étudie les « Considérations de politique publique », examine les problématiques phares que soulève la diffusion de l'IA. Il expose les Principes de l'OCDE en matière d'IA, adoptés en mai 2019, dont il détaille d'abord les valeurs : croissance inclusive, développement durable et bien-être ; valeurs centrées sur l'humain et équité ; transparence et explicabilité ; robustesse, sûreté et sécurité ; et responsabilité. Dans un deuxième temps, il s'intéresse aux politiques nationales à mettre en place pour promouvoir

des systèmes d'IA dignes de confiance et se penche en particulier sur des domaines d'action phares : l'investissement dans des activités de recherche et de développement responsables en matière d'IA ; l'instauration d'un écosystème numérique propice à l'IA ; la mise en place d'un cadre d'action favorisant l'innovation dans l'IA ; l'accompagnement des travailleurs afin de les préparer à la transformation des emplois et le développement des compétences ; et la mesure des progrès.

Le chapitre 5, consacré aux « Politiques et initiatives dans le domaine de l'IA », illustre la place de plus en plus importante qu'elle occupe dans les priorités d'action des acteurs aux niveaux tant national qu'international. Tous les groupes de parties prenantes – pouvoirs publics et organisations intergouvernementales, mais aussi entreprises, organismes techniques, universitaires, société civile et syndicats – mènent une réflexion sur la façon d'orienter le développement et le déploiement de l'IA afin qu'ils servent l'ensemble de la société.

La présente étude a été déclassifiée, selon la procédure écrite, par le Comité de la politique de l'économie numérique de l'OCDE (CPEN) le 10 avril 2019 et préparée en vue de sa publication par le Secrétariat de l'OCDE.

Remerciements

La publication *L'intelligence artificielle dans la société* a été préparée sous l'égide du Comité de la politique de l'économie numérique de l'OCDE (CPEN), avec le concours de ses groupes de travail. Les délégués auprès du CPEN ont apporté une importante contribution, que ce soit en faisant part de leurs observations et modifications, ou en partageant l'expérience de leurs pays respectifs et en passant en revue les stratégies nationales en matière d'IA.

Les auteurs principaux de la publication sont Karine Perset, Nobuhisa Nishigata et Luis Aranda, de la Division de la politique de l'économie numérique de l'OCDE – Karine Perset en ayant assuré l'édition et la coordination générales. Anne Carblanc, Chef de la Division de la politique de l'économie numérique de l'OCDE ; Andrew Wyckoff et Dirk Pilat, respectivement Directeur et Directeur adjoint de la Direction de la science, de la technologie et de l'innovation, ont piloté et supervisé le projet. Les recherches et la rédaction de certaines parties de l'étude ont été menées à bien par Doaa Abu Elyounes, Gallia Daor, Lawrence Pacewicz, Alistair Nolan, Elettra Ronchi, Carlo Menon et Christian Reimsbach-Kounatze. Des experts de l'ensemble de l'OCDE, dont Laurent Bernat, Stijn Broecke, Dries Cuijpers, Marie-Agnès Jouanjean, Caroline Paunov, Luke Slawomirski, Mariagrazia Squicciarini, Barbara Ubaldi, Joao Vasconcelos et Jeremy West, ont fourni des orientations, des contributions et des commentaires.

L'étude a bénéficié de la contribution de Taylor Reynolds et Jonathan Frankle, de l'Internet Policy Research Initiative du MIT ; de Douglas Frantz, consultant indépendant ; d'Avi Goldfarb, de l'université de Toronto ; de Karen Scott, de l'université de Princeton ; de la Commission syndicale consultative auprès de l'OCDE ; d'Amar Ashar, de Ryan Budish, de Sandra Cortesi, de Finale Doshi-Velez, de Mason Kortz et de Jessi Whitby, du Berkman Klein Center for Internet and Society de l'université de Harvard ; Joanna Bryson de l'Université de Bath ; et des membres du Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO). L'équipe de rédaction tient à remercier tout particulièrement Nozha Boujemaa, Marko Grobelnik, James Kurose, Michel Morvan, Carolyn Nguyen, Javier Juárez Mojica et Matt Chensen pour leurs contributions et commentaires précieux.

L'étude s'est également nourrie des travaux menés actuellement à l'échelle de l'OCDE. Citons notamment ceux du Comité de la politique scientifique et technologique et de son Groupe de travail sur la politique de l'innovation et de la technologie ; du Comité de la politique à l'égard des consommateurs et de son Groupe de travail sur la sécurité des produits de consommation ; du Comité de l'industrie, de l'innovation et de l'entrepreneuriat et de son Groupe de travail sur l'analyse de l'industrie ; du Comité de l'emploi, du travail et des affaires sociales ; du Comité des politiques d'éducation ; et de l'initiative e-leaders du Comité de la gouvernance publique et du Comité de la concurrence, sans oublier le Comité de la politique de l'économie numérique et ses groupes de travail, en particulier le Groupe de travail sur la sécurité et la vie privée dans l'économie numérique.

Les auteurs remercient en outre Mark Foss, pour son travail d'édition, et Alice Weber et Angela Gosmann pour leur soutien éditorial. Leur travail a grandement contribué à la qualité globale de la publication.

Enfin, de vifs remerciements vont au ministère japonais des Affaires Intérieures et des Communications (MIC), pour le soutien qu'il a apporté à ce projet.

Table des matières

Tableaux

Graphiques

Encadrés

Acronymes, abréviations et monnaies

3D	Tridimensionnel
A.L.I.C.E.	*Artificial Linguistic Internet Computer Entity*
ACM	*Association for computing machinery*
AI HLEG	Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle (Commission européenne)
AIGO	Groupe d'experts sur l'intelligence artificielle à l'OCDE
AINED	Partenariat public-privé « L'intelligence artificielle pour les Pays-Bas » (Pays-Bas)
AUD	Dollar australien
CAD	Dollar canadien
CE	Commission européenne
CEA	Commissariat à l'énergie atomique et aux énergies alternatives
CESE	Comité économique et social européen
CHF	Franc suisse
CIFAR	Institut canadien de recherches avancées (Canada)
CITI	Classification internationale type, par industrie, de toutes les branches d'activité économique
CMS	Calcul multipartite sécurisé
CNRC	Conseil national de recherches Canada (Canada)
CNY	Yuan ren min bi (Chine)
COMEST	Commission mondiale d'éthique des connaissances scientifiques et des technologies
CPC	Comité de la politique à l'égard des consommateurs
CPEN	Comité de la politique de l'économie numérique
CPST	Comité de la politique scientifique et technologique
DARPA	*Defense Advanced Research Projects Agency*
DKK	Couronne danoise
DPI	Droits de propriété intellectuelle
DSI	Dossier de santé informatisé
EPIC	*Electronic Privacy Information Center*
EUR	Euro
FAO	Organisation des Nations Unies pour l'alimentation et l'agriculture

FATML	*Fairness, accountability and transparency in machine learning*
FDA	*Food and Drug Administration*
FICO	*Fair, Isaac and Company*
FIT	Forum international des transports
G20	Groupe des vingt
G2IA	Groupe international d'experts en intelligence artificielle
G7	Groupe des sept
GBP	Livre sterling
GM	General Motors
HoME	*Household Multimodal Environment*
IA	Intelligence artificielle
IAE	Intelligence artificielle étroite
IAG	Intelligence artificielle générale
IdO	Internet des objets
IEC	Commission électrotechnique internationale
IEEE	*Institute for Electrical and Electronics Engineers*
IPRI	*Internet Policy Research Initiative* (Massachusetts Institute of Technology)
ISO	Organisation internationale de normalisation
ITI	*Information Technology Industry Council*
KRW	Won coréen
MIT	*Massachusetts Institute of Technology*
MPI	Institut Max Planck pour l'innovation et la concurrence
NOK	Couronne norvégienne
OAI	*Office for Artificial Intelligence* (Royaume-Uni)
ODD	Objectifs de développement durable (Nations Unies)
PAI	*Partnership on Artificial Intelligence to Benefit People and Society*
PIAAC	Programme pour l'évaluation internationale des compétences des adultes (OCDE)
PISA	Programme international pour le suivi des acquis des élèves
PME	Petites et moyennes entreprises
Po	Pétaoctet
RATP	Régie autonome des transports parisiens
RCM	Réunion du Conseil au niveau des Ministres (OCDE)
R-D	Recherche et développement
RGPD	Règlement général sur la protection des données (Union européenne)

SAE	*Society of Automotive Engineers*
SAI	Systèmes autonomes et intelligents
SAR	Riyal saoudien
SNCF	Société nationale des chemins de fer français
STIM	Sciences, technologies, ingénierie et mathématiques
TIC	Technologies de l'information et des communications
TLN	Traitement du langage naturel
UE	Union européenne
UGAI	*Universal Guidelines on Artificial Intelligence*
UIT	Union internationale des télécommunications
USD	Dollar américain

Résumé

L'apprentissage automatique, les données massives et la puissance de calcul ont impulsé les progrès récents de l'IA

Le paysage technique de l'intelligence artificielle (IA) s'est métamorphosé depuis 1950, lorsqu'Alan Turing s'est interrogé pour la première fois sur la capacité des machines à penser. L'intelligence artificielle, expression consacrée en 1956, a évolué au fil des décennies : à une IA symbolique, marquée par la conception de systèmes fondés sur la logique, a succédé un temps d'arrêt (l'« hiver de l'IA ») dans les années 70, avant la naissance de l'ordinateur Deep Blue doté d'un programme de jeu d'échecs, dans les années 90. Depuis 2011, les progrès décisifs réalisés dans le domaine de l'« apprentissage automatique », branche de l'IA qui s'appuie sur une approche statistique, ont permis d'accroître la capacité des machines à formuler des prévisions à partir de données historiques. La maturité d'une technique de modélisation de l'apprentissage automatique dénommée « réseaux neuronaux », conjuguée à des ensembles de données volumineux et à l'augmentation de la puissance de calcul, ont contribué à l'accélération du développement de l'IA.

Les systèmes d'IA établissent des prévisions, formulent des recommandations ou prennent des décisions influant sur les environnements

Selon la description proposée par le Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO), un système d'intelligence artificielle (ou système d'IA) est un

> *système automatisé qui, pour un ensemble donné d'objectifs définis par l'homme, est en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur des environnements réels ou virtuels. Pour ce faire, il se fonde sur des entrées machine et/ou humaines pour percevoir les environnements réels et/ou virtuels ; transcrire ces perceptions en modèles (par des moyens automatisés, en s'appuyant par exemple sur l'apprentissage automatique, ou manuels) ; et utiliser des inductions des modèles pour formuler des possibilités de résultats (informations ou actions à entreprendre). Les systèmes d'IA sont conçus pour fonctionner à des niveaux d'autonomie divers.*

Le cycle de vie d'un système d'IA comporte les phases suivantes : i) une phase de planification et de conception, de collecte et de traitement des données, de construction du modèle et d'interprétation ; ii) une phase de vérification et de validation ; iii) une phase de déploiement ; et iv) une phase d'exploitation et de suivi. Une taxinomie de recherche sur l'IA opère une distinction entre les applications de l'IA, comme le traitement du langage naturel ; les techniques d'entraînement des systèmes d'IA, avec par exemple les réseaux neuronaux ; les solutions d'optimisation telles que l'apprentissage à partir d'un exemple unique (*one-shot learning*) ; et la recherche axée sur les considérations sociétales, comme le besoin de transparence.

L'IA peut contribuer à améliorer la productivité et aider à résoudre des problèmes complexes

Le paysage économique de l'IA évolue à mesure que l'intelligence artificielle s'impose comme une nouvelle technologie générique. L'IA, qui offre un moyen de produire des prévisions, des recommandations ou des décisions plus fiables à moindre coût, promet de générer des gains de productivité, d'améliorer le bien-être et d'aider à relever des défis complexes. Son exploitation exige de réaliser des investissements complémentaires dans les données, les compétences et les flux de travail numériques, et de modifier les processus organisationnels. C'est pourquoi son adoption varie selon les entreprises et les secteurs.

L'IA est un domaine où les investissements et le développement des entreprises progressent rapidement

Après cinq années de croissance régulière, le capital-investissement dans des startups spécialisées dans l'IA s'est accéléré à partir de 2016. Le volume des investissements a en effet doublé entre 2016 et 2017, pour atteindre 16 milliards USD en 2017. Les startups spécialisées dans l'IA ont attiré 12 % du capital-investissement mondial au cours du premier semestre de 2018, en nette progression par rapport à 2011, où elles n'en concentraient que 3 % – une tendance observée dans toutes les grandes économies. Ces investissements correspondent, en règle générale, à des opérations de grande envergure se montant à plusieurs millions de dollars. À mesure que les technologies et les modèles économiques gagnent en maturité, on s'achemine vers un déploiement à grande échelle de l'IA.

Les applications de l'IA sont légion, des transports à la science, en passant par la santé

Les applications de l'IA font une percée rapide dans les secteurs où elles sont à même de détecter des schémas dans des volumes considérables de données et de modéliser des systèmes complexes interdépendants en vue d'améliorer la prise de décisions et de réduire les coûts.

- Dans les transports, les véhicules autonomes, grâce aux systèmes de conduite virtuelle, aux cartes haute définition et aux itinéraires optimisés, laissent entrevoir des avantages en termes d'économies, de sécurité, de qualité de vie et de protection de l'environnement.
- Dans la recherche scientifique, l'IA est utilisée pour collecter et traiter des données à grande échelle, reproduire les expérimentations et en réduire les coûts, et accélérer la découverte scientifique.
- Dans le secteur de la santé, les systèmes d'IA aident à diagnostiquer et prévenir les maladies et les épidémies le plus tôt possible, découvrir des traitements et des médicaments, ou encore proposer des interventions personnalisées ; en outre, ils ouvrent la voie à des outils d'autosurveillance.
- Dans le domaine de la justice pénale, l'IA est utilisée à des fins de police prédictive et d'évaluation des risques de récidive.
- Les applications de sécurité numérique font appel aux systèmes d'IA pour automatiser la détection des menaces et la réponse aux incidents, de plus en plus souvent en temps réel.
- Dans le domaine de l'agriculture, l'IA aide à surveiller l'état des cultures et des sols et à prévoir l'impact des facteurs environnementaux sur le rendement des cultures.

- Les services financiers s'appuient sur l'IA pour la détection de la fraude, l'évaluation de la solvabilité des emprunteurs, la réduction des coûts des services à la clientèle, la négociation automatisée et la mise en conformité réglementaire.
- Dans les domaines du marketing et de la publicité, l'IA est utilisée pour explorer les données sur le comportement des consommateurs en vue de cibler et de personnaliser les contenus, la publicité, les biens et les services, ainsi que les recommandations et les tarifs.

La confiance dans l'IA est une condition essentielle pour en tirer le meilleur parti

L'IA présente certes des avantages, mais appelle également une réflexion sur l'action des pouvoirs publics. De fait, des efforts doivent être déployés afin de veiller à ce que les systèmes d'IA soient dignes de confiance et centrés sur l'humain. L'IA – en particulier certaines techniques d'apprentissage automatique – suscite des inquiétudes inédites en termes d'éthique et d'équité. Posent particulièrement problème les questions liées au respect des droits de l'homme et des valeurs démocratiques, ainsi que les risques de transposition des biais du monde analogique vers le monde numérique. Certains systèmes d'IA sont si complexes qu'il peut s'avérer impossible d'expliquer comment ils sont parvenus aux décisions qui ont été prises. Il est par conséquent essentiel de veiller à la transparence sur le recours à l'IA et de permettre la détermination des responsabilités quant aux résultats qui en découlent. Les systèmes d'IA doivent par ailleurs fonctionner convenablement et être sûrs et sécurisés.

Des politiques nationales doivent être mises en place afin de promouvoir des systèmes d'IA dignes de confiance et, notamment, d'encourager l'investissement dans la recherche et le développement responsables. Outre les technologies liées à l'IA et la puissance de calcul, l'intelligence artificielle exploite des volumes considérables de données. D'où l'impérieuse nécessité de bâtir un environnement numérique permettant l'accès aux données, avec des mécanismes solides de protection des données et de la vie privée. Les écosystèmes propices au développement de l'IA peuvent également aider les petites et moyennes entreprises à prendre le virage de l'IA et favoriser un environnement concurrentiel.

L'IA va transformer la nature du travail à mesure qu'elle remplace et/ou modifie les composantes du travail humain. Les politiques devront par conséquent aider les travailleurs à passer d'un emploi à l'autre et garantir que soient déployés des efforts continus en termes d'éducation, de formation et de développement des compétences.

L'IA est une priorité croissante pour toutes les parties prenantes

Compte tenu des mutations induites par l'intelligence artificielle, porteuses à la fois d'avantages et de risques, l'IA constitue une priorité d'action croissante pour l'ensemble des parties prenantes. De nombreux pays disposent d'ores et déjà de stratégies dédiées, qui abordent l'IA comme un moteur de croissance et de bien-être, visent à favoriser la formation et le recrutement de la prochaine génération de chercheurs, et s'attachent à identifier les meilleures approches pour affronter les défis connexes. Des mesures sont également prises par les acteurs non gouvernementaux – entreprises, organisations techniques, milieux universitaires, société civile et syndicats – et des instances internationales telles que le G7, le G20, l'OCDE, la Commission européenne et les Nations Unies.

En mai 2019, l'OCDE a adopté ses Principes sur l'intelligence artificielle, premier ensemble de normes internationales convenu par les pays pour favoriser une approche responsable au service d'une IA digne de confiance, élaboré avec le concours d'un groupe d'experts multipartite.

1. Paysage technique de l'IA

Ce chapitre décrit les caractéristiques du paysage technique de l'intelligence artificielle (IA), qui s'est métamorphosé depuis 1950, lorsqu'Alan Turing s'est interrogé pour la première fois sur la capacité des machines à penser. Depuis 2011, des progrès décisifs ont été réalisés dans une branche de l'IA dénommée « apprentissage automatique », qui permet à des machines de s'appuyer sur des approches statistiques pour apprendre à partir de données historiques et formuler des prévisions dans des situations nouvelles. La maturité des techniques d'apprentissage automatique, conjuguée à des ensembles de données volumineux et à l'augmentation de la puissance de calcul, ont contribué à l'accélération du développement de l'IA. Ce chapitre donne également un aperçu général d'un système d'IA, qui établit des prévisions, formule des recommandations ou prend des décisions influant sur l'environnement. Il décrit ensuite le cycle de vie type d'un système d'IA, qui se décompose en quatre phases, à savoir : i) la phase de « conception, données et modèles », qui comprend la planification et la conception, la collecte et le traitement des données, ainsi que la construction et l'interprétation du modèle ; ii) la phase de « vérification et validation » ; iii) la phase de « déploiement » ; et iv) la phase d'« exploitation et (de) suivi ». Enfin, il propose une taxinomie de recherche à l'intention des décideurs.

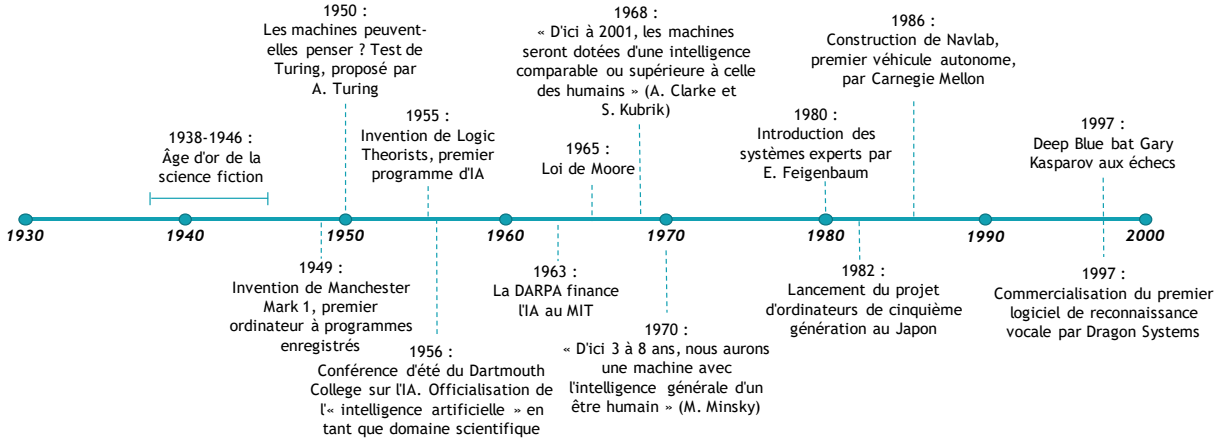
Genèse de l'intelligence artificielle

En 1950, Alan Turing, mathématicien britannique, publie un article sur l'ordinateur et l'intelligence, intitulé *Computing Machinery and Intelligence* (Turing, 1950[1]), dans lequel il s'interroge sur la capacité des machines à penser. Il développe alors une heuristique simple pour tester son hypothèse : un ordinateur pourrait-il mener une conversation et répondre à des questions d'une manière qui puisse conduire une personne suspicieuse à penser que l'ordinateur est en réalité un humain[1] ? De là naît le « test de Turing », encore utilisé de nos jours. La même année, Claude Shannon propose la création d'une machine à laquelle on pourrait apprendre à jouer aux échecs (Shannon, 1950[2]). L'entraînement de la machine pouvait alors se faire en recourant à la force brute ou en évaluant un ensemble réduit de déplacements stratégiques de l'adversaire (UW, 2006[3]).

Nombreux sont ceux qui considèrent le *Dartmouth Summer Research Project*, mené à l'été 1956, comme le point de départ de l'intelligence artificielle (IA). Lors de cet atelier, John McCarthy, Alan Newell, Arthur Samuel, Herbert Simon et Marvin Minsky ont conceptualisé le principe de l'IA. Si les recherches dans le domaine de l'IA n'ont cessé de progresser au cours des 60 dernières années, les promesses de ses précurseurs se révèlent à l'époque par trop optimistes. L'IA connaît alors, dans les années 70, un temps d'arrêt (on parle de l'« hiver de l'IA »), marqué par une chute des financements et de l'intérêt pour la recherche connexe.

On observe dans les années 90 un regain sur ces deux fronts, à la faveur des progrès en termes de puissance de calcul (UW, 2006[3]). Le Graphique 1.1 propose une chronologie de l'évolution de l'IA depuis sa naissance.

Graphique 1.1. Chronologie de l'évolution de l'IA (des années 50 à 2000)



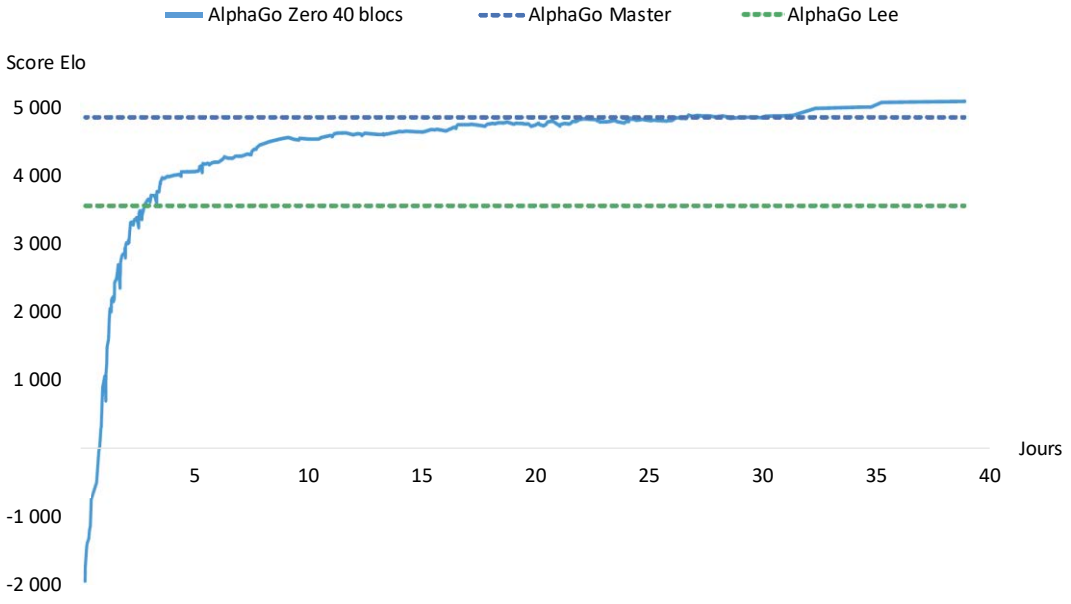
Source : D'après Anyoha (28 août 2017[4]), « The history of artificial intelligence », http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

Après l'« hiver de l'IA », qui prend fin dans les années 90, les progrès de la puissance de calcul et des capacités de stockage des données rendent possible l'exécution de tâches complexes. En 1995, l'IA franchit une étape décisive, avec le développement, par Richard Wallace, d'A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*), un programme capable de tenir une conversation basique. Toujours dans les années 90, IBM met au point un ordinateur nommé Deep Blue, qui s'appuie sur une approche fondée sur la force brute pour affronter le champion du monde d'échecs, Gary Kasparov. Deep Blue est alors capable d'anticiper

six étapes ou plus et de calculer 330 millions de positions par seconde (Somers, 2013[5]). En 1996, Deep Blue perd contre Kasparov, avant de remporter la revanche un an plus tard.

En 2015, DeepMind, filiale d'Alphabet, lance un logiciel à même d'affronter au jeu ancestral de Go les meilleurs joueurs mondiaux. Pour ce faire, il fait appel à un réseau neuronal artificiel qui a appris à jouer en s'entraînant sur des milliers de parties exécutées par des professionnels et des amateurs humains. En 2016, AlphaGo bat le champion du monde de l'époque, Lee Sedol, quatre jeux à un. Ses développeurs laissent alors le programme jouer contre lui-même en s'appuyant uniquement sur un apprentissage par essai et erreur, et en commençant par des parties totalement aléatoires, sur la base de quelques règles simples. De là naît le programme AlphaGo Zero, capable, avec un entraînement accéléré, de battre le programme AlphaGo initial par 100 jeux à 0. Entièrement autodidacte – sans intervention humaine ni utilisation de données d'historique –, AlphaGo Zero parvient, en 40 jours, à surpasser toutes les autres versions d'AlphaGo (Silver et al., 2017[6]) (Graphique 1.2).

Graphique 1.2. Auto-apprentissage rapide d'AlphaGo pour devenir le champion du monde de jeu de Go en 40 jours



Source : D'après Silver et al. (2017[6]), « Mastering the game of Go without human knowledge », http://dx.doi.org/10.1038/nature24270.

Situation actuelle

Au cours des dernières années, la montée en puissance des données massives, de l'infonuagique et des capacités de calcul et de stockage connexes, alliée aux progrès d'une branche de l'IA nommée « apprentissage automatique », ont dopé la puissance, la disponibilité, le développement et l'impact de l'IA.

Par ailleurs, les avancées technologiques constantes ouvrent la voie à des capteurs plus performants et abordables, qui capturent des données plus fiables venant nourrir les systèmes d'IA. Les volumes de données auxquels accèdent les systèmes d'IA continuent de croître à mesure que la taille et le coût réduits des capteurs en favorisent le déploiement. Cela permet, par ricochet, de réaliser des progrès majeurs dans de nombreux domaines de recherche en IA fondamentale, notamment :

- le traitement du langage naturel
- les véhicules autonomes et la robotique
- la vision par ordinateur
- l'apprentissage des langues.

Certaines des évolutions les plus intéressantes de l'IA ont lieu non pas dans l'informatique, mais dans des domaines comme la santé, la médecine, la biologie et la finance. La transition vers l'IA ressemble à bien des égards au processus de diffusion des ordinateurs qui, après avoir été l'apanage de quelques entreprises spécialisées, a gagné l'ensemble de l'économie et de la société dans les années 90. Elle n'est pas non plus sans rappeler l'expansion de l'accès à l'internet, des entreprises multinationales à une majorité de la population de nombreux pays, dans les années 2000. Les économies vont avoir de plus en plus besoin de personnes disposant de doubles spécialités, c'est-à-dire spécialisées dans une discipline comme l'économie, la biologie ou le droit, mais disposant également de compétences dans les techniques d'IA telles que l'apprentissage automatique. Le présent chapitre s'intéresse aux applications qui sont utilisées ou se profilent à court et moyen termes, plutôt qu'à de possibles évolutions à plus long terme, comme l'intelligence générale artificielle (en anglais, *artificial general intelligence*, ou AGI) (Encadré 1.1).

Encadré 1.1. Intelligence étroite artificielle et intelligence générale artificielle

L'intelligence étroite artificielle (en anglais *artificial narrow intelligence*, ou ANI), également dénommée IA « appliquée », est conçue pour accomplir une tâche de raisonnement ou de résolution de problème spécifique. Elle correspond à l'état actuel de la technologie. Les systèmes d'IA les plus perfectionnés disponibles aujourd'hui, comme AlphaGo de Google, n'en restent pas moins « étroits ». Ils peuvent, dans une certaine mesure, généraliser la reconnaissance de schémas et de formes, en transférant par exemple des connaissances apprises dans le domaine de la reconnaissance d'images vers celui de la reconnaissance de la parole. Toutefois, l'esprit humain est bien plus polyvalent.

On oppose souvent à l'IA appliquée l'intelligence générale artificielle (*artificial general intelligence*, ou AGI), qui reste pour l'heure hypothétique. Dans ce cas, les machines autonomes deviendraient capables d'actions relevant de l'intelligence générale. Comme les humains, elles seraient en mesure de généraliser et de synthétiser des apprentissages acquis par le biais de différentes fonctions cognitives. L'AGI serait assortie d'une mémoire associative développée et d'une capacité de raisonnement et de prise de décisions. Elle pourrait résoudre des problèmes multifacettes, apprendre par la lecture ou l'expérience, créer des concepts, percevoir le monde mais aussi elle-même, inventer et faire preuve de créativité, réagir aux imprévus dans des environnement complexes, ou encore anticiper. Les points de vue divergent sensiblement sur les perspectives d'AGI. Les experts sont d'avis qu'il convient d'être réaliste en termes de calendrier. Ils s'accordent dans l'ensemble sur le fait que l'ANI donnera lieu à des opportunités, des risques et des défis nouveaux importants, et conviennent que l'avènement éventuel de l'AGI, peut-être au cours du XXIe siècle, décuplerait ces conséquences.

Source : OCDE (2018[7]), *Perspectives de l'économie numérique de l'OCDE 2017*, https://doi.org/10.1787/9789264282483-fr.

L'IA, qu'est-ce que c'est ?

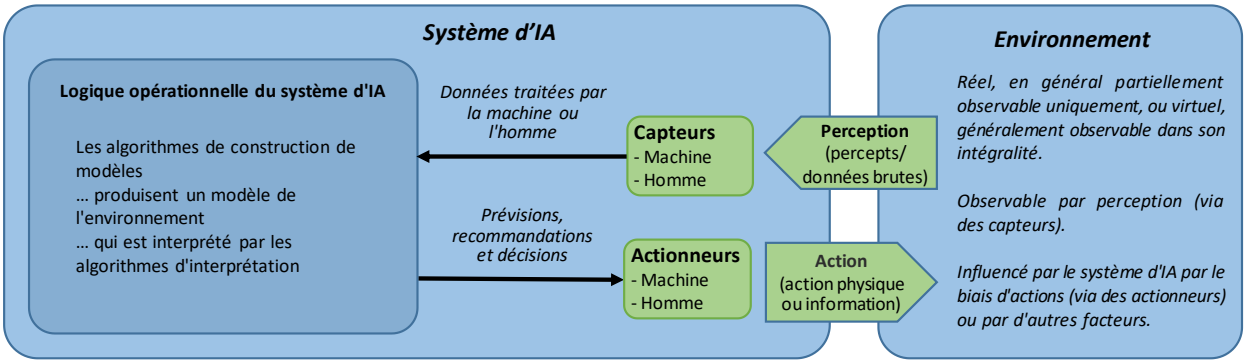
Il n'existe pas de définition universellement admise de l'IA. En novembre 2018, le Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO) a créé un sous-groupe chargé d'élaborer une description d'un système d'IA. Cette description entend être compréhensible, techniquement juste, neutre du point de vue de la technologie et applicable aux horizons à court et long termes. Elle est suffisamment large pour couvrir nombre de définitions de l'IA couramment utilisées par la communauté scientifique, les entreprises et les pouvoirs publics. Elle a également nourri l'élaboration de la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle* (OCDE, 2019[8]).

Vision conceptuelle d'un système d'IA

La présente description d'un système d'IA s'appuie sur la vision conceptuelle de l'IA exposée dans l'ouvrage *Artificial Intelligence: A Modern Approach* (Russel et Norvig, 2009[9]). Celle-ci est cohérente avec la définition fréquente de l'IA comme « ensemble des mécanismes permettant à un agent de percevoir, de raisonner et d'agir » (Winston, 1992[10]) et avec des définitions générales similaires (Gringsjord et Govindarajulu, 2018[11]).

Une première vision conceptuelle de l'IA donne à voir une structure de haut niveau d'un système d'IA générique (également dénommé « agent intelligent ») (Graphique 1.3). Un système d'IA comporte trois éléments principaux : des capteurs, une logique opérationnelle et des actionneurs. Les capteurs collectent des données brutes à partir de l'environnement, tandis que les actionneurs agissent de manière à modifier l'état de l'environnement. La véritable puissance d'un système d'IA réside dans sa logique opérationnelle. Pour un ensemble déterminé d'objectifs et à partir de données d'entrée issues des capteurs, la logique opérationnelle produit des résultats en sortie à l'intention des actionneurs. Ceux-ci prennent la forme de recommandations, de prévisions ou de décisions susceptibles d'influer sur l'état de l'environnement.

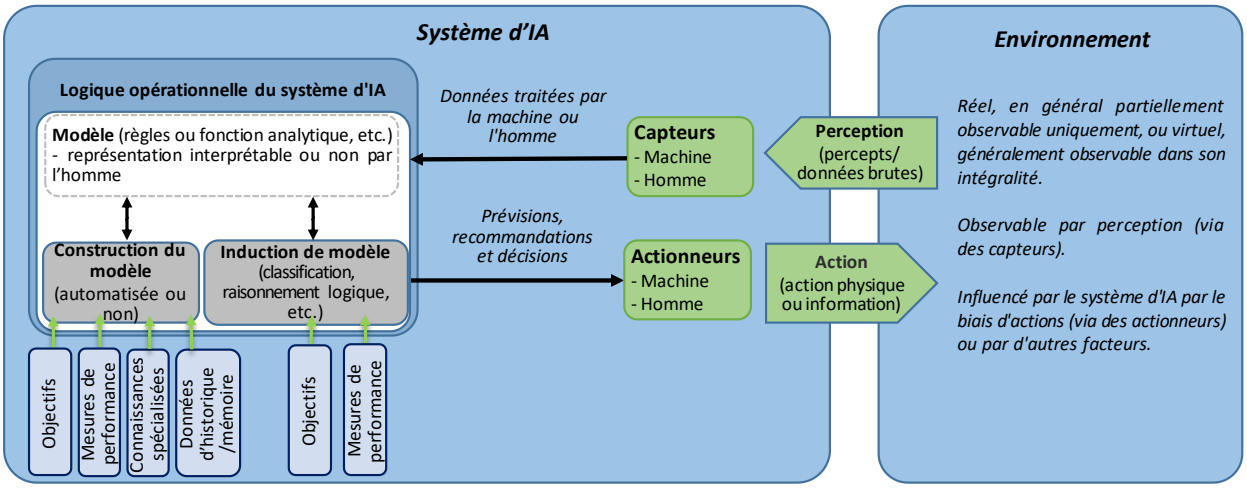
Graphique 1.3. Vision conceptuelle de haut niveau d'un système d'IA



Source : Tel que défini et approuvé par l'AIGO en février 2019.

Une structure plus détaillée rend compte des principaux éléments liés aux dimensions des systèmes d'IA intéressant l'action des pouvoirs publics (Graphique 1.4). Pour couvrir différents types de systèmes d'IA et divers scénarios, le diagramme distingue le processus de construction du modèle (comme l'apprentissage automatique) du modèle lui-même. La construction du modèle est également séparée du processus d'interprétation, qui utilise le modèle pour établir des prévisions, formuler des recommandations et prendre des décisions ; les actionneurs utilisent ces résultats pour influer sur l'environnement.

Graphique 1.4. Vision conceptuelle détaillée d'un système d'IA



Source : Tel que défini et approuvé par l'AIGO en février 2019.

Environnement

Dans le contexte d'un système d'IA, un environnement est un espace observable par le biais de perceptions (via des capteurs) et influencé au moyen d'actions (via des actionneurs). Les capteurs et les actionneurs sont soit des machines, soit des hommes. Les environnements sont quant à eux soit réels (environnement physique, social, mental, etc.) et, en règle générale, partiellement observables uniquement, soit virtuels (à l'instar des jeux, par exemple) et généralement observables dans leur intégralité.

Système d'IA

Un système d'IA est un système automatisé qui, pour un ensemble donné d'objectifs définis par l'homme, est en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur des environnements réels ou virtuels. Pour ce faire, il se fonde sur des entrées machine et/ou humaines pour : i) percevoir les environnements réels et/ou virtuels ; ii) transcrire ces perceptions en modèles grâce à une analyse manuelle ou automatisée (s'appuyant par exemple sur l'apprentissage automatique) ; et iii) utiliser des inductions des modèles pour formuler des possibilités de résultats (informations ou actions à entreprendre). Les systèmes d'IA sont conçus pour fonctionner à des niveaux d'autonomie divers.

Modèle d'IA, construction et interprétation de modèle

Au cœur d'un système d'IA se trouve le modèle d'IA, représentation de tout ou partie de l'environnement externe du système qui en décrit la structure et/ou la dynamique. Un modèle peut être fondé sur des connaissances spécialisées et/ou des données, émanant d'humains et/ou d'outils automatisés (des algorithmes d'apprentissage automatique, par exemple). Les objectifs (tels que les variables de sortie) et les mesures de performance (fiabilité, ressources d'entraînement, représentativité de l'ensemble de données) guident le processus de construction. L'induction est le processus par lequel les humains et/ou les outils automatisés déduisent des résultats à partir du modèle. Ceux-ci prennent la forme de recommandations, de prévisions ou de décisions. Les objectifs et les mesures de performance guident l'exécution. Dans certains cas (règles déterministes), le modèle peut générer une seule recommandation. Dans d'autres (modèles probabilistes), il peut en proposer une variété. Ces recommandations sont associées à différents niveaux, par exemple de mesures de performance telles que le niveau de confiance,

de robustesse ou de risque. Il peut être possible, au cours du processus d'interprétation, d'expliquer pourquoi des recommandations particulières ont été formulées ; parfois, c'est impossible.

Exemples de systèmes d'IA

Système d'évaluation des risques-clients

Un système d'évaluation des risques-clients est un exemple de système automatisé influant sur son environnement (octroi ou non d'un prêt à une personne). Il émet des recommandations (cotes de crédit) pour un ensemble donné d'objectifs (solvabilité). Pour ce faire, il utilise à la fois des entrées machine (données sur les profils des personnes et sur leur historique de remboursement d'emprunts) et des entrées humaines (ensemble de règles). À partir de ces deux jeux d'entrées, le système perçoit les environnements réels (à savoir si les personnes remboursent régulièrement leurs emprunts), puis transcrit automatiquement ces perceptions en modèles. Un algorithme d'évaluation des risques-clients pourrait par exemple utiliser un modèle statistique. Enfin, il a recours à des inductions (algorithme d'évaluation des risques) pour formuler des recommandations (cotes de crédit) sur les possibilités de résultats (accorder ou refuser le prêt).

Assistant pour malvoyants

Un assistant pour les personnes souffrant d'une déficience visuelle illustre la manière dont un système automatisé influe sur son environnement. Il émet des recommandations (comment un malvoyant peut éviter un obstacle ou traverser une rue) pour un ensemble donné d'objectifs (se déplacer d'un endroit à un autre). Pour ce faire, il utilise des entrées machine et/ou humaines (de volumineuses bases de données contenant des images étiquetées d'objets, de mots écrits, voire de visages humains) à trois fins. Premièrement, il perçoit les images de l'environnement (une caméra capture une image de ce qui se trouve devant une personne et la transmet à une application). Deuxièmement, il transcrit automatiquement ces perceptions en modèles (algorithmes de reconnaissance d'objets, capables de reconnaître un feu de signalisation, un véhicule ou un obstacle sur le trottoir). Troisièmement, il recourt à des inductions pour recommander des possibilités de résultats (en fournissant une description audio des objets détectés dans l'environnement) afin que la personne puisse décider de l'action à entreprendre et, par conséquent, influer sur son environnement.

AlphaGo Zero

AlphaGo Zero est un système d'IA capable de battre au jeu de Go n'importe quel joueur professionnel. L'environnement du jeu est virtuel et observable dans son intégralité. Les positions sont contraintes par les objectifs et les règles du jeu. Le système AlphaGo Zero utilise à la fois des entrées humaines (les règles du jeu de Go) et des entrées machine (apprentissage par le jeu itératif contre lui-même, en commençant par des parties entièrement aléatoires). Il transcrit alors les données en modèle (stochastique) d'actions (« déplacements » dans le jeu) entraîné à l'aide d'une technique dite d'« apprentissage par renforcement ». Enfin, il utilise le modèle pour proposer un nouveau déplacement d'après l'état d'avancement de la partie.

Système de conduite automatisée

Le système de conduite automatisée est un exemple de système automatisé capable d'influer sur son environnement (en décidant si le véhicule accélère, ralentit ou opère un virage). Il formule des prévisions (prévoit si un objet ou un panneau correspond à un obstacle ou à une instruction) et/ou prend des décisions (accélérer, freiner, etc.) pour un ensemble donné

d'objectifs (aller d'un point A à un point B en toute sécurité, le plus rapidement possible). Pour ce faire, il utilise à la fois des entrées machine (données d'historique de conduite) et des entrées humaines (ensemble de règles de conduite). Ces entrées sont utilisées pour créer un modèle du véhicule et de son environnement. Il permet ainsi au système d'atteindre trois objectifs. Premièrement, il peut percevoir les environnements réels (via des capteurs comme des caméras et des sonars). Deuxièmement, il peut transcrire automatiquement ces perceptions en modèles (reconnaissance d'objet ; vitesse et détection de trajectoire ; et données géodépendantes). Troisièmement, il peut recourir à l'induction. Par exemple, l'algorithme de conduite automatisée peut comporter de nombreuses simulations de possibilités à court terme pour le véhicule et son environnement. Il peut ainsi recommander des possibilités de résultats (s'arrêter ou avancer).

Cycle de vie d'un système d'IA

En novembre 2018, l'AIGO a créé un sous-groupe dont les travaux sont destinés à éclairer l'élaboration de la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle* (OCDE, 2019[8]) en détaillant le cycle de vie d'un système d'IA. Ce cadre n'a pas vocation à devenir une nouvelle norme sur le cycle de vie de l'IA[2] ni à proposer des directives. En revanche, il peut aider à définir un contexte pour d'autres initiatives internationales sur les principes de l'IA[3].

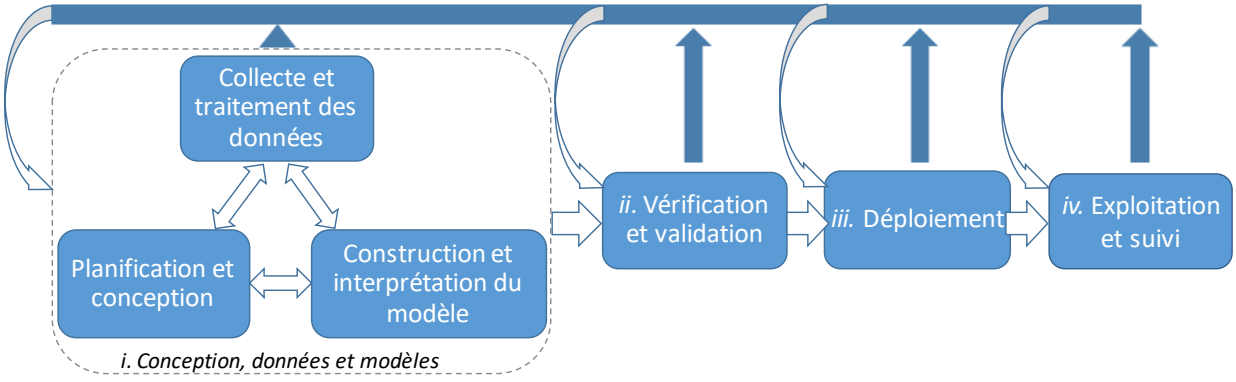
Un système d'IA présente de nombreuses phases communes avec les cycles traditionnels de développement des logiciels et, plus généralement, des systèmes. En revanche, le cycle de vie type d'un système d'IA comporte quatre phases spécifiques. La phase de conception, données et modèles est une séquence dépendante du contexte comprenant la planification et la conception, la collecte et le traitement des données, ainsi que la construction et l'interprétation du modèle. Elle est suivie des phases de vérification et validation, puis de déploiement et, enfin, d'exploitation et de suivi (Graphique 1.5. Cycle de vie d'un système d'IA). Ces phases présentent souvent un caractère itératif et ne respectent pas nécessairement un ordre séquentiel. La décision de mettre un terme à l'utilisation d'un système d'IA peut intervenir à n'importe quel stade de la phase d'exploitation et de suivi.

Les phases du cycle de vie d'un système d'IA peuvent être décrites comme suit :

1. La phase de **conception, données et modèles** comprend plusieurs activités, dont l'ordre peut varier selon les systèmes d'IA.
    - La **planification et la conception** du système d'IA couvrent la définition du concept et des objectifs du système, des principes sous-jacents, du contexte et du cahier des charges, ainsi que la construction éventuelle d'un prototype.
    - La **collecte et le traitement des données** englobent les tâches visant à recueillir et nettoyer les données, réaliser les vérifications d'exhaustivité et de qualité, et documenter les métadonnées et les caractéristiques de l'ensemble de données. Les métadonnées intègrent les informations relatives aux modalités de création de l'ensemble de données, à sa composition, aux usages prévus et à sa maintenance au fil du temps.
    - La **construction et l'interprétation du modèle** couvrent la création ou le choix des modèles ou des algorithmes, leur calage et/ou leur entraînement, ainsi que leur interprétation.
2. La phase de **vérification et validation** comprend l'exécution et le réglage des modèles, avec des tests visant à évaluer les performances au regard de diverses dimensions et considérations.

3. La phase de **déploiement** (mise en production) englobe le pilotage, la vérification de la compatibilité avec les systèmes existants, la mise en conformité réglementaire, la gestion des changements organisationnels et l'évaluation de l'expérience des utilisateurs.
4. La phase d'**exploitation et** de **suivi** couvre l'exploitation du système d'IA et l'évaluation permanente de ses recommandations et de ses effets (attendus et imprévus) au regard des objectifs et des considérations éthiques. C'est au cours de cette phase que l'on identifie les problèmes, opère les ajustements en revenant aux autres phases, voire, si nécessaire, abandonne la production du système d'IA.

Graphique 1.5. Cycle de vie d'un système d'IA



Source : Tel que défini et approuvé par l'AIGO en février 2019.

De par la centralité des données et des modèles dont l'entraînement et l'évaluation dépendent des données, le cycle de vie de nombreux systèmes d'IA se distingue du cycle traditionnel de développement des systèmes. Certains systèmes d'IA faisant appel à l'apprentissage automatique peuvent fonctionner par itérations et évoluer au fil du temps.

Recherche en matière d'IA

Cette section passe en revue certaines évolutions techniques qui ont marqué la recherche en matière d'intelligence artificielle dans les secteurs universitaire et privé et favorisent la transition vers l'IA. L'IA, en particulier sa branche dénommée « apprentissage automatique », est devenue un domaine de recherche active de l'informatique. Un nombre croissant de disciplines universitaires mettent à profit les techniques d'IA pour un large éventail d'applications.

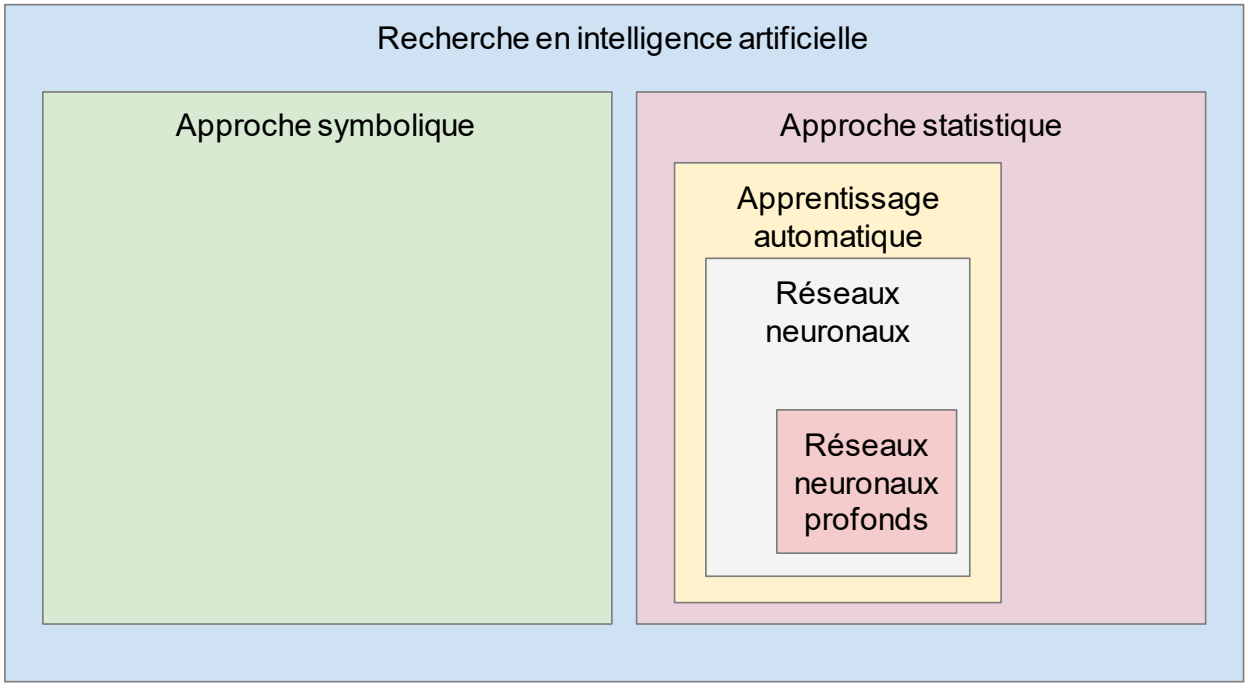
Il n'existe pas de système de classification communément admis pour la répartition des activités d'IA en domaines de recherche, à l'image par exemple des 20 grandes catégories de recherche économique du système de classification *Journal of Economic Literature*. Cette section entend proposer une taxinomie de recherche sur l'IA ayant vocation à aider les décideurs à décrypter certaines tendances récentes de l'IA et identifier les enjeux en termes d'action des pouvoirs publics.

Par le passé, les travaux de recherche ont opéré une distinction entre l'IA symbolique et l'IA statistique. L'IA symbolique s'appuie sur des représentations logiques pour aboutir à une conclusion à partir d'un ensemble de contraintes. Elle exige que les chercheurs construisent des structures décisionnelles détaillées, compréhensibles par l'homme, pour traduire la complexité du monde réel et aider les machines à parvenir à des décisions semblables à celles des humains. L'IA symbolique est encore aujourd'hui couramment utilisée, par exemple pour l'optimisation et la planification des outils. L'IA statistique, qui

permet aux machines d'inférer une tendance à partir de schémas, connaît depuis peu un engouement grandissant. Un certain nombre d'applications allient les approches symbolique et statistique. Ainsi, il n'est pas rare que des algorithmes de traitement du langage naturel conjuguent des approches statistiques (qui exploitent d'importants volumes de données) et des approches symboliques (qui tiennent compte de considérations comme les règles de grammaire). La combinaison de modèles s'appuyant à la fois sur les données et sur l'expertise humaine pourrait aider à lever les contraintes liées à chacune des deux approches.

Les systèmes d'IA font un usage croissant de l'apprentissage automatique. Il s'agit là d'un ensemble de techniques permettant aux machines d'apprendre de manière automatisée à partir de schémas et d'inductions, plutôt qu'en suivant les instructions explicites d'un humain. Les approches fondées sur l'apprentissage automatique entraînent souvent les machines à atteindre un résultat en leur présentant de nombreux exemples de résultats corrects. Toutefois, ils peuvent aussi définir un ensemble de règles et laisser la machine apprendre par essai et erreur. L'apprentissage automatique sert généralement à construire ou ajuster un modèle, mais pas seulement : il peut également être utilisé pour en interpréter les résultats (Graphique 1.6). Il fait appel à de nombreuses techniques employées depuis des décennies par des économistes, des chercheurs et des technologues. Ces techniques vont des régressions linéaires et logistiques aux arbres de décision, en passant par l'analyse en composantes principales, sans oublier les réseaux neuronaux profonds.

Graphique 1.6. Relation entre l'IA et l'apprentissage automatique



Source : Fourni par le programme *Internet Policy Research Initiative* (IPRI) du *Massachusetts Institute of Technology* (MIT).

En économie, les modèles de régression utilisent des données d'entrée pour établir des prévisions de telle sorte que les chercheurs puissent interpréter les coefficients (pondérations) appliqués aux variables d'entrée, souvent dans une optique d'action publique. Avec l'apprentissage automatique, il arrive que les humains ne soient pas en mesure de comprendre les modèles eux-mêmes. De plus, les problèmes d'apprentissage automatique tendent à recourir à un nombre bien plus important de variables que celles utilisées couramment en

économie. Ces variables, nommées « caractéristiques », se chiffrent généralement en milliers, voire plus. Des ensembles de données plus volumineux peuvent aller de dizaines de milliers à des centaines de millions d'observations. À une telle échelle, les chercheurs ont recours à des techniques plus sophistiquées et plus méconnues, comme les réseaux neuronaux, pour établir des prévisions. Il est intéressant de noter que l'un des domaines fondamentaux de recherche en matière d'apprentissage automatique tente de réintroduire le type d'explicabilité employé par les économistes dans ces modèles à grande échelle (voir Volet 4 ci-après).

La véritable technologie derrière la vague actuelle d'applications d'apprentissage automatique correspond à une technique de modélisation statistique perfectionnée, dénommée « réseaux neuronaux ». Elle s'accompagne d'une augmentation de la puissance de calcul et d'une disponibilité accrue d'ensembles de données colossaux (les « données massives »). Les réseaux neuronaux établissent des interconnexions répétées entre des milliers, voire des millions de transformations simples pour aboutir à une machine statistique plus importante, capable d'apprendre des relations élaborées entre les entrées et les sorties. En d'autres termes, ils modifient leur propre code pour trouver et optimiser les liens entre les entrées et les sorties. L'apprentissage profond désigne quant à lui des réseaux neuronaux particulièrement volumineux ; aucun seuil n'est défini pour déterminer à quel stade un réseau neuronal devient « profond ».

Cette dynamique évolutive de la recherche en matière d'IA va de pair avec des progrès constants en termes de capacités de calcul, de disponibilité des données et de conception des réseaux neuronaux. Sous leurs effets conjugués, l'approche statistique de l'IA devrait continuer de jouer un rôle important dans la recherche sur l'IA à court terme. Par conséquent, les décideurs devraient concentrer leur attention sur les évolutions de l'IA qui sont susceptibles d'avoir les incidences les plus marquées au cours des années à venir et représentent certains des défis les plus difficiles à surmonter pour les pouvoirs publics. Au nombre de ces défis figurent le décryptage des décisions des machines et le renforcement de la transparence du processus décisionnel. Les décideurs devraient également garder à l'esprit que les approches de l'IA les plus dynamiques – l'IA statistique et plus particulièrement les réseaux neuronaux – ne sont pas adaptées à tous les types de problèmes. D'autres approches, ainsi que le couplage des méthodes symbolique et statistique, restent importantes.

Il n'existe pas de taxinomie largement admise de la recherche sur l'IA ou de l'apprentissage automatique. La taxinomie exposée dans la sous-section suivante couvre 25 axes de recherche sur l'IA. Ils sont organisés en quatre grandes catégories (ou volets) et neuf sous-catégories, principalement centrées sur l'apprentissage automatique. S'il arrive que les chercheurs en économie se penchent sur un domaine de recherche restreint, les chercheurs en IA, pour leur part, travaillent généralement sur différents volets simultanément dans le but de résoudre des problématiques de recherche ouvertes.

Volet 1 : Applications d'apprentissage automatique

La première grande catégorie de recherche a trait à la mise en œuvre des méthodes d'apprentissage automatique pour résoudre diverses problématiques pratiques touchant l'économie et la société. L'émergence des applications d'apprentissage automatique est comparable à la façon dont l'accès à l'internet a commencé par transformer certains secteurs, avant de déferler sur le reste de l'économie. Le chapitre 3 expose différents exemples d'applications d'IA qui voient le jour dans les pays de l'OCDE. Les axes de recherche figurant dans le Tableau 1.1 représentent les principaux domaines de recherche liés au développement d'applications pour le monde réel.

Les domaines fondamentaux de recherche appliquée qui utilisent l'apprentissage automatique vont du traitement du langage naturel à la vision par ordinateur, en passant par la navigation robotique. Chacune de ces trois disciplines représente un champ de recherche riche et en expansion. Les problématiques de recherche peuvent être limitées à un seul domaine ou

couvrir plusieurs axes. Par exemple, aux États-Unis, des chercheurs allient, d'une part, le traitement du langage naturel pour des mammographies et des notes de pathologie en texte libre et, d'autre part, la vision par ordinateur des mammographies afin d'aider au dépistage du cancer du sein (Yala et al., 2017[12]).

Tableau 1.1. Volet 1 : Domaines d'application

Domaines d'application	Utilisation de l'apprentissage automatique	Traitement du langage naturel
		Vision par ordinateur
		Navigation robotique
		Apprentissage des langues
	Contextualisation de l'apprentissage automatique	Théorie algorithmique des jeux et choix social computationnel
		Systèmes collaboratifs

Source : Fourni par le programme IPRI du MIT.

Deux lignes de recherche étudient les moyens de contextualiser l'apprentissage automatique. La théorie algorithmique des jeux se situe à l'intersection de l'économie, de la théorie des jeux et de l'informatique. Elle utilise les algorithmes pour analyser et optimiser des jeux sur plusieurs périodes. Les systèmes collaboratifs permettent une approche des grands défis où plusieurs systèmes d'apprentissage automatique s'associent pour traiter différentes parties de problèmes complexes.

Volet 1: Intérêt pour l'action des pouvoirs publics

Plusieurs questions relevant des pouvoirs publics sont liées aux applications d'IA. Tel est le cas notamment de l'avenir du travail, des incidences potentielles de l'IA, du développement du capital humain et des compétences, ou encore de la compréhension des situations dans lesquelles le recours à des applications de l'IA pourrait ou non s'avérer adapté dans des contextes sensibles. À cela s'ajoutent les répercussions de l'IA sur les acteurs et la dynamique de l'industrie, les politiques en matière de données publiques ouvertes, la réglementation de la navigation robotique et les politiques en faveur de la protection de la vie privée qui régissent la collecte et l'utilisation des données.

Volet 2 : Techniques d'apprentissage automatique

La deuxième grande catégorie de recherche porte sur les techniques et paradigmes utilisés dans le domaine de l'apprentissage automatique. Cette ligne de recherche, similaire aux travaux de recherche sur les méthodes quantitatives dans les sciences sociales, développe et fournit les outils techniques et les approches employés dans les applications d'apprentissage automatique (Tableau 1.2).

Tableau 1.2. Volet 2 : Techniques d'apprentissage automatique

Techniques d'apprentissage automatique	Techniques	Apprentissage profond
		Apprentissage par simulation
		Production participative et calcul humain
		Calcul évolutif
		Techniques au-delà des réseaux neuronaux
	Paradigmes	Apprentissage supervisé
		Apprentissage par renforcement
		Modèles génératifs/réseaux antagonistes génératifs

Source : Fourni par le programme IPRI du MIT.

Cette catégorie est dominée par les réseaux neuronaux (dont l'apprentissage profond est une sous-catégorie) et constitue aujourd'hui le socle de la majeure partie de l'apprentissage automatique. Les techniques d'apprentissage automatique intègrent également divers paradigmes utilisés pour aider les systèmes à apprendre. L'apprentissage par renforcement entraîne le système d'une façon qui imite l'apprentissage humain, par essai et erreur. Plutôt que d'attribuer des tâches explicites aux algorithmes, ceux-ci apprennent en essayant différentes options qu'ils enchaînent à un rythme rapide. Ils s'adaptent alors en fonction des résultats, qui prennent la forme de récompenses et de pénalités. Certains parlent d'« expérimentation sans relâche » (Knight, 2017[13]).

Les modèles génératifs, notamment les réseaux antagonistes génératifs, entraînent un système à produire de nouvelles données à l'image d'un ensemble de données existant. Ces réseaux constituent un domaine de recherche sur l'IA intéressant car ils mettent en compétition au moins deux réseaux neuronaux non supervisés dans un jeu à somme nulle. En théorie des jeux, cela signifie qu'ils fonctionnent et apprennent comme une suite de jeux répétés à un rythme rapide. Les systèmes ainsi mis en compétition et dotés de vitesses de calcul élevées peuvent apprendre des stratégies utiles. Ils sont particulièrement adaptés aux environnements structurés assortis de règles claires, comme le jeu de Go, avec AlphaGo Zero.

Volet 2: Intérêt pour l'action des pouvoirs publics

Plusieurs questions relevant des pouvoirs publics sont liées au développement et au déploiement des technologies d'apprentissage automatique. Citons notamment le soutien en faveur de l'amélioration des ensembles de données d'entraînement ; le financement de la recherche universitaire et de la science fondamentale ; les politiques visant à former des personnes dotées de doubles spécialités, à savoir disposant de compétences à la fois en matière d'IA et dans une autre discipline ; et l'enseignement informatique. Par exemple, le financement de la recherche par le gouvernement canadien a permis des avancées qui ont conduit au formidable succès des réseaux neuronaux modernes (Allen, 2015[14]).

Volet 3 : Solutions d'amélioration de l'apprentissage automatique/optimisations

La troisième grande catégorie de recherche porte sur les moyens d'améliorer et d'optimiser les outils d'apprentissage automatique. Les axes de recherche y sont décomposés selon l'horizon temporel des résultats (actuels, émergents et futurs) (Tableau 1.3). La recherche à court terme est axée sur l'accélération du processus d'apprentissage profond, soit en améliorant la collecte des données, soit en utilisant des systèmes informatiques distribués pour entraîner les algorithmes.

Les chercheurs étudient comment équiper de fonctions d'apprentissage automatique des appareils à faible puissance comme des téléphones mobiles et d'autres appareils connectés. Des progrès significatifs ont été réalisés sur ce front. Des projets, à l'instar de la *Teachable Machine* de Google, offrent désormais des outils à code source libre suffisamment légers pour fonctionner avec un simple navigateur (Encadré 1.2). Le projet *Teachable Machine* n'est qu'un exemple parmi d'autres d'outils émergents de développement de l'IA destinés à étendre le rayonnement et accroître l'efficience de l'apprentissage automatique. À cela s'ajoutent des avancées significatives dans le développement de puces d'IA dédiées aux appareils mobiles.

La recherche dans l'apprentissage automatique à plus long terme porte sur l'étude des mécanismes permettant aux réseaux neuronaux d'apprendre efficacement. Bien que les réseaux neuronaux se soient révélés être une puissante technique d'apprentissage automatique, la compréhension de leur mode de fonctionnement reste limitée. Mieux appréhender ces processus permettrait de concevoir des réseaux neuronaux plus profonds. La recherche à plus long terme

s'intéresse également aux moyens d'entraîner les réseaux neuronaux à utiliser des ensembles de données d'entraînement plus réduits, une technique parfois dénommée « apprentissage à partir d'un exemple unique » (*one-shot learning*). Qui plus est, on cherche généralement à améliorer l'efficience du processus d'entraînement. De fait, les modèles de grande envergure peuvent nécessiter des semaines voire des mois d'entraînement et requièrent des centaines de millions d'exemples.

Tableau 1.3. Volet 3 : Solutions d'amélioration de l'apprentissage automatique/optimisations

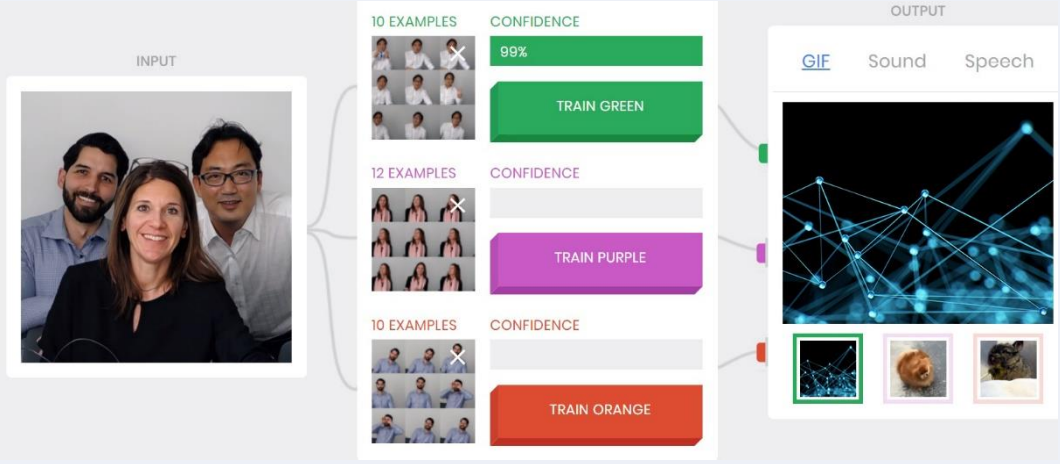
Solutions pour améliorer l'apprentissage automatique	Facteurs favorables (actuels)	Accélération de l'apprentissage profond
		Amélioration de la collecte des données
		Systèmes distribués pour l'entraînement des algorithmes
	Facteurs favorables (émergents)	Performances sur des appareils à faible puissance
		Apprendre à apprendre/méta-apprentissage
		Outils de développement de l'IA
	Facteurs favorables (futurs)	Comprendre les réseaux neuronaux
		Apprentissage à partir d'un exemple unique

Source : Fourni par le programme IPRI du MIT.

Encadré 1.2. Projet *Teachable Machine*

Le projet *Teachable Machine* est une expérience menée par Google permettant aux utilisateurs d'entraîner une machine à détecter différents scénarios à l'aide de l'appareil photo d'un téléphone ou de la caméra d'un ordinateur. Pour ce faire, l'utilisateur prend une série de photos dans trois situations distinctes (par exemple, trois expressions faciales). La machine analyse alors les photos au sein de l'ensemble de données d'entraînement et peut les utiliser pour détecter différents scénarios. Elle peut ainsi émettre un son à chaque fois que l'utilisateur sourit dans le champ de la caméra. Le projet *Teachable Machine* se démarque des autres projets d'apprentissage automatique par le fait que le réseau neuronal fonctionne exclusivement via le navigateur de l'utilisateur, sans avoir besoin de recourir à des calculs externes ni au stockage des données (Graphique 1.7).

Graphique 1.7. Entraînement d'une machine à l'aide de la caméra d'un ordinateur



Source : https://experiments.withgoogle.com/ai/teachable-machine.

Volet 3: Intérêt pour l'action des pouvoirs publics

Les questions liées au troisième volet intéressant les pouvoirs publics tiennent notamment aux incidences de l'utilisation de l'apprentissage automatique sur des appareils autonomes, sans impliquer nécessairement de partage de données dans le nuage. Autres sujets d'intérêt : les perspectives de réduction de la consommation d'énergie et la nécessité de développer des outils d'IA plus perfectionnés afin d'en optimiser les usages bénéfiques.

Volet 4 : Prise en compte du contexte sociétal

La quatrième grande catégorie de recherche examine le contexte technique, juridique et social dans lequel s'inscrit l'apprentissage automatique. Les systèmes d'apprentissage automatique s'appuient de plus en plus fréquemment sur des algorithmes pour prendre des décisions majeures. D'où l'importance de comprendre comment des biais peuvent être introduits, comment ils peuvent se propager et comment les éliminer des résultats. L'un des domaines de recherche les plus actifs en matière d'apprentissage automatique concerne la transparence et la responsabilité dans le cadre des systèmes d'IA (Tableau 1.4). Les approches statistiques de l'IA ont conduit à l'utilisation de calculs moins compréhensibles par l'homme pour la prise de décisions algorithmiques. Or ces dernières peuvent avoir des incidences non négligeables sur la vie des individus – qu'il s'agisse de prêts bancaires ou de décisions de libération conditionnelle de prisonniers (Angwin et al., 2016[15]). Les étapes visant à assurer la sécurité et l'intégrité de ces systèmes sont un autre type de recherche tenant compte du contexte. Les chercheurs commencent à peine à entrevoir de quelle façon les réseaux neuronaux parviennent à leurs décisions. Les réseaux peuvent souvent être piégés à l'aide de méthodes simples, par exemple en changeant quelques pixels sur une photo (Ilyas et al., 2018[16]). Ces axes de recherche visent à protéger les systèmes de l'introduction intempestive d'informations indésirables et d'attaques. Le but est également de vérifier l'intégrité des systèmes d'apprentissage automatique.

Volet 4: Intérêt pour l'action des pouvoirs publics

Plusieurs questions relevant des pouvoirs publics sont liées au contexte dans lequel s'inscrit l'apprentissage automatique. Citons notamment les exigences de responsabilité algorithmique, la lutte contre les biais, les incidences des systèmes d'apprentissage automatique, la sécurité des produits, la responsabilité des personnes et la sécurité des systèmes (OCDE, 2019[8]).

Tableau 1.4. Volet 4 : Affiner l'apprentissage automatique en tenant compte du contexte

Affiner l'apprentissage automatique en tenant compte du contexte	Explicabilité	Transparence et responsabilité
		Explication des décisions individuelles
		Simplification afin de tendre vers des algorithmes compréhensibles par l'homme
		Équité/biais
		Capacité de débogage
	Sécurité et fiabilité	Exemples de risques
		Vérification
		Autres classes d'attaques

Source : Fourni par le programme IPRI du MIT.

Références

Allen, K. (2015), « How a Toronto professor's research revolutionized artificial intelligence, », *The Star,* 17 April, https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html. [14]

Angwin, J. et al. (2016), « Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks », *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [15]

Anyoha, R. (28 août 2017), « The history of artificial intelligence », Harvard University Graduate School of Arts and Sciences Blog, 28 août 2017, http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/. [4]

Gringsjord, S. et N. Govindarajulu (2018), *Artificial Intelligence*, The Stanford Encyclopedia of Philosophy Archive, https://plato.stanford.edu/archives/fall2018/entries/artificial-intelligence/. [11]

Ilyas, A. et al. (2018), *Blackbox Adversarial Attacks with Limited Queries and Information*, exposé présenté à la 35e Conférence internationale sur l'apprentissage automatique, 2018, Stockholmsmässan Stockholm, du 10 au 15 juillet 2018, pp. 2142–2151. [16]

Knight, W. (2017), « 5 big predictions for artificial intelligence in 2017 », *MIT Technology Review,* 4 janvier, https://www.technologyreview.com/s/603216/5-big-predictions-for-artificial-intelligence-in-2017/. [13]

OCDE (2019), *Recommandation du Conseil sur l'intelligence artificielle*, OCDE, Paris, https://legalinstruments.oecd.org/api/print?ids=648&lang=fr. [8]

OCDE (2018), *Perspectives de l'économie numérique de l'OCDE 2017*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264282483-fr. [7]

Russel, S. et P. Norvig (2009), *Artificial Intelligence: A Modern Approach, 3ème édition*, Pearson, http://aima.cs.berkeley.edu/. [9]

Shannon, C. (1950), « XXII. Programming a computer for playing chess », *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, vol. 41/314, pp. 256-275. [2]

Silver, D. et al. (2017), « Mastering the game of Go without human knowledge », *Nature*, vol. 550/7676, pp. 354-359, http://dx.doi.org/10.1038/nature24270. [6]

Somers, J. (2013), « The man who would teach machines to think », *The Atlantic,* November, https://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/. [5]

Turing, A. (1950), « Computing machinery and intelligence », dans *Parsing the Turing Test*, Springer, Dordrecht, pp. 23-65. [1]

UW (2006), *History of AI*, University of Washington, History of Computing Course (CSEP 590A), https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf. [3]

Winston, P. (1992), *Artificial Intelligence*, Addison-Wesley, Reading, MA, https://courses.csail.mit.edu/6.034f/ai3/rest.pdf. [10]

Yala, A. et al. (2017), « Using machine learning to parse breast pathology reports », *Breast Cancer Research and Treatment*, vol. 161/2, pp. 201-211. [12]

Notes

[1] Ces tests ont été réalisés à l'aide de messages saisis ou relayés, et non par la voix.

[2] Des travaux sur le cycle de développement d'un système ont été menés, entre autres, par le *National Institute of Standards*. Plus récemment, des organisations de normalisation comme l'Organisation internationale de normalisation (ISO), par le biais de son sous-comité SC 42, ont commencé à se pencher sur le cycle de vie des systèmes d'IA.

[3] L'initiative mondiale sur l'éthique dans la conception de systèmes autonomes et intelligents (*Global Initiative on Ethics of Autonomous and Intelligent Systems*) de l'*Institute of Electrical and Electronics Engineers* (IEEE), en est un exemple.

2. Paysage économique de l'IA

Ce chapitre décrit les caractéristiques économiques de l'intelligence artificielle (IA), qui s'impose comme une technologie émergente à visée générique ouvrant la voie à une diminution du coût des prévisions et une optimisation de la prise de décisions. L'IA, qui offre un moyen de produire des prévisions, des recommandations ou des décisions plus fiables à moindre coût, promet de générer des gains de productivité, d'améliorer le bien-être et d'aider à relever des défis complexes. Le rythme d'adoption de l'IA varie selon les entreprises et les secteurs, puisque son exploitation exige de réaliser des investissements complémentaires dans les données, les compétences et la transformation numérique des flux de travail, et d'être à même d'adapter les processus organisationnels. De plus, l'IA est un domaine porteur en termes d'investissements et de développement des entreprises. Le capital-investissement dans des startups spécialisées dans l'IA s'est en effet accéléré à compter de 2016 – il a même doublé entre 2016 et 2017, pour atteindre 16 milliards USD en 2017. Les startups spécialisées dans l'IA ont attiré 12 % du capital-investissement mondial au cours du premier semestre de 2018, en nette progression par rapport à 2011 où elles ne représentaient que 3 %. Les investissements dans les technologies liées à l'IA devraient continuer d'augmenter à mesure qu'elles gagnent en maturité.

Les données statistiques concernant Israël sont fournies par et sous la responsabilité des autorités israéliennes compétentes. L'utilisation de ces données par l'OCDE est sans préjudice du statut des hauteurs du Golan, de Jérusalem-Est et des colonies de peuplement israéliennes en Cisjordanie aux termes du droit international.

Caractéristiques économiques de l'intelligence artificielle

L'intelligence artificielle facilite la production de prévisions

Du point de vue économique, les progrès récents de l'intelligence artificielle (IA) entraînent soit une réduction du coût des prévisions, soit une amélioration de leur qualité moyennant un coût stable. De nombreux aspects de la prise de décisions sont certes indépendants des prévisions. Pour autant, le recours à l'IA pour produire des prévisions de meilleure qualité, à moindre coût et largement accessibles pourrait avoir des effets transformateurs puisque les prévisions sont à la base d'une kyrielle d'activités humaines.

Alors que l'IA rend les prévisions moins onéreuses à produire, leurs usages se multiplient, comme ce fut le cas jadis avec les ordinateurs. Les premières applications de l'IA ont longtemps été reconnues comme étant dédiées aux problèmes de prévision – à l'image des techniques d'apprentissage automatique, qui permettent de prévoir les risques d'insolvabilité et d'assurance. À mesure que les coûts diminuent, certaines activités humaines s'apparentent de plus en plus à des exercices de prévision. C'est ainsi que pour établir un diagnostic médical, le médecin utilise désormais les données sur les symptômes du patient et apporte les informations manquantes sur leur cause. Le processus qui consiste à utiliser des données pour fournir des informations manquantes relève de la prévision. La classification des objets est également une question de prévision : les yeux d'une personne reçoivent les données sous forme de signaux lumineux et le cerveau complète l'information manquante en leur associant un libellé.

L'IA, en ouvrant la voie à l'établissement de prévisions à moindre coût, offre un nombre incalculable d'applications, les prévisions étant une composante essentielle du processus décisionnel. En d'autres termes, les prévisions facilitent la prise de décisions, qui est présente dans tous les domaines. Ainsi, les responsables sont appelés à prendre des décisions cruciales en termes de recrutement, d'investissement et de stratégie, et d'autres, plus triviales, sur les réunions auxquelles ils doivent assister et le rôle qu'ils doivent y jouer. Les juges prennent des décisions importantes quant à la culpabilité ou l'innocence de prévenus, aux procédures et aux peines à prononcer, et d'autres, qui le sont beaucoup moins, sur un paragraphe ou une requête spécifique. Enfin, les individus prennent en permanence des décisions – du menu du dîner à une demande en mariage, en passant par la musique qu'ils souhaitent écouter. L'enjeu phare de la prise de décisions tient à la gestion de l'incertitude. Dans la mesure où la prévision réduit l'incertitude, elle nourrit toutes ces décisions et peut ainsi ouvrir le champ des possibles.

La prévision générée par la machine est un substitut de la prévision humaine

Autre notion d'ordre économique : la substitution. Lorsque le prix d'un produit de base (à l'instar du café) chute, non seulement les individus en achètent davantage, mais ils délaissent également les produits de substitution (comme le thé). De la même façon, si les machines sont capables d'établir des prévisions à moindre coût, elles se substitueront à l'homme pour l'exécution de ce type de tâches. La réduction de la main-d'œuvre dans ce domaine deviendra dès lors une conséquence majeure de l'IA sur le travail humain.

Tout comme l'avènement des ordinateurs signifie qu'aujourd'hui, rares sont les travailleurs qui exécutent des opérations arithmétiques dans le cadre professionnel, l'IA aura le même effet sur les tâches de prévision. Par exemple, la transcription – à savoir la conversion d'un discours oral en texte – s'apparente à de la prévision en ce qu'elle consiste à trouver les informations manquantes sur les symboles écrits correspondant aux paroles prononcées. L'IA donne d'ores et déjà des résultats plus rapides et plus fiables que de nombreuses personnes dont le travail implique des tâches de transcription.

Données, actions et jugement complètent les prévisions des machines

Lorsque le prix d'un produit de base (le café) diminue, les individus tendent à acheter en plus grande quantité les produits complémentaires (le lait et le sucre, par exemple). L'identification des « produits complémentaires » de la prévision constitue par conséquent un enjeu de taille compte tenu des progrès récents de l'IA. Si la prévision est une composante essentielle de la prise de décisions, elle ne fait pas tout. Les autres aspects d'une décision sont des compléments de l'IA, qu'il s'agisse des données, des actions ou du jugement.

Les **données** désignent l'information qui vient alimenter une prévision. Nombre des avancées récentes de l'IA dépendent de volumes considérables de données numériques dont les systèmes d'IA ont besoin pour établir des prévisions à partir d'exemples passés. En règle générale, plus ces exemples sont nombreux, plus les prévisions sont fiables. Par conséquent, grâce à l'IA, l'actif que représente l'accès à de vastes quantités de données renferme davantage de valeur pour les organisations. La valeur stratégique des données s'avère toutefois difficile à appréhender, puisqu'elle dépend de deux types de considérations : dans quelle mesure les données vont-elles aider une organisation à prévoir des éléments qui lui sont importants, et l'organisation a-t-elle à sa disposition uniquement des données rétrospectives ou est-elle en mesure de les enrichir avec des données collectées au fil du temps ? L'aptitude à poursuivre l'apprentissage grâce à de nouvelles données peut dès lors être une source d'avantage concurrentiel durable (Agrawal, Gans et Goldfarb, 2018[1]).

Des tâches nouvelles peuvent être menées à bien grâce aux autres éléments de décision : les **actions** et le **jugement**. Certaines *actions* revêtent par nature davantage de valeur lorsqu'elles sont exécutées par des humains plutôt que par une machine (qu'il s'agisse d'athlètes de haut niveau, de professionnels de la petite enfance, ou de commerciaux et vendeurs). Néanmoins, l'aspect le plus important reste probablement le *jugement*, à savoir le processus de détermination de l'intérêt d'une action particulière dans un environnement donné. Lorsque l'on recourt à l'IA pour établir des prévisions, un humain doit décider de ce que l'on va prévoir et de l'usage qui en sera fait.

La mise en œuvre de l'IA dans les organisations nécessite de réaliser des investissements complémentaires et d'adapter les processus

Tout comme l'informatique, l'électricité ou les moteurs à vapeur, l'IA peut être considérée comme une technologie générique (Bresnahan et Trajtenberg, 1992[2] ; Brynjolfsson, Rock et Syverson, 2017[3]). Ce qui signifie qu'elle est à même de conduire à des gains de productivité notables dans un éventail plus large de secteurs. Dans le même temps, son déploiement nécessite des investissements dans un certain nombre de facteurs complémentaires – et peut amener les organisations à repenser leur stratégie globale.

Pour que le recours à l'IA soit synonyme de gains de productivité significatifs, les organisations doivent réaliser des investissements complémentaires. Ceux-ci portent sur l'infrastructure de collecte de données en continu, le recrutement de spécialistes capables d'exploiter les données, ou encore l'adaptation des processus afin de mettre à profit les nouvelles possibilités offertes par la réduction de l'incertitude.

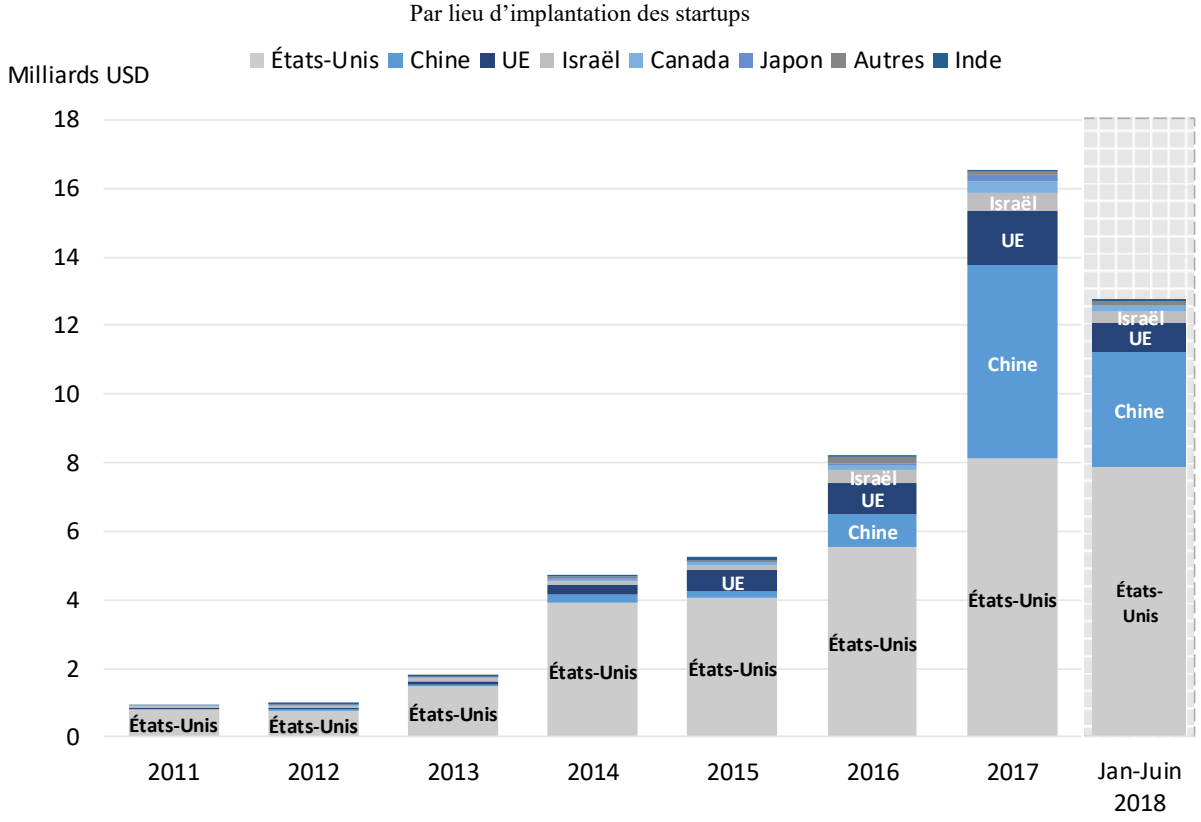
Chaque organisation dispose de nombreux processus dont le but est plus de tirer le meilleur parti de situations données face à l'incertitude, que d'offrir le meilleur service possible aux clients. Les salles d'attente des aéroports, par exemple, offrent aux passagers un espace confortable où patienter avant le départ de leur avion. Or si les passagers disposaient de prévisions fiables sur la durée du trajet jusqu'à l'aéroport et le temps nécessaire aux contrôles de sécurité, ils n'auraient peut-être plus besoin de ces salles d'attente.

Le champ des opportunités qu'offriraient de meilleures prévisions devrait varier selon les entreprises et les secteurs. Google, Baidu et d'autres entreprises exploitant de vastes plateformes électroniques sont bien placées pour tirer avantage d'investissements de grande ampleur dans l'IA. Du côté de l'offre, ils disposent d'ores et déjà de systèmes pour la collecte des données. Pour ce qui est de la demande, ils commencent à se constituer une clientèle suffisante pour justifier les coûts fixes élevés des investissements dans l'IA. Pour autant, nombreuses sont les entreprises qui n'ont pas intégralement informatisé leurs flux de travail et ne sont pas encore en mesure d'appliquer directement les outils d'IA à leurs processus existants. Toutefois, avec la diminution des coûts, elles prendront peu à peu conscience des opportunités que peut leur conférer la réduction de l'incertitude. Motivées par la quête de solutions pour satisfaire leurs besoins, elles suivront les traces des entreprises pionnières et investiront dans l'IA.

Capital-investissement dans les startups spécialisées dans l'IA

Les investissements dans l'IA progressent à un rythme soutenu, de sorte que l'IA produit d'ores et déjà des effets notables sur les entreprises. MGI (2017[4]) estime ainsi qu'en 2016, 26 à 39 milliards USD ont été investis dans l'IA à l'échelle mondiale. Environ 70 % de ces investissements étaient réalisés en interne ; 20 % environ correspondaient à des investissements dans des startups spécialisées dans l'IA, et autour de 10 % à des acquisitions d'entreprises d'IA (Dilda, 2017[5]). Les trois quarts de ces investissements sont le fait des grandes entreprises de technologie.

Graphique 2.1. Investissements totaux estimés dans des startups spécialisées dans l'IA, 2011-17 et premier semestre de 2018



Note : Les estimations pour 2018 pourraient être sous-évaluées, certaines données pouvant être manquantes du fait des délais de déclaration (voir Encadré 2.1. Note méthodologique).
Source : Estimations de l'OCDE d'après Crunchbase (juillet 2018), www.crunchbase.com.

En dehors du secteur des technologies, l'adoption de l'IA n'en est qu'à ses prémices et rares sont les entreprises qui ont déployé des solutions d'IA à grande échelle. Les acteurs majeurs des autres secteurs affichant une maturité numérique avancée et disposant de données à exploiter, à l'image de la finance et de l'automobile, se tournent eux aussi peu à peu vers l'IA.

Les géants des technologies multiplient les acquisitions de startups spécialisées dans l'IA. Selon CBI (2018[6]), Google, Apple, Baidu, Facebook, Amazon, Intel, Microsoft, Twitter et Salesforce sont les entreprises qui ont réalisé le plus d'acquisitions de ce type depuis 2010. Par ailleurs, plusieurs startups spécialisées dans l'IA appliquée à la cybersécurité ont été rachetées en 2017 et début 2018. Tel est le cas de Sqrrl et de Zenedge, acquises respectivement par Amazon et Oracle.

Les startups spécialisées dans l'IA sont également une cible privilégiée pour les entreprises opérant dans des secteurs plus traditionnels, comme l'automobile, la santé – avec par exemple Roche Holding ou Athena Health –, l'assurance et la vente au détail.

Après cinq années de croissance ininterrompue, le capital-investissement dans des startups spécialisées dans l'IA s'est accéléré à compter de 2016. Le volume des investissements a même doublé entre 2016 et 2017 (Graphique 2.1). On estime que plus de 50 milliards USD ont été investis dans ce type de startups entre 2011 et mi-2018 (Encadré 2.1).

Encadré 2.1. Note méthodologique

On expose dans cette section des estimations du capital investi dans des startups spécialisées dans l'IA, d'après les données de la plateforme Crunchbase (version de juillet 2018). Créée en 2007, Crunchbase est une base de données commerciales sur des entreprises innovantes ; elle contient des informations sur plus de 500 000 entités implantées dans 199 pays. Breschi, Lassébie et Menon (2018[7]) proposent une analyse comparative de Crunchbase et d'autres sources de données agrégées. Ils observent des schémas concordants pour un large éventail de pays, notamment la plupart des membres de l'OCDE (à l'exception du Japon et de la Corée). On retrouve également des schémas homogènes en Afrique du Sud, au Brésil, en Fédération de Russie, en Inde et en République populaire de Chine (ci-après dénommée la « Chine »). Crunchbase classe les entreprises dans un ou plusieurs domaine(s) technologique(s) pris dans une liste de 45 groupes.

Il convient toutefois de rester prudent lorsque l'on utilise Crunchbase et ce, pour des questions liées au champ extrêmement large de la base de données, à la fiabilité des informations autodéclarées et à la sélection de l'échantillon. En particulier, l'enregistrement des nouvelles opérations dans la base de données peut prendre du temps et les délais peuvent varier selon les pays. En outre, il se peut que les startups aient tendance à s'autodéclarer en tant que startups spécialisées dans l'IA du fait de l'intérêt croissant des investisseurs pour cette catégorie.

Aux fins de la présente étude, les « startups spécialisées dans l'IA » correspondent aux entreprises fondées après 2000, référencées dans le domaine technologique « artificial intelligence » (intelligence artificielle) de la base Crunchbase (soit 2 436 entreprises). Elles englobent également les entreprises ayant indiqué des mots-clés liés à l'IA dans la description courte de leurs activités (ce qui représente 689 entreprises supplémentaires). Trois types de mots-clés sont considérés comme étant liés à l'IA : premièrement, les mots-clés génériques de l'IA, à savoir « artificial intelligence » (intelligence artificielle), « AI » (IA), « machine learning » (apprentissage automatique) et « machine intelligence » (intelligence des machines) ; deuxièmement, les mots-clés liés aux techniques d'IA, tels « neural

network » (réseau neuronal), « deep learning » (apprentissage profond) et « reinforcement learning » (apprentissage par renforcement). Enfin, le troisième type a trait aux champs d'application de l'IA ; on y retrouve des mots-clés comme « computer vision » (vision par ordinateur), « predictive analytics » (analyse prédictive), « natural language processing » (traitement du langage naturel), « autonomous vehicles » (véhicule autonome), « intelligent system » (système intelligent) et « virtual assistant » (assistant virtuel).

Plus d'un quart (26 %) des opérations d'investissement dans des startups spécialisées dans l'IA référencées dans la base de données ne font pas état des investissements des apporteurs de capital-risque. Aux fins de la présente analyse, on a estimé les montants de ces opérations sur la base de la valeur moyenne des opérations de plus petite envergure (en tenant compte uniquement des opérations de moins de 10 millions USD) pour la même période et dans le même pays. La raison pour laquelle les opérations plus importantes sont exclues tient au fait que les montants correspondants sont généralement des informations publiques. La valeur des opérations non divulguées au public est estimée à environ 6 % de la valeur totale des investissements réalisés entre 2011 et mi-2018, un taux qui pourrait toutefois être sous-évalué. Les chiffres du premier semestre de 2018 sont partiels, les opérations n'étant pas déclarées immédiatement.

L'IA représente aujourd'hui plus de 12 % du capital-investissement dans les startups

Les startups spécialisées dans l'IA ont attiré environ 12 % de l'ensemble du capital-investissement mondial au cours du premier semestre de 2018, en nette progression par rapport à 2011, où elles représentaient seulement 3 % (Graphique 2.2). La part des investissements dans les startups spécialisées dans l'IA a augmenté dans tous les pays analysés. Au premier semestre de 2018, quelque 13 % des investissements dans des startups aux États-Unis et en Chine visaient des entreprises spécialisées dans l'IA. Surtout, Israël a vu la part des investissements dans ce type d'entreprises bondir de 5 % à 25 % entre 2011 et le premier semestre de 2018 ; les véhicules autonomes ont capté 50 % des investissements en 2017.

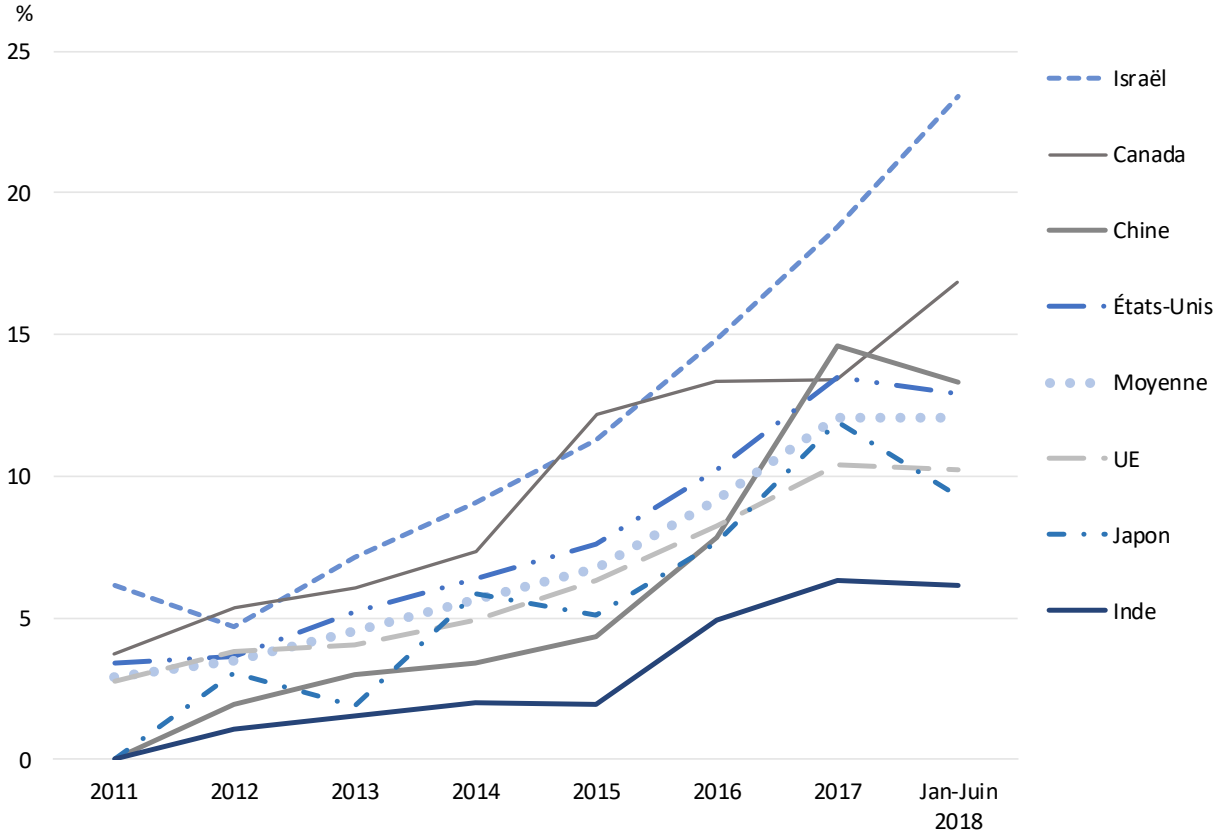
Les États-Unis et la Chine concentrent la majeure partie des investissements dans des startups spécialisées dans l'IA

Les startups implantées aux États-Unis captent la majeure partie du capital-investissement mondial dans les startups spécialisées dans l'IA. Ce constat vaut à la fois pour le nombre d'opérations d'investissement et pour les montants investis, qui comptent pour les deux tiers de la valeur totale investie depuis 2011 (Graphique 2.1). Rien de surprenant à cela, si l'on considère que les États-Unis représentent 70 à 80 % des investissements mondiaux de capital-risque, toutes technologies confondues (Breschi, Lassébie et Menon, 2018[7]).

En Chine, l'investissement dans les startups spécialisées dans l'IA connaît un essor spectaculaire depuis 2016. À tel point que la Chine s'est hissée au deuxième rang mondial en termes de valeur du capital-investissement dans l'IA. Les entreprises chinoises ont attiré 36 % du capital-investissement mondial dans l'IA en 2017, contre seulement 3 % en 2015, la moyenne s'établissant à 21 % de 2011 à mi-2018.

Graphique 2.2. Part de l'IA dans le capital investi dans des startups, 2011 à 2017 et premier semestre de 2018

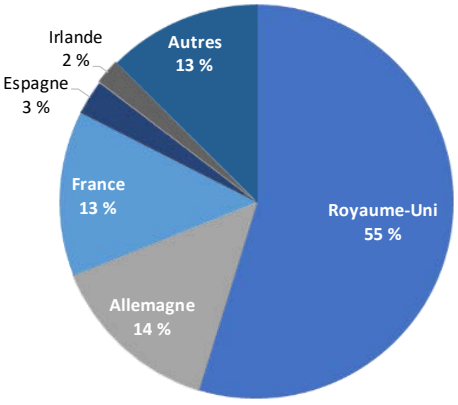
En pourcentage du nombre total d'opérations d'investissement



Note : Le pourcentage pour 2018 couvre uniquement le premier semestre de l'année (voir Encadré 2.1. Note méthodologique).
Source : Estimations de l'OCDE d'après Crunchbase (juillet 2018), www.crunchbase.com.

Graphique 2.3. Capital-investissement dans des startups spécialisées dans l'IA implantées dans l'Union européenne, 2011 à mi-2018

En pourcentage des montants totaux investis dans des startups implantées dans l'UE au cours de la période considérée



Note : Le pourcentage pour 2018 couvre uniquement le premier semestre de l'année.
Source : Estimations de l'OCDE d'après Crunchbase (juillet 2018), www.crunchbase.com.

L'Union européenne a attiré 8 % du capital-investissement mondial dans l'IA en 2017. Ce qui représente une forte hausse pour la région prise dans son ensemble, puisqu'elle affichait un taux de seulement 1 % en 2013. En revanche, les volumes d'investissement varient sensiblement selon les États membres. Les startups implantées au Royaume-Uni ont capté 55 % de l'investissement total observé dans l'Union européenne entre 2011 et mi-2018, suivies par les jeunes pousses allemandes (14 %) et françaises (13 %). Par conséquent, les 25 pays restants se sont partagé moins de 20 % du capital-investissement total dans l'IA reçu dans l'Union européenne (Graphique 2.3).

Les États-Unis, la Chine et l'Union européenne représentent à eux trois plus de 93 % du capital-investissement dans l'IA totalisé entre 2011 et mi-2018. Au-delà de ce peloton de tête, il convient de souligner également les taux observés en Israël (3 %) et au Canada (1.6 %).

Les opérations dans le domaine de l'IA ont augmenté jusqu'en 2017, non seulement en nombre, mais aussi en taille

Le nombre d'opérations d'investissement a augmenté à l'échelle mondiale, passant de moins de 200 à plus de 1 400 transactions au cours de la période 2011-17. Cela équivaut à un taux de croissance annuel composé de 35 % entre 2011 et le premier semestre de 2018 (Graphique 2.4). Les startups implantées aux États-Unis ont attiré une part non négligeable des opérations, qui ont bondi de 130 à environ 800 transactions sur la période 2011-17. Même constat dans l'Union européenne, où le nombre d'opérations a progressé de 30 à 350 environ pendant la période considérée.

Les startups basées en Chine ont conclu un nombre moins élevé de transactions que celles opérant aux États-Unis ou dans l'Union européenne, passant de 0 à environ 60 entre 2011 et 2017. En revanche, la valeur totale élevée des investissements réalisés en Chine signifie que la valeur moyenne de ces opérations était bien supérieure à celle observée dans l'Union européenne.

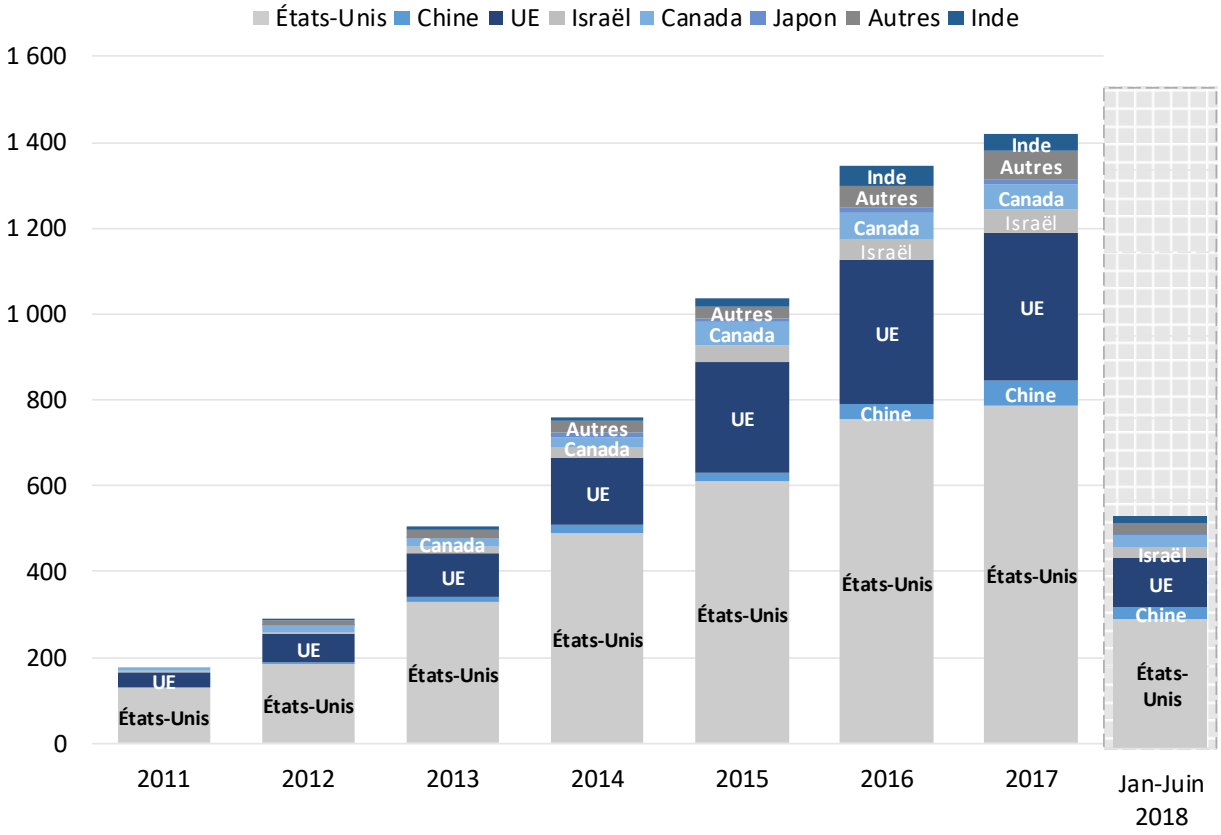
La valeur moyenne importante des investissements observés en Chine s'inscrit dans une tendance générale d'augmentation de la valeur par transaction. En 2012 et 2013, près de neuf opérations d'investissement déclarées sur dix portaient sur moins de 10 millions USD. Seule une sur dix se situait dans une fourchette allant de 10 à 100 millions USD, et aucune ne dépassait 100 millions USD. En 2017, plus de deux opérations sur dix portaient sur un montant supérieur à 10 millions USD et près de 3 % dépassaient 100 millions USD. Cette tendance s'est accentuée au premier semestre de 2018, 40 % des opérations déclarées franchissant la barre des 10 millions USD et 4.4 % celle des 100 millions USD.

En termes de valeur, les « méga-opérations » (portant sur un montant supérieur à 100 millions USD) représentaient 66 % des montants totaux investis dans des startups spécialisées dans l'IA au premier semestre de 2018. Ces chiffres reflètent le niveau de maturité croissant des technologies de l'IA, ainsi qu'une évolution des stratégies des investisseurs, qui tendent à privilégier des investissements de plus grande envergure dans un nombre réduit d'entreprises spécialisées dans l'IA – à l'image de la startup chinoise Toutiao, qui a attiré en 2017 l'investissement le plus élevé (d'une valeur de 3 milliards USD). L'entreprise a mis au point un système de recommandation de contenu fondé sur l'IA, qui s'appuie sur l'exploration de données pour proposer des informations pertinentes et personnalisées aux utilisateurs chinois d'après une analyse des réseaux sociaux.

Depuis 2016, Israël (Voyager Labs), la Suisse (Mindmaze), le Canada (LeddarTech et Element AI) et le Royaume-Uni (Oaknorth et Benevolent AI) sont autant de pays qui ont vu se conclure des opérations de 100 millions USD ou plus. Ce qui témoigne du dynamisme des activités autour de l'IA au-delà des États-Unis et de la Chine.

Graphique 2.4. Nombre d'opérations de capital-investissement dans des startups spécialisées dans l'IA, par lieu d'implantation des startups

2011-17 et premier semestre de 2018



Note : Les estimations pour 2018 pourraient être sous-évaluées, certaines données pouvant être manquantes du fait des délais de déclaration (voir Encadré 2.1. Note méthodologique).
Source : Estimations de l'OCDE d'après Crunchbase (juillet 2018), www.crunchbase.com.

Les schémas d'investissement varient selon les pays et régions

Si l'on observe depuis 2011 une augmentation notable des montants totaux investis et du nombre d'opérations à l'échelle mondiale, de fortes disparités demeurent dans les schémas d'investissement selon les pays et régions.

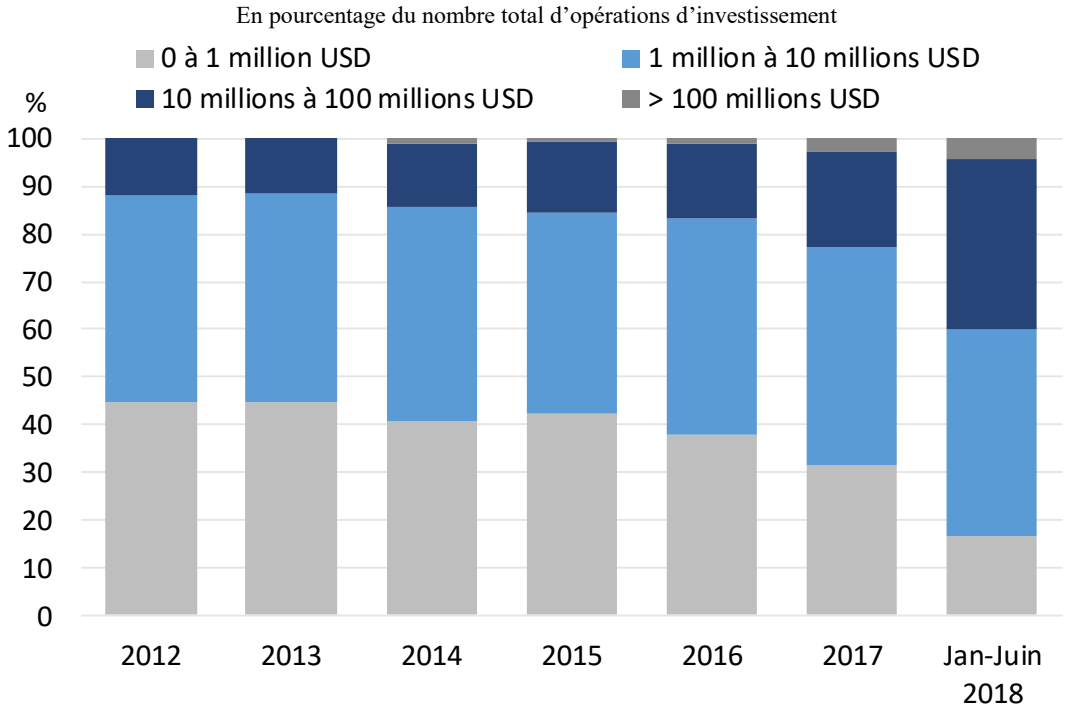
Surtout, le profil des investissements réalisés dans des startups chinoises diffère de celui observé dans le reste du monde. Les opérations de capital-investissement dans des startups chinoises spécialisées dans l'IA, enregistrées dans la base Crunchbase, représentaient une enveloppe moyenne de 150 millions USD en 2017 et au premier semestre de 2018. À titre de comparaison, la valeur moyenne des investissements réalisés en 2017 dans les autres pays atteignait à peine un dixième de ce montant.

Trois grands schémas d'investissement se dessinent. Premièrement, en Chine, les opérations portent sur un nombre réduit de startups mais impliquent des montants élevés. Deuxièmement, les startups implantées dans l'UE attirent un nombre sans cesse croissant d'opérations de plus petite envergure. La valeur moyenne par opération est passée de 3.2 millions USD en 2016 à 5.5 millions USD en 2017, puis 8.5 millions USD au premier semestre de 2018. Troisièmement, les États-Unis se caractérisent par un nombre sans cesse croissant d'investissements de plus

grande envergure. La valeur moyenne par opération y est passée de 9.5 millions USD en 2016 à 13.2 millions USD en 2017, pour atteindre 32 millions USD au premier semestre de 2018. Ces différentes tendances restent notables même si l'on exclut de l'échantillon les opérations d'une valeur supérieure à 100 millions USD (Tableau 2.1 et Tableau 2.2).

Autre constat, ces schémas d'investissement, loin de se limiter aux startups spécialisées dans l'IA, valent également pour les autres secteurs. En 2017, les startups chinoises, tous secteurs confondus, ont levé en moyenne 200 millions USD par cycle d'investissement. Dans le même temps, celles implantées aux États-Unis et dans l'Union européenne ont levé en moyenne respectivement 22 millions USD et 10 millions USD.

Graphique 2.5. Taille des opérations d'investissement, 2012-17 et premier semestre de 2018



Note : Les pourcentages pour 2018 couvrent uniquement le premier semestre de l'année.
Source : Estimations de l'OCDE d'après Crunchbase (juillet 2018), www.crunchbase.com.

Tableau 2.1. Montants moyens levés par opération d'investissement, pour les opérations d'une valeur allant jusqu'à 100 millions USD

En millions USD

	Canada	Chine	UE	Israël	Japon	États-Unis
2015	2	12	2	4	4	6
2016	4	20	3	6	5	6
2017	2	26	4	12	14	8

Source : Estimations de l'OCDE d'après Crunchbase (avril 2018), www.crunchbase.com.

Tableau 2.2. Montants moyens levés par opération d'investissement, pour l'ensemble des opérations réalisées dans le domaine de l'IA

En millions USD

	Canada	Chine	UE	Israël	Japon	États-Unis
2015	2	12	3	4	4	8
2016	4	73	3	6	5	10
2017	8	147	6	12	14	14

Source : Estimations de l'OCDE d'après Crunchbase (avril 2018), www.crunchbase.com.

Les startups spécialisées dans les véhicules autonomes attirent d'importants investissements

Les volumes de capital-investissement dans l'IA varient considérablement selon les domaines d'application. Les véhicules autonomes attirent une part croissante du capital-investissement dans les startups spécialisées dans l'IA. Jusqu'en 2015, ils captaient moins de 5 % des investissements totaux dans des startups de ce type. En 2017, leur part était passée à 23 %, pour atteindre 30 % mi-2018. La majeure partie des investissements de capital-risque dans les startups spécialisées dans les véhicules autonomes ont été destinés à des entreprises implantées aux États-Unis (80 % entre 2017 et mi-2018). Suivaient les startups basées en Chine (15 %), en Israël (3 %) et dans l'Union européenne (2 %). La progression s'explique par une hausse notable des montants par opération d'investissement, le nombre réel d'opérations étant resté relativement stable (87 en 2016 et 95 en 2017). Aux États-Unis, le montant moyen par opération d'investissement dans le secteur des véhicules autonomes a été multiplié par dix, passant de 20 millions USD à près de 200 millions USD entre 2016 et le premier semestre de 2018. Cette progression est essentiellement due à l'investissement de 3.35 milliards USD de Softbank dans Cruise Automation. Cette société, division du groupe General Motors spécialisée dans les véhicules autonomes, développe des systèmes de conduite automatisée pour des véhicules existants. En 2017, Ford a, pour sa part, investi 1 milliard USD dans la startup de véhicules autonomes Argo AI.

Tendances plus larges en matière de développement et de diffusion de l'IA

Les efforts déployés actuellement pour définir des mesures empiriques de l'IA se heurtent à un certain nombre de problématiques, notamment à l'absence de définitions claires. Or celles-ci constituent une condition indispensable pour compiler des mesures fiables et comparables. Des travaux expérimentaux menés conjointement par l'OCDE et l'Institut Max Planck (MPI) pour l'innovation et la concurrence ont débouché sur l'élaboration d'une approche axée sur trois volets en vue de mesurer i) les évolutions de l'IA dans la science, en se fondant sur les publications scientifiques ; ii) les évolutions technologiques de l'IA, en s'appuyant sur une mesure indirecte : les brevets ; et iii) les évolutions dans le domaine des logiciels d'IA, en particulier des logiciels libres. Cette approche implique de faire appel aux conseils d'experts pour identifier les ressources (publications, brevets et logiciels) explicitement liées à l'IA. Celles-ci sont ensuite utilisées comme référence pour évaluer le degré de corrélation à l'IA d'autres ressources (Baruffaldi et al., à paraître[8]).

Les publications scientifiques servent depuis longtemps à la mesure indirecte des résultats des efforts de recherche et des progrès de la science. L'OCDE utilise des données bibliométriques issues de Scopus, grande base de données de citations et de résumés provenant de la documentation examinée par les pairs (notamment des actes de conférences). Ces derniers constituent une ressource particulièrement utile dans le cas des domaines émergents comme

l'IA. De fait, ils donnent un aperçu immédiat des nouveautés présentées dans les actes de conférences examinés par les pairs, avant publication des travaux. L'établissement d'une liste de mots-clés liés à l'IA et leur validation avec des experts en IA permettent de repérer les ressources ayant trait à l'IA dans n'importe quel domaine scientifique.

L'approche fondée sur les brevets, mise au point par l'OCDE et la branche du MPI travaillant sur les brevets, vise à identifier et cartographier les inventions liées à l'IA et d'autres évolutions technologiques intégrant des composantes d'IA, quel que soit le domaine technologique. Elle s'appuie pour ce faire sur diverses méthodes en vue de recenser les inventions, notamment des recherches par mots-clés dans les résumés ou les demandes de brevets ; l'analyse des portefeuilles de brevets des startups spécialisées dans l'IA ; et l'analyse des brevets citant des ressources scientifiques liées à l'IA. Cette approche a été affinée à la lumière de travaux menés sous l'égide du Groupe de réflexion sur les statistiques de propriété intellectuelle, piloté par l'OCDE[1].

Les données issues de GitHub, la plus grande plateforme d'hébergement de logiciels à code source libre, sont utilisées afin de repérer les évolutions en matière d'IA. Les codes afférents à l'IA sont classés sous différents thèmes grâce à une analyse par modélisation thématique faisant apparaître les grands domaines de l'IA. Les domaines génériques couvrent l'apprentissage automatique (y compris l'apprentissage profond), les statistiques, les mathématiques et les méthodes computationnelles. Les domaines et applications spécifiques comprennent l'exploration de texte, la reconnaissance d'image ou la biologie.

Références

Agrawal, A., J. Gans et A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business School Press. [1]

Baruffaldi, S. et al. (à paraître), « Identifying and measuring developments in artificial intelligence », *OECD Science, Technology and Industry Working Papers*, Éditions OCDE, Paris. [8]

Breschi, S., J. Lassébie et C. Menon (2018), « A portrait of innovative start-ups across countries », *OECD Science, Technology and Industry Working Papers*, n° 2018/2, Éditions OCDE, Paris, https://dx.doi.org/10.1787/f9ff02f4-en. [7]

Bresnahan, T. et M. Trajtenberg (1992), « General purpose technologies: « Engines of growth? » », *document de travail*, n° 4148, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w4148. [2]

Brynjolfsson, E., D. Rock et T. Syverson (2017), « Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics », *National Bureau of Economic Research*, vol. 24001, http://dx.doi.org/10.3386/w24001. [3]

CBI (2018), *The Race For AI: Google, Intel, Apple In A Rush To Grab Artificial Intelligence Startups*, CBI Insights, 27 février 2018, https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/. [6]

Dilda, V. (2017), *AI: Perspectives and Opportunities*, exposé présenté à la conférence AI: Intelligent Machines, Smart Policies, Paris, les 26 et 27 octobre 2017, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-dilda.pdf. [5]

MGI (2017), *Artificial intelligence: The next digital frontier?*, McKinsey Global Institute, juin 2017. [4]

Note

[1] Ces travaux ont bénéficié des conseils d'experts et d'examinateurs de brevets des offices de propriété intellectuelle de l'Australie et du Canada, de l'Office européen des brevets, de l'Office israélien des brevets, de l'office italien des brevets et des marques, de l'Institut national de la propriété industrielle du Chili, de l'Office britannique de la propriété intellectuelle et de l'Office américain des brevets.

3. Applications de l'intelligence artificielle

Le présent chapitre décrit les nouvelles possibilités offertes dans plusieurs secteurs où les technologies de l'intelligence artificielle connaissent une percée rapide, à savoir les transports, l'agriculture, la finance, le marketing et la publicité, la science, les soins de santé, la justice pénale, la sécurité, le secteur public et les applications de réalité augmentée et de réalité virtuelle. Les systèmes d'IA mis au point dans ces domaines peuvent détecter des schémas dans de gigantesques volumes de données et modéliser des systèmes complexes interdépendants pour produire des résultats synonymes d'amélioration de l'efficience de la prise de décision, de réduction des coûts et d'optimisation des ressources. La section relative à l'IA dans les transports a été préparée par l'Internet Policy Research Initiative du Massachusetts Institute of Technology. Plusieurs autres sections ont été rédigées sur la base de travaux de l'OCDE, notamment ceux du Comité de la politique de l'économie numérique et de son Groupe de travail sur la sécurité et la vie privée dans l'économie numérique, du Comité de la politique scientifique et technologique, de l'initiative e-Leaders du Comité de la gouvernance publique, ainsi que du Comité de la politique à l'égard des consommateurs et de son Groupe de travail sur la sécurité des produits de consommation.

L'IA dans le secteur des transports avec les véhicules autonomes

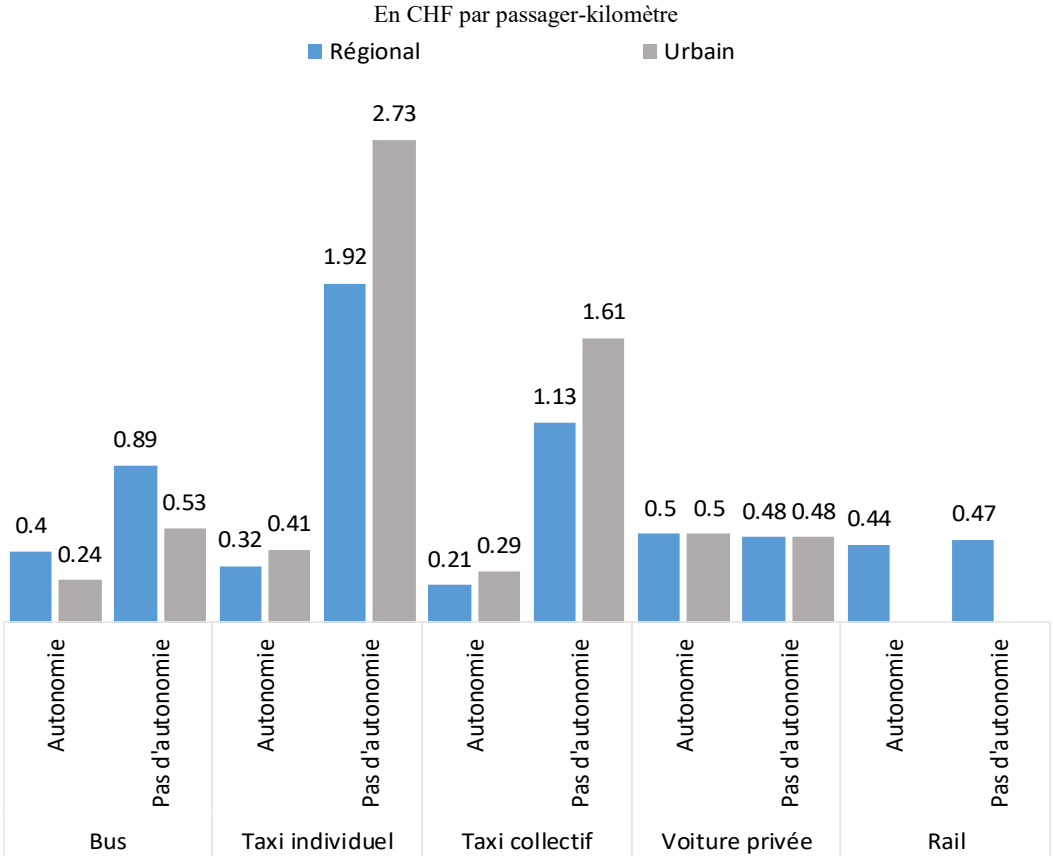
Les systèmes d'IA font leur entrée dans tous les secteurs de l'économie. Mais c'est dans celui des transports qu'est en train de se produire l'un des plus grands changements de paradigme, avec la transition vers les véhicules autonomes.

Impacts économiques et sociaux des véhicules autonomes

Le secteur des transports est l'un des plus importants de la zone OCDE : en 2016, il a totalisé 5.6 % de son produit intérieur brut (OCDE, 2018[1])[1]. Le déploiement des véhicules autonomes pourrait avoir un impact économique majeur en réduisant le nombre d'accidents et les problèmes de congestion et en générant d'autres bénéfices. On estime qu'un taux d'adoption de 10 % de véhicules autonomes aux États-Unis permettrait de sauver 1 100 vies et d'économiser 38 milliards USD par an. À un taux de 90 %, on atteindrait 21 700 vies sauvées et 447 milliards USD d'économies par an (Fagnant et Kockelman, 2015[2]).

Des travaux de recherche plus récents calculent, pour plusieurs modes de transport en Suisse, des différences significatives de coût par kilomètre avec ou sans automatisation des véhicules (Bösch et al., 2018[3]). Les services de taxi économiseraient le plus. Les propriétaires de voitures particulières bénéficieraient d'une moindre baisse des coûts (Graphique 3.1). Les gains des premiers découleraient bien sûr principalement de l'élimination des salaires des chauffeurs.

Graphique 3.1. Coûts avec ou sans automatisation des véhicules pour plusieurs modes de transport



Source : D'après Bösch et al. (2018[3]), « Cost-based analysis of autonomous mobility services », https://doi.org/10.1016/j.tranpol.2017.09.005.

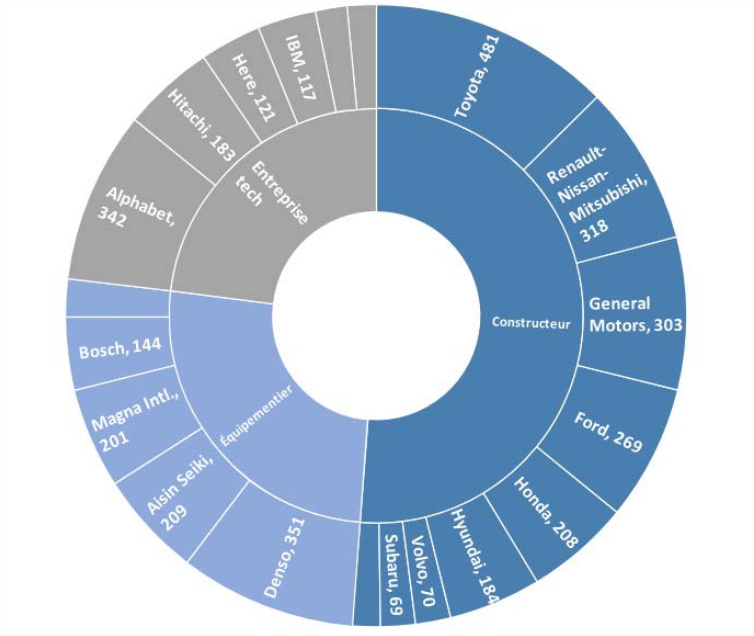
Évolution du marché

Le marché des transports est actuellement très fluctuant sous l'effet de trois récentes évolutions majeures : la mise au point des véhicules autonomes, l'essor des services de partage de véhicule ou de trajet (autopartage, covoiturage, transport avec chauffeur, etc.), et la transition au profit des véhicules électriques. Deux tendances expliquent que les constructeurs automobiles historiques peinent à redéfinir leurs stratégies. Premièrement, les services de partage de véhicule ou de trajet sont une solution de transport que les usagers, en particulier les plus jeunes, jugent de plus en plus viable. Deuxièmement, il n'est pas sûr que le principe traditionnel de la propriété privée d'un véhicule continue d'exister à long terme. Les constructeurs de modèles haut de gamme expérimentent déjà de nouveaux modèles économiques fondés par exemple sur des services d'abonnement : en contrepartie d'une redevance fixe mensuelle, la clientèle des programmes « Access by BMW », « Mercedes Collection » ou « Porsche Passport » peut se procurer un véhicule ou en changer quand elle le souhaite.

Les entreprises technologiques, des multinationales aux startups, se lancent dans la conception de systèmes de véhicules autonomes, de solutions de partage de véhicule ou de trajet, ou de véhicules électriques – ou une combinaison des trois. Selon Morgan Stanley, la valorisation de la division Waymo d'Alphabet pourrait atteindre 175 milliards USD à elle seule, du fait principalement de son potentiel dans le secteur des services de livraison et de transport routier autonome de marchandises (Ohnsman, 2018[4]). De son côté, la récente startup Zoox, spécialisée dans les systèmes d'IA destinés à la conduite en zone urbaine dense, a pu lever 790 millions USD, ce qui situe actuellement sa valorisation à 3.2 milliards USD[2] avant même la production de revenus (voir chapitre 2, section « Capital-investissement dans les startups spécialisées dans l'IA »). Ces démarches des entreprises technologiques viennent s'ajouter aux investissements des constructeurs et équipementiers historiques dans les applications automobiles des technologies de l'IA.

Graphique 3.2. Dépôts de brevets relatifs aux véhicules autonomes, par entreprise, 2011-16

Entreprises ayant déposé plus de 50 brevets relatifs aux véhicules autonomes



Source : D'après Lippert et al. (2018[5]), « Toyota's vision of autonomous cars is not exactly driverless », https://www.bloomberg.com/news/features/2018-09-19/toyota-s-vision-of-autonomous-cars-is-not-exactly-driverless.

Étant donné la complexité des systèmes de véhicules autonomes, les entreprises tendent à agir dans leur propre domaine d'expertise, puis à nouer des partenariats avec des spécialistes d'autres domaines. Par exemple, Waymo a mis à profit sa maîtrise des gros ensembles de données et de l'apprentissage automatique pour devenir l'une des entreprises leaders du marché des véhicules autonomes mais, comme elle ne construit pas elle-même de voitures, elle a choisi de s'associer à des partenaires comme General Motors (GM) et Jaguar (Higgins et Dawson, 2018[6]).

Les grands constructeurs automobiles concluent également des accords avec des startups plus petites afin d'avoir accès à des technologies de pointe. Ainsi, en octobre 2018, le groupe Honda a annoncé qu'il investissait 2.75 milliards USD dans l'entreprise à risque GM's Cruise, engagée dans le développement de véhicules autonomes (Carey et Lienert, 2018[7]). Les sociétés de transport avec chauffeur comme Uber investissent également beaucoup dans les véhicules autonomes, et mettent en place des partenariats avec des universités techniques de premier plan (CMU, 2015[8]). Cependant, ces évolutions posent la question de la responsabilité en cas d'accident, en particulier quand plusieurs parties prenantes ont la charge de différentes parties du système.

La diversité des acteurs prêts à miser sur les capacités des véhicules autonomes est confirmée par le nombre de brevets déposés par les différents groupes d'entreprises dans le domaine (Graphique 3.2). Les grands constructeurs investissent considérablement dans la propriété intellectuelle ; ils sont suivis de près par les équipementiers et les entreprises technologiques.

Évolution de la technologie

À la base, les véhicules autonomes sont dotés de systèmes de capteurs et de processeurs de calcul d'un genre nouveau, qui rendent d'autant plus complexe le processus d'extraction, de transformation et de chargement de leur système de données. Dans tous les domaines clés connexes, l'innovation, soutenue par de hauts niveaux d'investissement, est en plein essor. Par exemple, il est désormais possible de cartographier l'environnement du véhicule à l'aide de détecteurs de lumière et de distance moins coûteux. De nouvelles technologies de vision par ordinateur permettent en outre de suivre les yeux et la concentration de la personne au volant, pour repérer les moments où elle pourrait être distraite. En bout de chaîne, après la collecte et le traitement des données, l'IA ajoute une étape supplémentaire : la prise de décisions opérationnelles en une fraction de seconde.

Pour mesurer les progrès du développement des véhicules autonomes, on utilise généralement la norme de référence à six niveaux élaborée par la Society of Automotive Engineers (SAE) (ORAD, 2016[9]). Les six niveaux en question sont les suivants :

Niveau 0 (pas d'automatisation) : La conduite est entièrement à la charge d'un être humain. Le véhicule ne contient aucun système automatisé de direction, d'accélération, de freinage, etc.

Niveau 1 (assistance à la conduite) : Il existe une automatisation des fonctions de base, mais le conducteur ou la conductrice garde en permanence le contrôle de la plupart des fonctions. À ce niveau, précise la SAE, le contrôle latéral (direction) ou le contrôle longitudinal (par exemple, accélération) peuvent être automatiques, mais pas simultanément.

Niveau 2 (automatisation partielle de la conduite) : Les déplacements latéraux et longitudinaux se font automatiquement, par exemple avec un régulateur de vitesse adaptatif ou un dispositif de maintien du véhicule dans sa file de circulation.

Niveau 3 (automatisation conditionnelle de la conduite) : La voiture est conduite par le système, mais celui-ci doit pouvoir dire au conducteur ou à la conductrice de reprendre la main en cas de besoin. La personne qui se trouve au volant est la solution de repli du système : elle doit rester alerte et prête à conduire.

Niveau 4 (automatisation élevée de la conduite) : La voiture est conduite par le système qui n'a pas besoin que l'être humain reprenne la main en cas de problème. Cependant, le système n'est pas autonome en toutes circonstances (son autonomie dépend de la situation, de la zone géographique, etc.).

Niveau 5 (automatisation complète de la conduite) : La voiture se conduit seule, sans intervention humaine attendue, dans toutes les situations de conduite.

Les avis sont partagés quant au chemin parcouru sur la voie de l'automatisation complète de la conduite. Les acteurs du domaine ne sont pas non plus d'accord sur l'approche à suivre pour intégrer des fonctionnalités d'autonomie aux véhicules.

Le débat s'articule autour des deux thèmes que sont le rôle du conducteur et la disponibilité technologique :

a) **Rôle du conducteur**

Conduite sans être humain : Certaines entreprises de développement de véhicules autonomes telles que Waymo et Tesla pensent qu'il sera bientôt possible d'éliminer la nécessité de la présence d'une personne au volant (propriétaire ou responsable de la sécurité). Tesla vend aujourd'hui des voitures autonomes de niveau 3. Waymo avait le projet de lancer un service de taxis entièrement autonomes (sans conducteur) en Arizona à la fin de 2018 (Lee, 2018[10]).

Assistance à la conduite : D'autres concepteurs pensent que l'objectif à court terme doit être d'éviter les accidents plutôt que de remplacer la personne au volant. Toyota, le premier constructeur automobile du monde par capitalisation boursière, donne la priorité au développement d'un véhicule incapable de causer un accident (Lippert et al., 2018[5]).

b) **Disponibilité technologique** : Le déploiement de systèmes d'automatisation embarqués peut se faire selon deux approches, décrites par Walker-Smith (2013[11]) et FIT (2018[12]).

Toutes les fonctionnalités dans quelques zones : La voiture est équipée de fonctionnalités de très haut niveau opérationnelles uniquement dans certaines zones géographiques ou sur certaines routes cartographiées en détail. C'est le cas du système Super Cruise de Cadillac, qui n'est disponible qu'à certains endroits (il ne fonctionne que sur les autoroutes à chaussées séparées préalablement cartographiées).

Quelques fonctionnalités dans toutes les zones : La voiture est équipée des seules fonctionnalités d'autonomie utilisables en toute circonstance et sur n'importe quelle route. L'ensemble de ces fonctionnalités est donc limité, mais exploitable partout. Cette approche semble être celle que beaucoup de constructeurs automobiles privilégient actuellement.

Les entreprises les plus optimistes se sont fixé l'échéance de 2020 ou 2021 pour la fourniture de véhicules autonomes de niveau 4. Tesla et Zoox visent 2020 tandis que les groupes Audi/Volkswagen, Baidu et Ford misent sur 2021 et que Renault Nissan prévoit une livraison en 2022. D'autres constructeurs, qui investissent aussi beaucoup dans la technologie, ont choisi de privilégier la prévention des accidents de la conduite humaine, ou estiment que la technologie n'est pas assez développée pour une autonomie de niveau 4 à court terme. C'est le cas, notamment, de BMW, Toyota, Volvo et Hyundai (Welsch et Behrmann, 2018[13]).

Questions pour l'action publique

Le déploiement des véhicules autonomes pose un certain nombre de questions législatives et réglementaires importantes (Inners et Kun, 2017[14]). Certaines portent spécifiquement sur la sécurité et la protection de la vie privée (Bose et al., 2016[15]), mais d'autres concernent plus généralement l'économie et la société (Surakitbanharn et al., 2018[16]). Les pays de l'OCDE devraient concentrer leur réflexion sur les axes prioritaires suivants.

Sécurité et réglementation

En plus d'assurer la sécurité (voir chapitre 4, sous-section « Robustesse, sûreté et sécurité »), les pouvoirs publics doivent se poser les questions de la responsabilité civile, de la réglementation des équipements de régulation et de signalisation, de la réglementation applicable aux conducteurs, du code de la route et des règles d'exploitation (Inners et Kun, 2017[14]).

Données

Le succès des véhicules autonomes, comme des autres systèmes d'IA, passe par l'accès à des données permettant d'entraîner et d'ajuster les algorithmes. C'est pourquoi les constructeurs collectent d'immenses quantités de données au cours de leurs essais. Fridman (8 octobre 2018[17]) estime par exemple que Tesla possède les données relatives à plus de 2.4 milliards de kilomètres conduits par son Autopilot. Ces données de conduite en temps réel récupérées par les développeurs des véhicules autonomes sont propriétaires et, de ce fait, non partagées entre les entreprises. Cependant, des initiatives comme celle du Massachusetts Institute of Technology (MIT) (Fridman et al., 2018[18]) visent à construire des ensembles de données accessibles permettant de comprendre le comportement des conducteurs. Leur accessibilité rend ces ensembles de données d'autant plus importants pour les équipes de recherche et de développement qui souhaitent améliorer les systèmes. Les responsables de l'élaboration des politiques devraient discuter, entre autres, de l'accès aux données collectées par différents dispositifs et du rôle des pouvoirs publics dans le financement d'ensembles de données ouvertes.

Sécurité et vie privée

Afin de fonctionner dans des conditions de fiabilité et de sécurité, les véhicules autonomes ont besoin de beaucoup de données concernant le système, le comportement des conducteurs et l'environnement. Ils doivent aussi pouvoir se connecter à divers réseaux pour relayer l'information. C'est pourquoi les données qu'ils collectent, consultent et utilisent doivent être suffisamment protégées de tout accès non autorisé. Parmi ces données peuvent figurer des informations sensibles – emplacement et comportement de l'utilisateur, par exemple – qu'il s'agit de gérer et de protéger (Bose et al., 2016[15]). Le Forum international des transports appelle à la mise en place de cadres complets de cybersécurité pour régir la conduite automatisée (FIT, 2018[12]). À cet effet, de nouveaux protocoles et systèmes cryptographiques promettent de mieux protéger la vie privée et sécuriser les données. Mais leur mise en œuvre pourrait augmenter le temps de calcul nécessaire au tâches critiques pour la mission et pour la sécurité. De plus, leur développement commence à peine, ce qui veut dire qu'ils ne sont pas encore disponibles aux échelles et aux vitesses requises pour le déploiement de véhicules autonomes en temps réel.

Perturbation du marché du travail

L'essor des véhicules autonomes pourrait avoir un impact substantiel sur les métiers des secteurs du transport de marchandises, des taxis et de la livraison et sur d'autres emplois de service. Aux États-Unis par exemple, on estime que 2.86 % des travailleurs ont un métier axé sur la conduite (Surakitbanharn et al., 2018[16]). Bösch et al. (2018[3]) mettent en avant les économies potentiellement importantes que ces secteurs pourraient enregistrer en basculant vers des systèmes autonomes. C'est pourquoi, dans une perspective de maximisation des profits, on peut s'attendre à une transition rapide vers les véhicules autonomes lorsque la technologie sera suffisamment avancée. Il faudrait néanmoins que soient levés les obstacles non techniques, et notamment réglementaires. La mutation technologique évincera certains travailleurs, d'où la nécessité d'une action publique centrée sur les compétences et l'emploi dans le contexte d'un environnement de travail en transition (OCDE, 2014[19]).

Infrastructures

Le déploiement des véhicules autonomes pourrait nécessiter de modifier les infrastructures pour qu'elles conviennent à un environnement de conduite mixte mêlant voitures conduites et systèmes automatisés. À terme, les véhicules autonomes pourraient être équipés de moyens de communiquer les uns avec les autres. Mais les automobiles plus anciennes, avec conducteur, demeureraient une source importante d'incertitude. Les véhicules autonomes devraient être en mesure d'ajuster leurs décisions en fonction de celles des personnes toujours au volant de leur voiture. On réfléchit actuellement à la possible création de voies ou d'infrastructures dédiées aux véhicules autonomes, ce qui permettrait de les séparer des voitures conduites (Surakitbanharn et al., 2018[16]). Il faudra prendre l'habitude de planifier les politiques d'infrastructure en tenant compte des véhicules autonomes, à mesure que leurs technologies et leur déploiement progresseront.

L'IA dans le secteur de l'agriculture

L'agriculture est en train de se transformer, sous l'effet de l'amélioration de la précision des technologies informatiques cognitives telles que la reconnaissance d'image. Jusqu'ici, elle dépendait de l'œil et des mains de fermiers expérimentés capables d'identifier les plantes à récolter. Aujourd'hui, des « robots cueilleurs » avec systèmes d'IA et données transmises par des caméras et des capteurs peuvent prendre cette décision en temps réel. Les robots de ce type peuvent réaliser de plus en plus de tâches autrefois dévolues à l'être humain et à son savoir.

Des startups technologiques élaborent des solutions innovantes pour tirer le meilleur parti de l'IA dans le secteur de l'agriculture (FAO, 2017[20]). Les innovations peuvent être réparties en trois catégories (Tableau 3.1) :

Les **robots agricoles** effectuent des tâches agricoles essentielles telles que la récolte. Comparés aux êtres humains, ils sont de plus en plus rapides et productifs.

Les **systèmes de surveillance des sols et des cultures** exploitent la vision par ordinateur et des algorithmes d'apprentissage profond pour surveiller l'état du sol et des cultures. Leurs performances se sont améliorées à mesure que les données par satellite sont devenues plus disponibles (Graphique 3.3).

L'**analyse prédictive** utilise des modèles d'apprentissage automatique pour suivre et prédire l'impact des facteurs environnementaux sur le rendement des cultures.

Tableau 3.1. Exemples de startups spécialistes de l'IA en agriculture

Catégorie	Entreprise	Description
Robots agricoles	Abundant Robotics	Conceptrice d'un robot cueilleur de pommes qui utilise la vision par ordinateur pour détecter et cueillir des pommes avec la même précision et le même soin qu'un être humain. L'entreprise affirme que le travail d'un robot équivaut à celui de dix personnes.
	Blue River Technology	Conceptrice du robot See & Spray dont la fonction est de surveiller les plantes et les sols et de pulvériser de l'herbicide sur les mauvaises herbes dans les plantations de laitue et de coton. Une pulvérisation de précision peut aider à empêcher la résistance aux herbicides et réduire de 80 % le volume de produits chimiques consommé. En septembre 2017, l'équipementier John Deere a acquis cette entreprise pour 305 millions USD.
	Harveset CROO Robotics	Conceptrice d'un robot de cueillette et d'emballage de fraises. Capable de couvrir une superficie de 3.2 hectares par jour et de remplacer 30 personnes, ce robot peut aider à gérer les pénuries de main d'œuvre dans certaines régions agricoles clés, et à empêcher les pertes de revenu associées.

Catégorie	Entreprise	Description
Surveillance du sol et des cultures	PEAT	Conceptrice d'une application d'apprentissage profond programmée pour identifier les éventuels défauts des sols ou déficiences nutritives. Cette application peut établir un diagnostic de santé d'une plante sur la base des images prises l'agriculteur.
	Resson	Conceptrice d'algorithmes de reconnaissance d'image programmés pour détecter et classer avec précision les parasites et les maladies qui affectent les plantes. Resson a noué un partenariat avec McCain Foods pour aider à réduire les pertes au niveau de la chaîne de production des pommes de terre.
	SkySquirrel Technologies	Conceptrice d'un système qui analyse la santé des vignes sur la base d'images. Les utilisateurs téléchargent les photos obtenues à l'aide de drones dans le système infonuagique de l'entreprise qui, en retour, produit un diagnostic de l'état des feuilles de vignes. L'entreprise affirme que sa technologie permet de scanner 20 hectares en 24 minutes et d'analyser les données avec une précision de 95 %.
Analyse prédictive	aWhere	Conceptrice d'algorithmes d'apprentissage automatique utilisant des données par satellite pour prédire les conditions météorologiques et transmettre des avis personnalisés aux agriculteurs, consultants semenciers et chercheurs. L'entreprise fournit aussi à ses utilisateurs l'accès à plus d'un milliard de points de données agronomiques par jour.
	FarmShots	Conceptrice d'un système d'analyse de données agricoles tirées d'images prises par des drones ou des satellites. Ce système peut détecter des maladies, des parasites ou des plantes dont la nutrition est insuffisante dans les exploitations agricoles et indiquer aux utilisateurs où, précisément, leurs champs ont besoin d'engrais. La consommation d'engrais serait ainsi réduite de près de 40 %.

Source : Descriptions des entreprises sur leurs sites web respectifs.

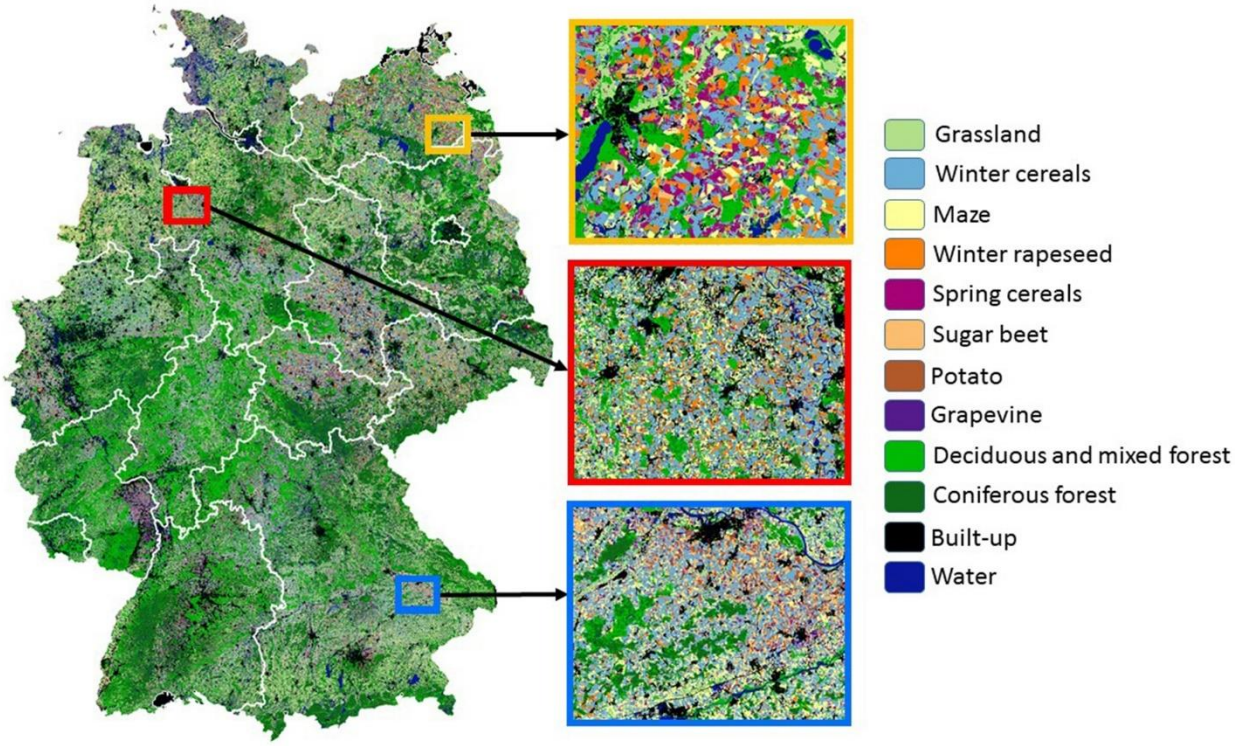
Obstacles à l'adoption de l'IA en agriculture

L'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) prédit une augmentation de la population mondiale de près de 30 % d'ici 2050, autrement dit de 7 milliards à 9 milliards de personnes. Pourtant, seuls 4 % de terres supplémentaires seront cultivés (FAO, 2009[21]). Dans ce contexte, l'OCDE étudie les opportunités et les défis de la transformation numérique dans le secteur agricole et alimentaire (Jouanjean, 2019[22]). Les applications de l'IA sont, de toutes les technologies numériques, celles qui se révèlent particulièrement prometteuses pour augmenter la productivité agricole. Cependant, les difficultés suivantes font obstacle à leur généralisation (Rakestraw, 2017[23]) :

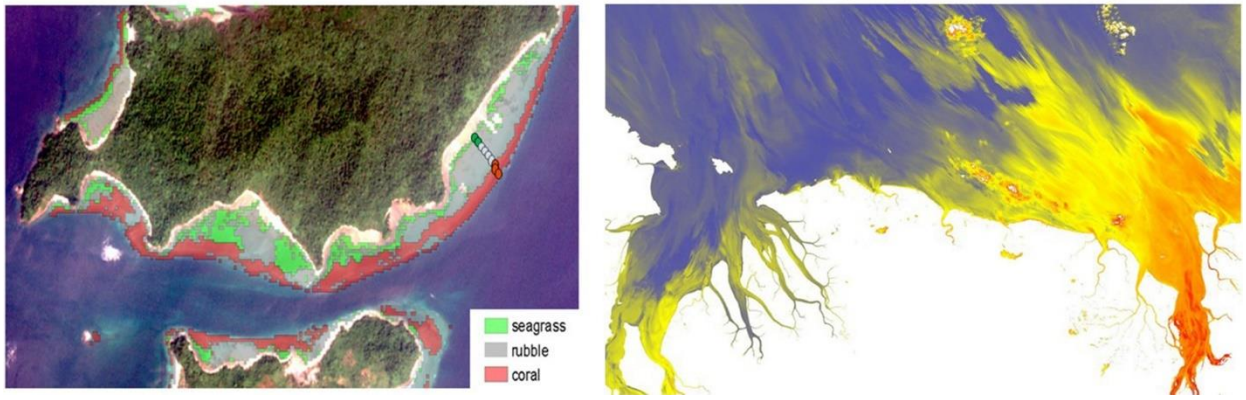
- **Manque d'infrastructures** : Les connexions au réseau restent mauvaises dans beaucoup de zones rurales. De plus, la construction d'applications robustes nécessiterait des systèmes d'entreposage des données.
- **Production de données de qualité** : Parce qu'elles ont pour but de reconnaître les plantes ou les feuilles, les applications agricoles de l'IA ont besoin de données de haute qualité. La collecte de ces données peut être coûteuse car elle ne peut se faire que pendant la saison de végétation.
- **Différences de perspectives entre les startups technologiques et les agriculteurs** : Les premières ont pour habitude de concevoir et de lancer rapidement leurs produits et services, mais les agriculteurs tendent à adopter de nouveaux procédés ou technologies à un rythme plus progressif. Les grosses entreprises agricoles elles-mêmes conduisent de longs essais au champ pour vérifier la régularité des performances et s'assurer que l'adoption d'une nouvelle technologie produira effectivement des bénéfices.
- **Coût, notamment pour les transactions** : Les équipements agricoles de haute technologie (par exemple, robots agricoles) exigent de gros investissements dans des capteurs et dispositifs d'automatisation. À titre d'exemple, la France élabore actuellement des politiques agricoles destinées à encourager l'investissement dans certaines applications agricoles de l'IA. Cela pourrait faciliter l'adoption des nouvelles technologies, y compris par les propriétaires de petites exploitations (OCDE, 2018[24]).

Graphique 3.3. Exemples de données par satellite utilisées pour améliorer la surveillance

A. Un programme d'apprentissage automatique analyse une série temporelle de données par satellite pour classer des terres cultivées (Allemagne)



B. Surveillance des zones côtières et marines (Australie) : fonds marins (à gauche) et turbidité (à droite)



Source : Roeland (2017[25]), *EC Perspectives on the Earth Observation*, www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-roeland.pdf ; Cooke (2017[26]), *Digital Earth Australia*, www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-cooke.pdf.

Moyens envisageables pour encourager l'adoption de l'IA en agriculture

On élabore actuellement des solutions pour relever les défis de l'IA en agriculture. Dans ce secteur comme dans d'autres, sont notamment développés des logiciels open source qui pourraient aider à résoudre les problèmes de coût. Par exemple, l'entreprise Connectra est partie de la suite logicielle open source TensorFlow de Google pour mettre au point un détecteur de mouvement qui s'attache au cou d'une vache et peut en surveiller la santé (Webb, 2017[27]). L'apprentissage par transfert (voir chapitre 4, sous-section « Accessibilité

et utilisation des données ») contribue à lever les difficultés liées aux données en donnant le moyen d'entraîner les algorithmes avec des ensembles de donnés beaucoup plus réduits. Ainsi, mettant à profit l'apprentissage effectué pour une autre variété de plante, des chercheurs ont conçu un système de détection des maladies susceptibles de toucher les plants de manioc : sur la base d'un ensemble de seulement 2 756 images de feuilles de manioc de Tanzanie, ces spécialistes en apprentissage automatique ont correctement identifié la présence d'une cercosporiose (maladie des taches foliaires brunes) avec une précision de 98 % (Simon, 2017[28]).

L'IA dans le secteur des services financiers

Dans le secteur de la finance, les grandes entreprises telles que JPMorgan, Citibank, State Farm et Liberty Mutual s'attachent à déployer rapidement l'intelligence artificielle. Il en est de même pour des startups telles que Zest Finance, Insurify, WeCash, CreditVidya et Aire. Les sociétés de services financiers combinent différentes pratiques d'apprentissage automatique, à l'instar de la startup française QuantCube Technology qui analyse plusieurs milliards de points de données collectés dans plus de 40 pays, puis utilise le traitement du langage, l'apprentissage profond, la théorie des graphes et d'autres méthodes encore pour concevoir des solutions d'IA d'aide à la décision destinées aux groupes financiers.

Déployer l'IA dans le secteur financier présente plusieurs avantages importants : amélioration de l'expérience client, identification rapide d'opportunités d'investissements pertinents, ou encore possibilité d'accorder à la clientèle davantage de crédits dans de meilleures conditions. Mais une telle évolution soulève aussi des questions concernant l'action publique, notamment sur les moyens d'assurer l'exactitude, d'empêcher les discriminations, et sur les impacts plus larges de l'automatisation sur l'emploi.

Cette section propose un tour d'horizon des applications de l'IA dans le secteur financier. Elle s'intéresse aux systèmes d'évaluation de la solvabilité des emprunteurs, à la technologie financière (fintech), à la négociation algorithmique, à la réduction des coûts des services financiers, à l'expérience client et à la conformité.

Systèmes d'évaluation de la solvabilité des emprunteurs

Le secteur des services financiers utilise depuis longtemps des méthodes statistiques à différentes fins, notamment le calcul des montants des apports personnels et l'estimation des risques d'insolvabilité. Dans ce dernier cas, les institutions financières conduisent une analyse statistique pour noter l'emprunteur en fonction de sa solvabilité. En d'autres termes, il s'agit d'évaluer le risque que cette personne ne puisse plus satisfaire à ses obligations de remboursement. Avec les modèles traditionnels, les analystes formulent des hypothèses concernant les attributs qui impactent ces notations et créent des segments de consommateurs.

Les techniques plus récentes des réseaux neuronaux permettent d'analyser de gros volumes de données issues des rapports de solvabilité et d'identifier dans le détail les facteurs les plus pertinents et les relations entre eux. Bâtis sur de grands ensembles de données, les algorithmes des systèmes d'IA déterminent automatiquement la meilleure configuration possible du réseau neuronal sous-jacent et, partant, les différents segments de consommateurs et leur pondération. Selon les agences de notation de crédit des États-Unis, les techniques d'apprentissage profond qui permettent d'analyser les données de manière inédite induiraient une hausse de la précision des prédictions pouvant atteindre 15 % (Press, 2017[29]).

Comme dans d'autres secteurs, il est problématique qu'il soit difficile d'expliquer les résultats des algorithmes fondés sur l'apprentissage automatique. Plusieurs législations nationales

exigent un haut niveau de transparence dans le secteur des services financiers. Citons notamment le *Fair Credit Reporting Act* (1970) et l'*Equal Credit Opportunity Act* (1974) qui, aux États-Unis, disposent que le déroulement et le résultat de tout algorithme soient explicables. Les entreprises semblent agir en ce sens. Par exemple, l'agence de notation de crédit Equifax et l'entreprise d'analyse de données SAS ont créé un outil d'évaluation de la solvabilité des emprunteurs, fondé sur l'apprentissage profond, qui reste interprétable.

Technologie financière et crédit instantané

La croissance des entreprises de technologie financière (fintech) est très rapide depuis quelques années. Les plateformes de crédit fintech offrent la possibilité de rechercher un prêt, de déposer une demande et d'obtenir la réponse en quelques clics seulement. Elles fournissent aux établissements prêteurs les données qu'on retrouve traditionnellement dans les rapports de solvabilité (historique des paiements, montants dus, ancienneté, nombre de comptes, etc.). Mais elles leur donnent aussi accès à d'autres sources de données très diverses, parmi lesquelles les sinistres d'assurance, les activités sur les réseaux sociaux, les informations relatives aux achats en ligne via certaines plateformes comme Amazon, les données d'expédition des services postaux, les profils de navigation sur le web, ou encore le type de téléphone ou de navigateur utilisé (Jagtiani et Lemieux, 2019[30]). Selon certaines recherches, les données alternatives que les entreprises fintech analysent avec l'IA peuvent donner aux personnes dépourvues d'historique bancaire classique un accès facilité au crédit. Elles peuvent aussi réduire les coûts associés au crédit, à la fois pour l'emprunteur et pour le prêteur (CSF, 2017[31]).

Une étude a cherché à comparer les performances des algorithmes de prédiction de la probabilité d'insolvabilité avec la notation FICO[3] habituelle aux États-Unis ou avec des données alternatives (Berg et al., 2018[32]). Si on utilise la seule notation FICO, le taux de précision est de 68.3 %, tandis que si on utilise les données alternatives, on atteint 69.6 %. Combinées, les deux sources de données permettent d'obtenir un taux de précision de 73.6 %. Ce résultat donne à penser que les données alternatives complètent les informations des agences de notation de crédit, mais sans s'y substituer. Un établissement prêteur peut donc prendre de meilleures décisions s'il a recours à la fois à des données traditionnelles (FICO) et à des données nouvelles.

En République populaire de Chine (ci-après, « la Chine »), la société Ant Financial a mis en exergue le rôle de l'IA dans le succès de sa stratégie de crédit (Zeng, 2018[33]), qui lui a permis de prêter plus de 13.4 milliards USD à près de trois millions de petites entreprises. Ant Financial utilise des algorithmes pour gérer l'important volume des données de transaction générées par les petites entreprises sur sa plateforme. Ces algorithmes analysent automatiquement, en temps réel, les données de transaction et les données comportementales relatives à tous les emprunteurs. Quelques minutes suffisent pour traiter une demande de prêt, dont le montant n'est parfois que de quelques centaines de CNY (environ 50 USD). Chaque action effectuée par une entreprise sur la plateforme Alibaba – transaction, communication entre vendeur et acheteur, ou connexion à d'autres services – a un impact sur sa notation de crédit. Mais les algorithmes qui calculent les notations évoluent eux aussi au cours du temps, ce qui permet une amélioration de la qualité des décisions à chaque itération. Ces opérations de microcrédit sont associées à un taux d'impayés d'environ 1 %, contre 4 % en moyenne à l'échelle de la planète selon une estimation de 2016 de la Banque mondiale.

Autre exemple, l'agence de notation de crédit Alipay utilise des points de données relatifs aux consommateurs pour calculer ses notations (O'Dwyer, 2018[34]). Les données en question peuvent être des historiques d'achats, le modèle de téléphone utilisé, les jeux pratiqués ou les amis sur les réseaux sociaux. La notation de crédit social mise en place en Chine peut

impacter, outre les traditionnelles décisions concernant l'attribution d'un prêt, le montant de la caution de location d'un appartement ou encore les mises en relation sur un site de rencontre. Par exemple, une personne qui jouerait chaque jour à des jeux vidéo pendant des heures pourrait se voir attribuer une notation de crédit social inférieure à celle d'une autre personne qui a acheté des couches, et dont on suppose donc qu'elle est un parent responsable (Rollet, 2018[35]). La Chine prévoit de déployer d'ici 2020 un système de crédit social élargi qui permettra de noter le « degré de confiance » que l'on peut accorder à un particulier, une entreprise ou un fonctionnaire.

L'utilisation de nouvelles sources de données donne la possibilité d'élargir l'accès au crédit. Mais elle soulève également des inquiétudes s'agissant des éventuelles disparités d'impact, de la protection de la vie privée, de la sécurité et de l'« explicabilité » (Gordon et Stewart, 2017[36]). C'est pourquoi, le Consumer Financial Protection Bureau des États-Unis a enquêté sur la façon dont ces données alternatives sont utilisées dans le calcul des notations de crédit (CFPB, 2017[37]).

Déployer l'IA pour réduire les coûts des services financiers

L'IA bénéficie aux clients et aux établissements financiers à tous les niveaux de l'interaction, c'est-à-dire à l'avant-plan (échanges avec le client, par exemple), au niveau intermédiaire (soutien aux échanges avec le client, par exemple) et à l'arrière-plan (règlements, ressources humaines, conformité, par exemple). Le déploiement de l'IA à ces trois niveaux devrait faire économiser aux entités financières une somme estimée à 1 000 milliards USD d'ici 2030 aux États-Unis, et impacter 2.5 millions d'employés des services financiers (Sokolin et Low, 2018[38]). En effet, plus l'intelligence artificielle progresse, moins l'intervention humaine est nécessaire.

À l'avant-plan de l'interaction avec le client (*front office*), les données financières et les opérations de gestion des comptes sont peu à peu intégrées à des agents logiciels fondés sur l'IA, qui peuvent dialoguer avec les clients via des plateformes telles que Facebook Messenger ou Slack grâce à des modules de traitement avancé du langage. Mais beaucoup d'entreprises du secteur financier, en plus d'améliorer leurs prestations habituelles de services, utilisent l'IA pour piloter des « robots conseillers ». Dans ce cas, les algorithmes proposent des offres et des conseils financiers automatisés (OCDE, 2017[39]).

Une autre évolution intéressante est le recours à l'analyse des émotions sur les plateformes des médias sociaux financiers. Des entreprises comme Seeking Alpha et StockTwits ont choisi de se concentrer sur le marché des actions : elles permettent aux utilisateurs de dialoguer entre eux et de consulter des professionnels pour faire fructifier leurs investissements. Les données produites sur ces plateformes peuvent ensuite être intégrées aux processus décisionnels (Sohangir et al., 2018[40]). L'IA facilite aussi les services bancaires en ligne et mobile en proposant des outils d'authentification par reconnaissance des empreintes digitales ou reconnaissance faciale grâce aux photos prises par les smartphones. Les banques utilisent aussi la reconnaissance vocale, plutôt que des mots de passe alphanumériques (Sokolin et Low, 2018[38]), pour autoriser l'accès à leurs services.

Au niveau intermédiaire (*middle office*), l'IA peut faciliter les processus de gestion des risques et de surveillance réglementaire. Elle aide les gestionnaires de portefeuilles à investir de façon plus précise et plus efficiente. Enfin, à l'arrière-plan (*back office*), elle élargit les sources de données exploitées pour évaluer les risques d'insolvabilité, les risques liés à la souscription d'une assurance, et les éventuels dommages (par exemple, évaluer un bris de pare-brise à l'aide d'un dispositif de vision par ordinateur).

Conformité juridique

Le secteur financier doit se conformer à des normes et exigences de déclaration réglementaire dont on sait qu'elles représentent un coût élevé. Les nouvelles réglementations entrées en vigueur aux États-Unis et dans l'Union européenne depuis une décennie ont encore augmenté les dépenses engagées par les banques pour se conformer à leurs obligations. En particulier, on estime à 70 milliards USD par an ce qu'elles ont dépensé ces dernières années pour s'assurer de leur conformité juridique et s'équiper en logiciels de gouvernance – un montant qui reflète le coût du travail des juristes bancaires, du personnel parajuridique et des autres agents chargés de vérifier la conformité des transactions. Le total de ces activités devait atteindre près de 120 milliards USD en 2020 (Chintamaneni, 26 juin 2017[41]). Or, le déploiement de l'IA, en particulier les technologies de traitement du langage, pourrait réduire de quelque 30 % les dépenses de conformité des banques, et considérablement raccourcir le temps de vérification de chaque transaction. L'IA peut aider à interpréter les documents réglementaires et à codifier les règles de conformité. Le programme Coin créé par JPMorgan Chase, par exemple, examine les documents en fonction de règles commerciales et de principes de validation des données : en quelques secondes, il peut passer en revue l'équivalent de ce qu'une personne mettrait 360 000 heurs à étudier (Song, 2017[42]).

Détection des fraudes

Les sociétés financières comptent aussi beaucoup sur l'IA pour la détection des fraudes. Les banques ont depuis toujours l'habitude de surveiller les profils d'activité des comptes. Avec l'apprentissage automatique, elles peuvent désormais espérer établir une surveillance en temps quasi-réel, ce qui leur permettrait, dès l'apparition d'une anomalie, de la repérer et de déclencher une procédure d'examen. Le fait que l'IA puisse continûment analyser de nouveaux schémas comportementaux et ajuster son programme en conséquence est une caractéristique spécifique fondamentale pour la détection des fraudes car, dans ce domaine, les profils évoluent rapidement. En 2016, l'établissement bancaire Credit Suisse Group AG a créé une coentreprise axée sur l'IA avec Palantir Technologies, une entreprise de la Silicon Valley spécalisée dans la surveillance et la sécurité. Pour aider les banques à détecter les négociations non autorisées, ils ont mis au point une solution destinée à repérer les employés aux comportements contraires à l'éthique avant qu'ils puissent mettre la banque en danger (Voegeli, 2016[43]). La détection des fraudes fondée sur des systèmes de sécurité biométrique à algorithme d'apprentissage automatique s'implante aussi progressivement dans le secteur des télécommunications.

Négociation algorithmique

On appelle négociation algorithmique le fait de confier à des algorithmes informatiques la tâche de décider automatiquement des transactions, de soumettre des ordres et de gérer ces ordres après leur soumission. En une décennie, cette pratique a spectaculairement gagné en popularité, au point qu'elle est aujourd'hui à l'origine de la majorité des ordres de bourse passés dans le monde. En 2017, JPMorgan estimait que seuls 10 % du volume des actions négociées sur les places boursières avaient fait l'objet d'une sélection spécifique en fonction de la valeur intrinsèque des titres, ce que l'on appelle le « *stock picking* » (Cheng, 2017[44]). Les possibilités accrues du calcul informatique favorisent également les « transactions à haute fréquence » (THF), c'est-à-dire la transmission quotidienne de millions d'ordres et l'étude simultanée de nombreux marchés. Si la plupart des négociateurs (« *traders* ») utilisent encore le même type d'outils de prédiction, l'IA permet la prise en compte d'un nombre plus élevé de facteurs.

L'IA dans le secteur du marketing et de la publicité

L'IA influe sur le marketing et la publicité de diverses manières. Il a pour effet premier de permettre une personnalisation de l'expérience en ligne de la clientèle, à laquelle on peut désormais proposer les contenus les plus susceptibles de l'intéresser. Les progrès de l'apprentissage automatique et l'augmentation concomitante des quantités de données générées donnent aux équipes publicitaires de plus en plus de moyens de cibler leurs campagnes. Elles peuvent aujourd'hui communiquer aux consommateurs des annonces personnalisées et dynamiques à une échelle sans précédent (Chow, 2017[45]). La publicité ciblée offre d'importants avantages aux entreprises et à leurs clientèles. Aux premières, elle pourrait apporter une hausse des ventes et du retour sur investissement des campagnes de marketing. Aux secondes, elle propose des services en ligne qui, financés par les revenus publicitaires, sont souvent gratuits dont permettent de fortement réduire les coûts de recherche.

Le panorama non exhaustif suivant donne une idée des progrès de l'IA qui pourraient impacter le plus les pratiques commerciales et publicitaires à l'échelle de la planète.

Traitement du langage : Le traitement du langage naturel (TLN) est l'un des sous-domaines majeurs de l'IA pour ce qui est de la personnalisation des messages publicitaires et commerciaux. Il permet d'adapter les campagnes au contexte linguistique, par exemple les messages postés sur les réseaux sociaux, les courriels, les interactions des clients avec le service après-vente, ou les avis relatifs aux produits. Les algorithmes de TNL « apprennent » à reconnaître les mots et à identifier les schémas récurrents des langues naturelles, et augmentent progressivement la précision de leurs prédictions. Ce faisant, ils peuvent en déduire les préférences ou les intentions d'achat des consommateurs (Hinds, 2018[46]). Ils permettent ainsi d'améliorer la qualité des résultats d'une recherche en ligne et de mieux aligner les publicités présentées sur les attentes personnelles de la clientèle, pour une plus grande efficacité publicitaire. Par exemple, si une personne a effectué une recherche en ligne en rapport avec une marque spécifique de chaussures, un algorithme publicitaire fondé sur l'IA peut afficher des publicités ciblées relatives à cette marque quand la personne effectue d'autres tâches en ligne. Il peut même lui envoyer une notification sur son téléphone si elle passe à proximité d'un magasin de chaussures proposant des réductions.

Analyse de données structurées : L'impact de l'IA dans le secteur du marketing va au-delà de la seule utilisation de modèles de TLN pour analyser des « données non structurées ». Avec l'IA, les algorithmes actuels de recommandation en ligne font beaucoup plus que ne le peuvent de simples ensembles de règles ou des historiques de notes attribuées par les utilisateurs. Ayant accès à des données très diverses, ils peuvent produire des recommandations très ciblées. Par exemple, Netflix propose des listes personnalisées de recommandations vidéo sur la base d'une analyse des films que chaque personne a déjà visionnés ou des notes qu'elle a données à ces films. Mais l'algorithme tient aussi compte du nombre de fois qu'un même film a été regardé et des actions (retour arrière ou avance rapide) en cours de visionnage (Plummer, 2017[47]).

Calcul d'une probabilité de succès : Le taux de clics – c'est-à-dire le nombre de personnes ayant cliqué sur un message publicitaire divisé par le nombre de personnes ayant vu le message – est un important indicateur de la performance d'une publicité en ligne. Des systèmes de prédiction du nombre de clics fondés sur des algorithmes d'apprentissage automatique ont donc été développés pour maximiser l'impact des publicités payantes et des campagnes de marketing en ligne. Dans la plupart des cas, c'est la technique de l'apprentissage automatique par renforcement qui est utilisée pour sélectionner la publicité dont les caractéristiques maximiseront le taux de clics dans la population cible. L'augmentation

du taux de clics peut substantiellement accroître les revenus d'une entreprise : une hausse de ce taux de 1 % suffirait à booster les ventes (Hong, 27 août 2017[48]).

Personnalisation des prix[4] : Grâce aux technologies de l'IA, les entreprises peuvent proposer des prix continuellement alignés sur les préférences et les comportements des consommateurs. Elles peuvent aussi réagir en fonction des lois de l'offre et de la demande, de l'exigence de profit et des externalités. Les algorithmes d'apprentissage automatique permettent de prédire le prix plafond que quelqu'un est prêt à payer pour un produit. Les prix sont ainsi calculés en fonction de chaque personne au point d'engagement, par exemple les plateformes en ligne (Waid, 2018[49]). S'il est vrai qu'on peut utiliser l'IA pour personnaliser les prix au service des clients, il ne faut pas oublier que la personnalisation des prix peut être une pratique nuisible si elle est fondée sur l'exploitation, la distorsion ou l'exclusion (Brodmerkel, 2017[50]).

Application combinée de la réalité augmentée et de l'IA : La réalité augmentée superpose à l'environnement réel perçu une représentation numérique d'un produit. En associant réalité augmentée et intelligence artificielle, on peut donner à quelqu'un une idée de ce à quoi ressemblerait le produit final une fois placé dans le contexte physique pour lequel il est prévu. Les systèmes de réalité augmentée qui fonctionnent avec l'IA peuvent « apprendre » des préférences individuelles, et ainsi adapter l'image du produit générée par ordinateur afin d'améliorer l'expérience du client et d'augmenter la probabilité d'achat (De Jesus, 2018[51]). Ils pourraient ainsi élargir le marché du commerce électronique et augmenter les revenus de la publicité en ligne.

L'IA dans le secteur de la science

Notre société est confrontée à des défis planétaires qui vont du changement climatique à la résistance bactérienne aux antibiotiques. Relever la plupart d'entre eux passe par l'approfondissement de nos connaissances scientifiques. À cet effet, l'IA pourrait augmenter la productivité des sciences – à l'heure où certains voix du monde universitaire soutiennent qu'il est de plus en plus difficile de trouver de nouvelles idées (Bloom et al., 2017[52]). L'IA promet aussi d'améliorer la productivité de la recherche, malgré les pressions croissantes qui pèsent sur les budgets publics qui lui sont affectés. L'émergence de nouveaux savoirs dépend de notre aptitude à donner sens aux gigantesques volumes de données produits par l'instrumentation scientifique moderne. C'est pourquoi, la science a absolument besoin de l'IA. Qui plus est, les scientifiques ayant sans doute atteint leur « pic de lecture », l'IA sera un complément nécessaire de l'être humain pour le dépouillement des articles scientifiques, publiés en nombre toujours plus élevé[5].

Appliquée à la science, l'IA pourrait aussi donner naissance à de nouvelles formes de découvertes et augmenter la reproductibilité de la recherche scientifique. Ses applications scientifiques et industrielles sont aujourd'hui nombreuses et de plus en plus déterminantes. Elle a permis, entre autres, de prédire le comportement de systèmes chaotiques, de résoudre des problèmes calculatoires complexes en génétique, d'améliorer la qualité des images astronomiques et de découvrir certaines règles de la synthèse chimique. Elle est actuellement déployée à d'autres fins qui peuvent aller de l'analyse de grands ensembles de données, l'abduction (la production d'hypothèses) et l'analyse et la compréhension de la littérature scientifique en vue de faciliter la collecte de données, à la conception expérimentale et à l'expérimentation elle-même.

Moteurs récents de l'IA en science

Cela fait déjà quelque temps qu'on applique diverses formes de l'IA à la découverte scientifique, même si ces tentatives sont sporadiques. Par exemple, dans les années 1960, le programme d'IA DENDRAL a aidé à identifier des structures chimiques. Une décennie plus tard, une IA connue sous le nom d'*Automated Mathematician* assistait la recherche en mathématique. Depuis ces premiers essais, les matériels et logiciels informatiques se sont spectaculairement améliorés, et les données sont beaucoup plus accessibles. Mais d'autres facteurs expliquent aussi l'application croissante de l'IA à la science : l'IA est bien financée, en particulier dans le secteur commercial ; les données scientifiques sont de plus en plus abondantes ; le calcul hautes performances s'améliore ; et le milieu de la recherche a désormais accès à des codes d'IA en open source.

Diversité des applications scientifiques de l'IA

De nombreuses disciplines font appel à l'IA pour faciliter leurs recherches. La physique des particules, par exemple, y a souvent recours quand elle cherche à repérer des configurations spatiales complexes dans les grands flux de données produits par les détecteurs de particules. En traitant les données collectées sur les réseaux sociaux, l'IA renseigne sur les relations entre utilisation des langues, psychologie et santé, et résultats économiques et sociaux. L'IA permet aussi, entre autres, de s'attaquer à des problèmes calculatoires complexes en génétique, d'améliorer la qualité des images astronomiques et de découvrir certaines règles de la synthèse chimique (OCDE, 2018[53]). La fréquence et l'éventail des applications de l'IA continueront probablement de croître. Plus les processus d'apprentissage automatique progresseront, plus la communauté scientifique, le secteur privé et d'autres utilisateurs prendront l'habitude de se tourner vers l'IA.

Des avancées ont aussi été enregistrées dans le domaine de l'abduction. Par exemple, IBM a produit un système prototype appelé KnIT qui explore les informations contenues dans les publications scientifiques, qui les représente explicitement sous la forme d'un réseau interrogeable, puis qui raisonne sur leur base pour produire de nouvelles hypothèses testables. En explorant la littérature publiée dans le domaine, KnIT a ainsi identifié de nouvelles kinases – des enzymes qui catalysent le transfert d'un ion phosphate de molécules à haut potentiel énergétique vers des substrats spécifiques. Ces kinases ont introduit un groupe phosphate dans une protéine de suppression tumorale (Spangler et al., 2014[54]).

De même, l'IA contribue à l'examen, la compréhension et l'analyse des publications scientifiques. Les techniques de traitement du langage naturel permettent aujourd'hui d'en extraire automatiquement à la fois des relations et du contexte. On a vu que dans le cas du système KnIT, l'exploration des textes publiés débouche sur une production automatisée d'hypothèses. La startup Iris.AI[6] propose quant à elle un outil gratuit d'extraction des concepts clés des résumés de recherche, qui donne à voir graphiquement ces concepts (c'est-à-dire que l'utilisateur peut voir les relations interdisciplinaires). Cet outil collecte aussi les articles pertinents dans une bibliothèque de plus de 66 millions de publications en accès ouvert.

L'IA peut en effet aider à collecter des données à grande échelle. Par exemple, les sciences participatives comptent sur des applications de l'IA pour aider les utilisateurs à identifier des spécimens inconnus de plantes ou d'animaux (Matchar, 2017[55]).

Combiner l'IA à la robotique pour mener des recherches scientifiques en boucle fermée

La science pourrait bénéficier à de multiples niveaux d'une convergence de l'IA et de la robotique. Les systèmes d'automatisation de laboratoire peuvent physiquement exploiter des techniques de l'IA pour la conduite d'expériences. Ainsi, dans le laboratoire de l'Université d'Aberystwyth (Pays de Galles), un robot du nom d'Adam utilise ces techniques pour effectuer automatiquement des cycles d'expérimentation scientifique. Il a été décrit comme la première machine à découvrir de nouvelles connaissances scientifiques de manière indépendante. Plus précisément, il a découvert un composé, le Triclosan, qui agit contre les espèces de type sauvage résistantes aux médicaments *Plasmodium falciparum* et *Plasmodium vivax* (King et al., 2004[56]). L'automatisation complète de la science aurait plusieurs avantages (OCDE, 2018[57]) :

- **Une découverte scientifique accélérée** : Les systèmes automatisés peuvent générer et tester des milliers d'hypothèses en parallèle, là où, du fait de leurs limites cognitives, les êtres humains ne peuvent examiner que quelques hypothèses à la fois (King et al., 2004[56]).
- **Des expérimentations moins coûteuses** : Les systèmes d'IA peuvent sélectionner les expériences dont la réalisation coûte moins cher (Williams et al., 2015[58]). Leur puissance rend possibles l'exploration et l'exploitation efficientes de paysages expérimentaux inconnus, ce qui pourrait conduire au développement de nouveaux médicaments (Segler, Preuss et Waller, 2018[59]), matériaux (Butler et al., 2018[60]) ou appareils (Kim et al., 2017[61]).
- **Des formations facilitées** : Enseignement initial compris, la formation complète d'un ou d'une scientifique dure plus de 20 ans et nécessite beaucoup de ressources. Les êtres humains ne peuvent absorber des connaissances que progressivement, par l'enseignement et l'expérience. Au contraire, les robots peuvent directement absorber les connaissances d'un autre.
- **Une amélioration des échanges de savoirs et de données et de la reproductibilité scientifique** : L'une des questions les plus importantes en biologie – et dans d'autres disciplines scientifiques – est celle de la reproductibilité. Les robots ont la capacité surhumaine d'enregistrer les tâches expérimentales et les résultats correspondants, lesquels sont, avec les métadonnées associées et les procédures appliquées, automatiquement et exhaustivement enregistrés conformément aux normes en vigueur et sans coût supplémentaire. Au contraire, l'enregistrement des données, des métadonnées et des procédures ajoute jusqu'à 15 % au coût total d'une expérience conduite par un être humain.

L'automatisation de laboratoire est essentielle dans la plupart des filières scientifiques et technologiques. Toutefois, coûteuse et difficile d'utilisation du fait du petit nombre d'unités vendues et de l'immaturité du marché, elle est surtout rentable quand elle est centralisée sur un seul et même grand site. C'est pourquoi les entreprises et les universités tendent de plus en plus à concentrer leurs systèmes d'automatisation de laboratoire. L'exemple le plus avancé en la matière est l'automatisation infonuagique. Cette pratique consiste à réunir de très nombreux équipements sur un même site, puis à proposer aux biologistes, par exemple, d'y envoyer leurs échantillons et de concevoir leurs expériences avec l'aide d'une application spécifique.

Considérations pour l'action publique

En ayant davantage recours aux systèmes d'IA, la science pourrait voir se modifier certains de ses aspects, notamment sociologiques et institutionnels : mode de transmission des connaissances, systèmes de crédit pour les découvertes scientifiques, mécanisme d'examen par les pairs, ou encore gestion des droits de propriété intellectuelle. À mesure que l'IA s'y généralisera, les politiques qui concernent l'accès aux données et le calcul hautes performances gagneront en importance. Qui plus est, la place croissante de l'IA dans le processus de découverte pose de nouvelles questions, dont on ignore encore la réponse. Faut-il inclure les machines dans les citations des publications ? Les systèmes de gestion des droits de propriété intellectuelle devront-ils être modifiés dans un monde où les machines peuvent inventer ? Qu'en est-il, enfin, de la question fondamentale de l'enseignement et de la formation (OCDE, 2018[57]) ?

L'IA dans le secteur de la santé

Contexte

Appliquée aux soins de santé et à l'industrie pharmaceutique, l'IA peut aider à détecter précocement des maladies, proposer des services de prévention, optimiser la prise de décision clinique et découvrir des traitements et des médicaments. Elle ouvre également la voie à une personnalisation des soins de santé et une médecine de précision, grâce aux outils, applications et moniteurs d'autosurveillance dans lesquels on la retrouve. Elle pourrait être avantageuse à la fois en termes de qualité et de coût des soins. Elle pose néanmoins certaines questions pour l'action publique, notamment en ce qui concerne l'accès aux données (de santé) et la protection de la vie privée (voir chapitre 4, sous-section « La protection des données personnelles »). Cette section se concentre sur les impacts spécifiques de l'IA sur les soins de santé.

D'une certaine façon, le secteur de la santé est une enceinte idéale pour le déploiement de systèmes d'IA et l'illustration parfaite de ses effets possibles. À forte intensité de connaissances, il ne peut améliorer ses thérapies et ses pratiques sans données ni capacités d'analyse. C'est pourquoi on y constate un élargissement considérable de l'éventail des informations collectées – qui peuvent être cliniques, génétiques, comportementales ou environnementales. Les professionnels de santé, les acteurs de la recherche biomédicale et les patients produisent chaque jour des quantités massives de données au moyen d'une multitude de dispositifs, parmi lesquels les dossiers de santé informatisés, les séquenceurs de gènes, les appareils d'imagerie médicale à haute résolution, les applications pour smartphones et les capteurs ubiquitaires, ainsi que tous les objets connectés, relevant de l'internet des objets, conçus pour surveiller l'état de santé de quelqu'un (OCDE, 2015[62]).

Effets positifs de l'IA sur les soins de santé

L'exploitation des données générées par l'IA pourrait être très utile pour les soins de santé et la recherche. Dans tous les pays, les secteurs de la santé se transforment en profondeur à mesure qu'ils mettent à profit les possibilités offertes par les technologies de l'information et de la communication. Ce processus de mutation obéit à des objectifs clés que sont l'amélioration de l'efficience, de la productivité et de la qualité des soins (OCDE, 2018[24]).

Exemples spécifiques

Prodiguer de meilleurs soins aux patients : Avec l'utilisation secondaire des données de santé, on peut espérer améliorer la qualité et l'efficacité des soins, que ce soit en milieu

hospitalier ou au domicile des patients. Par exemple, des systèmes d'IA peuvent alerter les administrateurs ou les soignants de première ligne quand des indicateurs liés à la qualité ou à la sécurité des patients s'écartent de la normale. Ils peuvent aussi mettre en évidence les déterminants possibles de ces déviations (Institut canadien d'information sur la santé, 2013[63]). L'un des volets spécifiques de l'amélioration des soins aux patients est celui de la **médecine de précision**, qui repose sur le traitement rapide d'une diversité de données complexes telles que celles du dossier médical, les réactions physiologiques et les données génétiques. La **santé mobile** en est un autre : les technologies mobiles fournissent un utile retour d'information en temps réel tout au long du continuum des soins – de la prévention au diagnostic, au traitement et au suivi. En association avec d'autres données personnelles, notamment sur le lieu de vie ou les préférences, les technologies de l'IA sont à même d'identifier les comportements à risque ou d'encourager les comportements bénéfiques. Elles peuvent alors produire des interventions ciblées pour promouvoir des comportements plus sains (par exemple, prendre les escaliers au lieu de l'ascenseur, boire de l'eau ou marcher plus) dans la perspective d'une amélioration de la santé. Ces technologies, comme les équipements de suivi qui utilisent des capteurs, offrent la possibilité d'une surveillance continue et directe et d'une intervention personnalisée. En tant que telles, elles sont particulièrement indiquées pour améliorer la qualité des soins aux personnes âgées et aux personnes en situation de handicap (OCDE, 2015[62]).

Gestion des systèmes de santé : Les données de santé peuvent venir étayer des décisions concernant les programmes, les politiques et les financements et, de cette façon, aider à gérer le système de santé, et à en améliorer l'efficacité et l'efficience. En déployant des systèmes d'IA, on peut identifier les interventions inefficaces, les opportunités manquées et les services dupliqués et donc réduire les coûts. Quatre angles d'action sont envisageables pour élargir l'accès aux soins et réduire les temps d'attente. Premièrement, les systèmes d'IA ont la capacité d'appréhender le parcours des patients le long du continuum des soins. Deuxièmement, ils peuvent faire en sorte que les patients reçoivent les services les mieux adaptés à leurs besoins. Troisièmement, ils peuvent prédire avec précision les futurs besoins de soins de santé de la population. Quatrièmement, ils constituent un moyen d'optimiser l'allocation des ressources à l'échelle du système (Institut canadien d'information sur la santé, 2013[63]). Dans le contexte du renforcement de la surveillance des thérapies et des événements causés par des produits pharmaceutiques ou des dispositifs médicaux (OCDE, 2015[62]), les administrations nationales peuvent déployer l'IA pour faire progresser l'identification des schémas récurrents à l'échelle du système, qu'ils s'agisse des erreurs ou des succès. De façon plus générale, l'innovation fondée sur les données donne à voir le système de santé comme un système « apprenant », c'est-à-dire à même d'intégrer en continu de nouvelles données issues des établissements de recherche, des prestataires de soins, ou des patients. Le système peut alors, sur cette base, améliorer les algorithmes cliniques généraux afin de mettre en évidence le type de soin préférable à certains nœuds de l'arbre décisionnel en vue d'appuyer la prise de décision clinique (OCDE, 2015[62]).

Comprendre et gérer les questions de santé publique : Les données peuvent aider non seulement à surveiller de plus près l'émergence de problèmes de santé publique comme une épidémie de grippe ou d'autres virus, mais aussi à identifier les effets secondaires imprévus et les contre-indications des nouveaux médicaments (Institut canadien d'information sur la santé, 2013[63]). Les technologies de l'IA permettent de détecter le plus tôt possible l'apparition d'une maladie et d'en surveiller la propagation. Grâce aux réseaux sociaux, par exemple, il est possible d'obtenir aussi bien que de diffuser des informations sur la santé publique. En effet, l'association de l'IA et des outils de traitement du langage naturel donne les moyens d'analyser les messages postés sur les réseaux sociaux pour en extraire des informations sur des effets secondaires potentiels (Comfort et al., 2018[64] ; Patton, 2018[65]).

Faciliter la recherche dans le domaine de la santé : Les données de santé peuvent étayer la recherche clinique et accélérer la découverte de nouvelles thérapies. L'analyse des données massives offre de nouvelles possibilités plus prometteuses de mesurer la progression des maladies et l'état de santé de la population, pour de meilleurs diagnostics et prestations de soins, ainsi qu'une meilleure recherche translationnelle et clinique, par exemple s'agissant de la mise au point de nouveaux médicaments. À titre d'illustration, l'entreprise pharmaceutique Atomwise a collaboré en 2015 avec des chercheurs de l'Université de Toronto et avec IBM pour appliquer l'IA à la recherche d'un traitement contre le virus Ebola[7]. L'IA est également de plus en plus souvent testée pour le diagnostic médical, et vient d'ailleurs de bénéficier d'une approbation remarquée, délivrée par la Food and Drug Administration des États-Unis. Cette décision autorise la commercialisation du premier dispositif médical à utiliser l'IA pour « détecter un état supérieur au niveau bénin de rétinopathie diabétique chez les adultes atteints de diabète » (FDA, 2018[66]). De la même façon, on peut se servir des techniques d'apprentissage automatique pour entraîner des modèles à classer des images de l'œil, ce qui pourrait conduire à intégrer des détecteurs de cataracte dans des smartphones que l'on pourrait alors emporter dans les zones reculées (Lee, Baughman et Lee, 2017[67] ; Patton, 2018[65]). Une étude récente a consisté à entraîner un algorithme d'apprentissage profond avec plus de 100 000 images de mélanomes (nævus malins) et de grains de beauté (nævus bénins) : en définitive, le programme a pu détecter un cancer de la peau avec une performance supérieure à celle d'un groupe international de 58 dermatologues (Mar et Soyer, 2018[68]).

Déployer l'IA dans le secteur de la santé – facteurs de risque et de succès

La pleine exploitation des capacités de l'IA dans le secteur de la santé passe par la mise en place des infrastructures suffisantes et des bons mécanismes d'atténuation des risques.

Les administrations nationales prennent de plus en plus l'habitude d'établir des dossiers de santé informatisés (DSI) et d'adopter des solutions de santé mobile (m-santé), c'est-à-dire des services mobiles à l'appui de la pratique de la médecine et de la santé publique (OCDE, s.d.[69]). De solides éléments de preuve montrent en quoi les DSI peuvent aider à réduire les erreurs de médication et mieux coordonner les soins (OCDE, 2018[24]). Cependant, toujours selon la même étude, l'intégration des outils n'est forte que dans quelques pays ayant capitalisé sur la possibilité d'extraire des données des DSI à des fins de recherche, à des fins statistiques, ou pour d'autres utilisations secondaires. Les systèmes de santé tendent encore à collecter les données en silos et à les analyser séparément. Exploiter le plein potentiel des DSI nécessite de relever les défis clés de la normalisation et de l'interopérabilité (OCDE, 2018[24]).

Il est également critique pour l'utilisation de l'IA dans le secteur de la santé de **réduire autant que possible les risques d'atteinte à la vie privée des personnes concernées par les données** (les « sujets »). Les risques liés à une augmentation de la collecte et du traitement des données personnelles sont décrits dans la sous-section « La protection des données personnelles » du chapitre 4. Cette sous-section porte spécifiquement sur la nature hautement sensible des données de santé. Les biais de fonctionnement des algorithmes de recommandation d'un traitement particulier pourraient créer de réels risques de santé au sein de certains groupes. D'autres risques d'atteinte à la vie privée sont propres au secteur de la santé. Ainsi, des questions liées à l'exploitation de données extraites de dispositifs médicaux implantables comme les pacemakers pourraient être présentées devant des tribunaux[8]. De plus, la sophistication croissante de ces dispositifs augmente les risques de sécurité, par exemple le risque qu'un tiers malveillant prenne le contrôle d'un appareil pour effectuer une action dangereuse. L'utilisation d'échantillons biologiques (tissus, notamment) pour l'apprentissage automatique soulève aussi des questions complexes de consentement et de propriété (OCDE, 2015[62] ; Ornstein et Thomas, 2018[70])[9].

Du fait de ces inquiétudes, de nombreux pays de l'OCDE déclarent disposer d'obstacles législatifs à l'utilisation des données personnelles de santé, parmi lesquels la désactivation des liens entre les données et des entraves au développement de bases de données fondées sur les DSI. La Recommandation de 2016 du Conseil sur la gouvernance des données de santé est une étape importante sur la voie du renforcement de la cohérence en matière de gestion et d'utilisation des données de santé (OCDE, 2016[71]). Son principal objectif est de promouvoir l'élaboration et le déploiement d'un cadre national de gouvernance des données de santé, qui encouragerait la mise à disposition et l'utilisation des données personnelles de santé au service de la santé publique, tout en demandant que soient protégées la vie privée, les données personnelles de santé et la sécurité des données. Adopter une démarche cohérente de gestion des données aiderait à éviter d'avoir à faire des compromis entre utilisation des données et sécurité.

Impliquer toutes les parties concernées est un moyen important de susciter la confiance et l'adhésion du public concernant l'utilisation de l'IA et la collecte des données à des fins de gestion de la santé. Dans le même ordre d'idée, les pouvoirs publics pourraient élaborer des cursus adaptés pour former les futurs experts en science des données de santé, ou associer des experts en science des données au personnel soignant pour qu'ils travaillent ensemble à l'approfondissement de la compréhension des possibilités et des risques de cette discipline émergente (OCDE, 2015[62]). La participation des soignants à la conception et au développement de systèmes de soins de santé fondés sur l'IA pourrait être déterminante pour obtenir la confiance des patients et des prestataires de soins dans les produits et services de santé fondés sur l'IA.

L'IA dans le secteur de la justice pénale

IA et algorithmes prédictifs pour la justice

L'IA offre la possibilité d'améliorer l'accès à la justice ainsi que l'impartialité et l'efficacité de la prise de décision. Cependant, elle suscite des inquiétudes du fait des problèmes qu'elle pourrait soulever en matière de participation citoyenne, de transparence, et de respect de la dignité, de la vie privée et de la liberté. Cette section portera principalement sur les progrès de l'IA appliquée à la justice pénale, même si certaines évolutions dans d'autres domaines de la justice sont également abordées.

L'IA apparaît de plus en plus à différents stades de la procédure pénale, qu'il s'agisse de prédire l'occurrence d'un crime ou le résultat d'une procédure pénale, de conduire une évaluation des risques posés par les prévenus, ou encore de gérer les procédures avec davantage d'efficience. Même si beaucoup d'applications sont encore expérimentales, quelques outils de prédiction plus avancés sont déjà utilisés par l'administration judiciaire et les forces de l'ordre. L'IA permet en effet de mieux établir des connexions, détecter des schémas récurrents, prévenir et résoudre les crimes (Wyllie, 2013[72]). Le recours croissant à de tels outils traduit une évolution plus générale qui tend à donner la préférence à des méthodes axées sur les faits, pour la raison qu'elles sont un moyen plus efficient, rationnel et rentable d'allouer les ressources limitées dont disposent les forces de l'ordre (Horgan, 2008[73]).

La justice pénale se situe à un carrefour sensible des échanges entre les pouvoirs publics et les citoyens, où l'asymétrie de l'information et des relations de pouvoir est particulièrement prononcée. Sans garde-fous suffisants, elle pourrait produire des résultats négatifs disproportionnés et renforcer les biais systémiques voire en créer de nouveaux (Barocas et Selbst, 2016[74]).

Police prédictive

On parle de police prédictive quand les forces de l'ordre se servent de l'IA pour identifier des schémas et ainsi faire des prédictions statistiques sur l'activité criminelle possible (Ferguson, 2014[75]). Les méthodes de police prédictive préexistent à l'IA : dans un cas notable, une analyse des données accumulées avait permis de cartographier des villes pour y repérer les quartiers à risque faible ou élevé (Brayne, Rosenblat et Boyd, 2015[76]). Avec l'IA, cependant, il est possible de mettre en relation de nombreux ensembles de données et de conduire des analyses plus complexes donnant des résultats plus fins, à même de donner des prédictions plus précises. En combinant par exemple les lecteurs automatiques de plaques d'immatriculation, les caméras ubiquitaires, les dispositifs de stockage à moindre coût et la puissance de calcul, les forces de police peuvent obtenir des informations importantes sur beaucoup de monde et, partant, identifier des schémas, notamment de comportements criminels (Joh, 2017[77]).

Il existe deux grandes méthodes de police prédictive. La **prédiction situationnelle** utilise des données rétrospectives de la criminalité pour prévoir quand et où de nouveaux crimes sont susceptibles de se produire. Les lieux pris en compte peuvent être des débits de boissons alcoolisées, des bars et des parcs où d'autres faits ont déjà été signalés. Dans ce cas, les services de police peuvent décider, pour prévenir de nouveaux crimes, d'envoyer un agent patrouiller dans la zone, à un moment précis de la journée ou de la semaine. La **prédiction axée sur la personne** utilise les statistiques de la criminalité pour prévoir quels individus ou groupes d'individus sont les plus susceptibles d'être concernés par un crime – soit parce qu'ils en seraient les victimes, soit parce qu'ils en seraient les auteurs.

Des initiatives de police prédictive fondée sur l'IA sont en cours d'expérimentation dans diverses villes du monde, dont Manchester, Durham, Bogota, Londres, Madrid, Copenhague et Singapour. Au Royaume-Uni, les services de police du Grand Manchester ont mis au point un système de cartographie prédictive de la criminalité dès 2012. Un an plus tard, la police du Kent a commencé à utiliser un système appelé PredPol. Ces deux systèmes estiment la probabilité d'occurrence de crimes à certains endroits et à certaines périodes. Ils reposent sur un algorithme développé à l'origine pour prédire les tremblements de terre.

En Colombie, la *Data-Pop Alliance* exploite des données de criminalité et de transport pour prédire les points chauds de la criminalité à Bogota. Des forces de police sont alors déployées sur les lieux précis et aux heures précises où le risque de crime est le plus élevé.

De nombreux services de police tirent également parti des réseaux sociaux à des fins très diverses, comme la découverte d'une activité criminelle, l'obtention d'une cause probable pour un mandat de recherche, la collecte d'éléments de preuve en vue d'une audience au tribunal, la localisation d'un criminel, la gestion de situations volatiles, l'identification de témoins, la diffusion d'informations ou l'appel au public pour la collecte d'informations (Mateescu et al., 2015[78]).

Mais l'IA soulève aussi des questions concernant l'utilisation des données personnelles (voir chapitre 4, sous-section « La protection des données personnelles ») et les risques de biais (voir chapitre 4, sous-section « Équité et éthique »). Le fait qu'elle puisse manquer de transparence et le fait qu'on ne puisse pas toujours comprendre son fonctionnement sont deux points d'inquiétude particulièrement sensibles quand il s'agit de justice pénale. L'une des méthodes retenues pour améliorer la transparence algorithmique, et qui est appliquée au Royaume-Uni, est un cadre appelé ALGO-CARE dont l'objectif est de faire en sorte que les forces de police qui ont recours à des outils algorithmiques d'évaluation des risques en envisagent les principaux aspects juridiques et pratiques (Burgess, 2018[79]). Ce cadre transpose

les principes clés du droit public et des droits humains, qui figurent dans les documents à haut niveau, en termes et en directives pratiques à l'intention des services de police.

IA pour l'autorité judiciaire

Dans plusieurs pays, les autorités judiciaires utilisent avant tout l'IA pour évaluer les risques. Les résultats de ces évaluations viennent étayer divers types de décisions pénales, dont la fixation du montant d'une caution ou d'autres conditions de libération ou l'éligibilité à la libération conditionnelle (Kehl, Guo et Kessler, 2017[80]). Quand elle est ainsi appliquée pour évaluer les risques, l'IA fait intervenir d'autres formes d'outils actuariels que ceux dont les juges se servent depuis des décennies (Christin, Rosenblat et Boyd, 2015[81]). Des chercheurs du Berkman Klein Center de l'Université de Harvard travaillent actuellement sur une base de données de tous les outils d'évaluation des risques utilisés dans le cadre de la justice pénale aux États-Unis pour étayer la prise de décision (Bavitz et Hessekiel, 2018[82]).

Les algorithmes d'évaluation des risques prédisent le niveau de risque sur la base d'un petit nombre de facteurs, généralement répartis en deux groupes : les antécédents criminels (par exemple, précédentes arrestations et condamnations, défauts de comparution) et les caractéristiques sociodémographiques (par exemple, âge, sexe, emploi et lieu de résidence). Les algorithmes prédictifs font la synthèse des informations pertinentes pour la prise de décision plus efficacement que le cerveau humain, d'une part parce qu'ils traitent davantage de données à une vitesse supérieure, et d'autre part parce qu'il se pourrait qu'ils soient moins exposés aux préjugés humains (Christin, Rosenblat et Boyd, 2015[81]).

Les outils d'évaluation des risques fondés sur l'IA que mettent au point les entreprises privées soulèvent des inquiétudes inédites en termes de transparence et d'explicabilité. En effet, les accords de non-divulgation empêchent souvent l'accès au code propriétaire pour protéger la propriété intellectuelle ou prévenir les actes de malveillance (Joh, 2017[77]). Or, sans accès au code, il reste peu de moyens d'examiner la validité et la fiabilité des outils.

L'organisation de presse à but non lucratif ProPublica a rapporté avoir testé la validité de l'outil propriétaire COMPAS utilisé dans certaines juridictions aux États-Unis : il ressort de ces tests que les prédictions de COMPAS sont exactes dans 60 % des cas, tous crimes confondus, mais que la précision de prédiction n'est plus que de 20 % dans le cas des crimes violents. Des disparités raciales ont également été mises au jour : l'algorithme a qualifié par erreur des accusés noirs de futurs criminels deux fois plus souvent qu'il ne l'a fait pour les accusés blancs (Angwin et al., 2016[83]). Cette enquête a fait parler d'elle dans les médias et ses résultats ont été remis en question sur la base d'erreurs statistiques (Flores, Bechtel et Lowenkamp, 2016[84]). COMPAS est un algorithme de type « boîte noir », ce qui signifie que personne, pas même ses utilisateurs, n'a accès au code source.

L'utilisation de COMPAS a été contestée au tribunal, ses opposants affirmant que sa nature propriétaire est en violation avec le droit des accusés à un procès équitable. La Cour suprême du Wisconsin a approuvé l'utilisation de COMPAS dans le cadre du prononcé de la peine. Cependant, l'outil doit rester un moyen d'assistance et le juge doit conserver l'entière liberté de déterminer quels sont les facteurs complémentaires à prendre en compte et avec quel poids ils doivent l'être[10]. La Cour suprême des États-Unis a refusé le recours lui demandant d'entendre l'affaire[11].

Dans le cadre d'une autre étude de l'impact de l'IA sur la justice pénale, Kleinberg et al. (2017[85]) ont construit un algorithme d'apprentissage automatique destiné à prédire si une personne accusée commettrait un nouveau crime dans l'intervalle jusqu'au procès ou chercherait à se soustraire au procès (manquements avant le procès). Les variables d'entrée

étaient connues. L'algorithme devait calculer les sous-catégories pertinentes et leurs pondérations respectives : par exemple, pour la variable « âge », l'algorithme a déterminé les intervalles les plus statistiquement pertinents, notamment les deux tranches 18-25 ans et 25-30 ans. Les auteurs ont constaté que cet algorithme pouvait considérablement réduire les taux d'incarcération, ainsi que les disparités raciales. L'IA a en outre réduit les biais humains : les auteurs ont conclu que toute information autre que les facteurs nécessaires à la prédiction pouvait distraire les juges et augmenter le risque de jugement biaisé.

Les outils avancés d'IA développés pour évaluer les risques sont aussi utilisés au Royaume-Uni. Les services de police de Durham ont mis au point un outil spécifique, le *Harm Assessment Risk Tool*, pour évaluer les risques de récidive. Les facteurs pris en compte sont notamment les antécédents criminels de la personne, son âge, son code postal et d'autres informations de contexte. Sur la base de ces indicateurs, l'algorithme affecte à la personne un degré de risque faible, moyen ou élevé.

IA pour prédire le résultat des procédures

Avec des techniques avancées de traitement du langage et d'analyse des données, des chercheurs ont construit des algorithmes pour réduire l'issue des procédures avec un taux élevé de précision. Ainsi, des équipes de l'University College de Londres et des Universités de Sheffield et de Pennsylvanie ont développé un algorithme d'apprentissage automatique capable de prédire l'issue des affaires entendues par la Cour européenne des droits de l'homme avec un taux de précision de 79 % (Aletras et al., 2016[86]). D'autres chercheurs de l'Illinois Institute of Technology de Chicago ont conçu un algorithme à même de prédire l'issue des affaires entendues par la Cour suprême des États-Unis avec un taux de précision de 79 % (Hutson, 2017[87]). De tels algorithmes pourraient aider les parties à évaluer la probabilité de succès du procès en première instance ou en appel (en fonction des résultats d'affaires analogues). Ils pourraient aussi aider les avocats à identifier les points à mettre en avant pour augmenter leurs chances de gagner.

Autres utilisations de l'IA dans le cadre des procédures juridiques

Les juridictions civiles peuvent appliquer l'IA à des fins plus larges. Les avocats s'en servent pour rédiger des contrats ou pour analyser des documents et en extraire des données dans le cadre des procédures d'enquête et de vérification (Marr, 2018[88]). L'utilisation de l'IA pourrait s'étendre à d'autres domaines similaires de la justice pénale comme les négociations de peine et les interrogatoires. Cependant, parce que les modalités de conception et d'utilisation des algorithmes peuvent influer sur leurs résultats, il convient de porter la plus grande attention aux répercussions de l'IA sur l'action publique.

L'IA dans le secteur de la sécurité

L'IA promet d'aider à résoudre les problèmes complexes de sécurité physique et numérique. En 2018, les dépenses mondiales de défense devraient atteindre 1 670 milliards USD, ce qui représente une augmentation de 3.3 % en glissement annuel (IHS, 2017[89]). Mais le secteur public n'est pas le seul à investir dans la sécurité : il était estimé que le secteur privé dépenserait 96 milliards USD, à l'échelle de la planète, pour répondre aux risques de sécurité en 2018, soit une hausse de 8 % par rapport à 2017 (Gartner, 2017[90]). En montrant que les violations des données peuvent avoir de larges conséquences économiques et sociales et des répercussions sur la sécurité nationale, les récentes attaques de sécurité numérique de grande ampleur ont rendu la société dans son ensemble plus sensible à la question. Dans ce contexte, les acteurs privés comme publics adoptent et déploient des

technologies de l'IA pour s'adapter à la nouvelle situation mondiale en matière de sécurité. Cette section s'intéresse à deux filières du secteur de la sécurité qui connaissent un essor particulier : la sécurité numérique et la surveillance[12,13].

IA et sécurité numérique

L'IA est déjà très présente dans les applications de sécurité numérique pour tout ce qui concerne la sécurité des réseaux, la détection des anomalies, l'automatisation des opérations de sécurité et la détection des menaces (OCDE, 2018[24]). Dans le même temps, on s'attend à une augmentation des utilisations malveillantes de l'IA, qui peuvent se manifester notamment à travers l'identification des vulnérabilités logicielles dans le but de les exploiter et ainsi de porter atteinte à l'accessibilité, l'intégrité ou la confidentialité des systèmes, des réseaux ou des données. Cette évolution modifiera la nature et le niveau global des risques de sécurité numérique.

Deux tendances rendent les systèmes d'IA d'autant plus pertinents pour la sécurité : le nombre croissant d'attaques de sécurité numérique et la pénurie de main d'œuvre dans le secteur de la sécurité numérique (ISACA, 2016[91]). Dans ce contexte, les outils d'apprentissage automatique et les systèmes d'IA gagnent en pertinence pour automatiser la détection des menaces et la réponse aux incidents (MIT, 2018[92]). Face à des logiciels malveillants en mutation constante, l'apprentissage automatique est devenu indispensable pour combattre les attaques telles que les virus polymorphiques, le déni de service et le hameçonnage[14]. Des fournisseurs de messagerie de premier plan, comme Gmail et Outook, utilisent, avec un succès variable, l'apprentissage automatique depuis plus de dix ans pour filtrer les courriels non désirés ou pernicieux. L'Encadré 3.1 illustre quelques-unes des utilisations possibles de l'IA pour protéger les entreprises des menaces de malveillance.

Encadré 3.1. Utiliser l'IA pour gérer les risques de sécurité numérique dans les entreprises

Beaucoup d'entreprises comme Darktrace ou Vectra appliquent l'apprentissage automatique et l'intelligence artificielle à la détection et à la gestion des attaques de sécurité numérique en temps réel. Le produit *Enterprise Immune System* de Darktrace n'a besoin d'aucune expérience préalable de menace pour comprendre les dangers potentiels. Les algorithmes d'IA découvrent par itération les schémas récurrents et les principes de cohérence propres à un réseau pour déceler les menaces émergentes qui, autrement, ne seraient pas repérées. D'un point de vue méthodologique, le système de Darktrace est analogue au système immunitaire humain, qui apprend le fonctionnement normal du corps, identifie automatiquement les situations anormales, et neutralise les menaces.

De son côté, Vectra a développé une plateforme automatisée toujours active et en apprentissage, appelée *Cognito Platform*, qui débusque les attaquants dans les environnements en nuage. Cette plateforme rend pleinement visibles les comportements des attaquants dans le nuage et les flux de travaux des centres de données vers les utilisateurs et les terminaux connectés. Il est ainsi de plus en plus difficile pour les attaquants de se cacher.

Source : www.darktrace.com/ ; https://vectra.ai/.

L'erreur humaine est fréquente en codage informatique. On estime à neuf sur dix le nombre d'attaques de sécurité numérique dont l'origine est due à une erreur dans le code logiciel, et ce en dépit du temps important de développement – de 50 % à 75 % – consacré aux tests (FT, 2018[93]). Étant donné les milliards de lignes de code écrites chaque année, et la

réutilisation des routines des bibliothèques logicielles propriétaires constituées par des tiers, la détection et la correction des erreurs de code sont des tâches éprouvantes pour l'œil humain. Des pays comme les États-Unis et la Chine financent des projets de recherche pour concevoir des systèmes d'IA capables de détecter les vulnérabilités de sécurité des logiciels. Des entreprises comme l'éditeur de jeux vidéo Ubisoft commencent à utiliser l'IA pour repérer les erreurs encore présentes dans le code avant son déploiement, ce qui a de fait permis de réduire le temps de test de 20 % (FT, 2018[93]). En pratique, les technologies de vérification du code fondées sur l'IA fonctionnent comme les correcteurs orthographiques et grammaticaux des logiciels de traitement de texte. Mais elles ont la capacité d'apprendre et de gagner en efficacité au fil des utilisations (FT, 2018[93]).

L'IA dans le secteur de la surveillance

Les villes s'équipent en infrastructures numériques, notamment de surveillance. Pour renforcer la sécurité publique, on déploie divers outils fondés sur l'IA. Les caméras intelligentes, par exemple, peuvent détecter une bagarre. Certains dispositifs peuvent aussi automatiquement repérer et enregistrer un tir d'arme à feu, le signaler et en fournir l'emplacement précis. Cette section examine la façon dont l'IA est en passe de transformer radicalement le monde de la surveillance et de la sécurité publique.

La vidéosurveillance est de plus en plus utilisée pour renforcer la sécurité publique. Au Royaume-Uni, une étude récente estime que les vidéos de sécurité ont fourni des éléments de preuve utiles dans 65 % des cas de crimes commis sur le réseau ferré britannique entre 2011 et 2015 quand un film de vidéosurveillance était disponible (Ashby, 2017[94]). Or, le nombre impressionnant de caméras de surveillance – 245 millions dans le monde en 2014 – signifie que le volume des données produites ne cesse de croître : de 413 pétaoctets (Po) d'information par jour en 2013, on est passé à environ 860 Po par jour en 2017 (Jenkins, 2015[95]) ; (Civardi, 2017[96]). Or, les capacités humaines ne suffisent pas à traiter de telles quantités d'informations. C'est pourquoi on utilise des technologies de l'IA pour traiter ces données massives et automatiser les processus mécaniques de détection et de contrôle. L'IA permet en outre aux systèmes de sécurité de repérer les crimes et d'intervenir en temps réel Encadré 3.2.

Encadré 3.2. Surveillance avec des caméras « intelligentes »

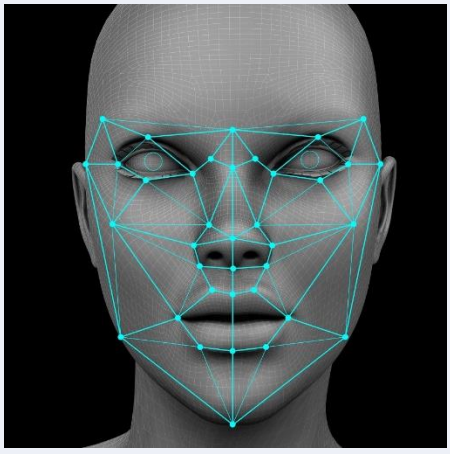
En France, le Commissariat à l'énergie atomique et aux énergies alternatives (CEA) en partenariat avec Thales, utilise des algorithmes d'apprentissage profond pour analyser et interpréter automatiquement les vidéos d'applications de sécurité. Un module de détection d'événement violent repère automatiquement les interactions violentes telles qu'une bagarre ou une agression filmées par des caméras de télévision en circuit fermé et alerte les opérateurs en temps réel. Un autre module aide à localiser les auteurs sur le réseau de caméras. Ces applications sont actuellement évaluées par deux sociétés françaises de transport en commun, la RATP et la SNCF, à Châtelet-Les Halles et Gare du Nord, deux des stations parisiennes de métro et de train les plus fréquentées. La municipalité française de Toulouse utilise aussi des caméras intelligentes pour signaler les comportements inhabituels et les bagages abandonnés dans les lieux publics. Des projets analogues sont en cours d'expérimentation à Berlin, Rotterdam et Shanghai.

Source : Démonstrations et informations fournies à l'OCDE par CEA Tech et Thales en 2018. Pour de plus amples informations, consulter la page www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/11/projet_voie_videoprotection_ouverte_et_integree_appel_a_projets.pdf.

Encadré 3.3. La reconnaissance faciale comme outil de surveillance

Les technologies de reconnaissance faciale sont de plus en plus utilisées à des fins de surveillance par les acteurs publics et privés (Graphique 3.4). L'IA améliore les systèmes traditionnels en permettant une identification plus rapide et plus précise dans les cas où ces derniers ne pourraient être efficaces, par exemple quand la luminosité est insuffisante ou quand la personne qu'on cherche à apercevoir est en partie cachée par quelque chose. Des entreprises comme FaceFirst ont combiné des outils de reconnaissance faciale et d'IA pour proposer des solutions de prévention du vol, de la fraude et de la violence. La conception de solutions de ce genre répond à des considérations spécifiques, dans le respect des plus hauts niveaux de sécurité et de protection de la vie privée : anti-profilage pour empêcher la discrimination, chiffrement des données des images, et intervalles de temps définis de façon stricte pour la purge des données. On retrouve ces outils de surveillance dans de nombreux secteurs, qui vont du commerce de détail (pour empêcher le vol à l'étalage) au secteur bancaire (pour empêcher l'usurpation d'identité), aux services de police (sécurité des frontières), à la gestion événementielle (pour reconnaître des personnes interdites d'accès) et aux casinos (pour repérer les personnalités).

Graphique 3.4. Illustration d'un logiciel de reconnaissance faciale



Source : www.facefirst.com.

Étant donné la dualité de l'IA, les outils de surveillance qui l'exploitent pourraient être utilisés à des fins illégitimes allant à l'encontre des principes décrits au chapitre 4. L'IA est légitime notamment quand il s'agit d'activités des services de police dans le cadre d'enquêtes criminelles, pour détecter et stopper les crimes dès que possible, ou quand il s'agit de lutte antiterroriste. Les technologies de reconnaissance vocale se sont révélées pertinentes à cet égard (Encadré 3.3). Cependant, l'impact de l'IA sur la surveillance va au-delà des systèmes de reconnaissance faciale. Elle contribue aussi de plus en plus à améliorer les technologies de reconnaissance sans visage, c'est-à-dire les techniques qui fondent l'identification sur d'autres informations relatives à la personne (la taille, le type de vêtements, la corpulence, la posture, etc.). De plus, l'IA est efficace lorsqu'on l'associe à des techniques de retouche numérique : on entraîne des réseaux de neurones, en leur présentant des millions d'images, à reconnaître les caractéristiques usuelles d'éléments physiques tels que la peau, les cheveux ou même les briques d'un mur. Le système peut ensuite reconnaître ces caractéristiques dans de nouvelles images et y ajouter des détails et des textures au moyen des connaissances

précédemment acquises. On peut alors combler les lacunes des images de mauvaise résolution et améliorer l'efficacité des systèmes de surveillance (Medhi, Scholkopf et Hirsch, 2017[97]).

L'IA dans le secteur public

L'IA peut être utilisée à des fins très diverses par les administrations publiques. Ses applications ont déjà un impact sur les modalités de fonctionnement du secteur public et l'élaboration des politiques au service des citoyens et des entreprises. Les services concernés sont notamment les services de santé, de transport et de sécurité[15].

Les pouvoirs publics des pays de l'OCDE expérimentent des solutions et mettent en œuvre des projets dans le but de mieux répondre, avec l'IA, aux besoin des utilisateurs des services publics. Ils veulent aussi améliorer la gestion de leurs ressources (par exemple, pour réduire le temps que les agents de la fonction public consacrent à l'assistance à la clientèle et aux tâches administratives). Justement, les technologies de l'IA pourraient augmenter l'efficience et la qualité de nombreuses procédures du secteur public. Par exemple, elles pourraient donner aux citoyens la possibilité de participer dès le début aux processus de conception d'un service et d'interagir avec les services de l'État de façon plus souple, efficace et personnalisée. Correctement conçues et déployées, elles pourraient être intégrées au processus d'élaboration des politiques dans son ensemble, soutenir les réformes du secteur public, et améliorer la productivité du secteur public.

Certains pays ont déjà déployé des systèmes d'IA pour renforcer leurs programmes médico-sociaux. L'IA permettrait par exemple d'optimiser les niveaux d'inventaire des bureaux des services d'action sanitaire et sociale. En effet, grâce à des technologies d'apprentissage automatique, on pourrait analyser les données des transactions et faire des prédictions de plus en plus précises concernant les réapprovisionnements. Cela faciliterait en retour la prévision et l'élaboration des politiques. Au Royaume-Uni, les pouvoirs publics utilisent aussi l'IA pour détecter la fraude aux prestations sociales (Marr, 2018[98]).

L'IA en association avec réalité augmentée et réalité virtuelle

Avec les technologies de l'IA et des tâches de reconnaissance visuelle à haut niveau comme la classification d'images et la détection d'objets, les entreprises développent des matériels et des logiciels de réalité augmentée et de réalité virtuelle. Ces nouveaux produits proposent de nouvelles formes d'expériences d'immersion, d'enseignement et de formation, d'aide aux personnes en situation de handicap ou encore de divertissement. La réalité augmentée et la réalité virtuelle se sont considérablement améliorées depuis le premier prototype de casque de réalité virtuelle mis au point en 1968 par Ivan Sutherland pour le visionnage d'images en 3D. Trop lourd à porter, le casque devait être fixé au plafond (Günger et Zengin, 2017[99]). Aujourd'hui, les entreprises de réalité virtuelle fournissent des expériences de flux vidéo (*streaming*) à 360 degrés avec des casques beaucoup plus légers. S'agissant de réalité augmentée, Pokemon GO a attiré l'attention des consommateurs en 2016 et les attentes restent très élevées. Des applications avec IA intégrée sont déjà commercialisées. IKEA propose à sa clientèle une appli mobile permettant de se rendre compte de l'effet et de la position d'un meuble dans un espace donné avec une précision pouvant aller jusqu'à 1 millimètre (Jesus, 2018[100]). Certaines entreprises technologiques développent aussi des applications pour les personnes malvoyantes[16].

L'IA pour des applications de réalité augmentée/virtuelle interactives

Le développement de la réalité augmentée et de la réalité virtuelle s'accompagne de celui de l'IA qui offre le moyen de les rendre interactives, avec des contenus plus attractifs et intuitifs. Les technologies de l'IA permettent aux applications de réalité augmentée/virtuelle de détecter les mouvements d'une personne, en particulier de ses yeux et de ses mains, et de les interpréter avec une très haute précision permettant de personnaliser le contenu en temps réel en fonction de la réaction de la personne (Lindell, 2017[101]). En associant IA et réalité virtuelle, on peut par exemple savoir quand l'utilisateur observe une portion spécifique de l'espace, et n'y afficher un contenu à haute résolution que dans ce cas précis. On économise ainsi des ressources de calcul, on réduit les délais et on évite les pertes d'images (Hall, 2017[102]). Le développement symbiotique des technologies de l'IA, de la réalité augmentée et de la réalité virtuelle devrait se faire dans divers domaines comme la recherche marketing, les simulations de formation, ou encore l'éducation (Kilpatrick, 2018[103]), (Stanford, 2016[104]).

Des systèmes de réalité virtuelle pour entraîner l'IA

Il faut de grandes quantités de données pour entraîner certains systèmes d'IA. C'est pourquoi l'indisponibilité des données demeure un problème important. Par exemple, les systèmes d'IA des voitures autonomes doivent être entraînés à la gestion de situations critiques, mais il existe peu de données réelles sur des situations telles que des enfants qui traversent une rue en courant. L'autre solution serait donc de développer des réalités numériques : les systèmes d'IA seraient entraînés dans des environnements simulés par ordinateur, qui reproduisent fidèlement les caractéristiques pertinentes du monde réel. De tels environnements simulés pourraient aussi servir à valider la performance des systèmes d'IA (« examen de permis de conduire » pour les véhicules autonomes, par exemple) (Slusallek, 2018[105]).

Ce type d'application intéresse d'autres domaines que celui des transports. De fait, une équipe de recherche a conçu une plateforme appelée *Household Multimodal Environment* (HoME) pour simuler un environnement où entraîner des robots ménagers. Cette plateforme est actuellement dotée d'une base de données de plus de 45 000 configurations diverses d'habitations en 3D, grâce à laquelle elle peut créer un environnement réaliste dans lequel des agents artificiels peuvent apprendre par l'intermédiaire de la vision, de l'audition, de la sémantique, de la physique et de l'interaction avec des objets et d'autres agents (Brodeur et al., 2017[106]).

En permettant aux systèmes d'IA d'apprendre de façon empirique (essai et erreur), les simulations de réalité virtuelle dans le nuage seraient idéales pour des sessions d'entraînement, en particulier d'entraînement aux situations de crise. Le développement des technologies infonuagiques devrait aider à donner corps à de tels environnements. En octobre 2017, NVIDIA a annoncé la création d'un simulateur de réalité virtuelle dans le nuage, capable de répliquer avec précision les lois de la physique dans des environnements représentant le monde réel. Il est probable que l'on mettra au point un nouveau terrain d'entraînement des systèmes d'IA dans quelques années (Solotko, 2017[107]).

Références

Aletras, N. et al. (2016), « Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective », *PeerJ Computer Science*, vol. 2, p. e93, http://dx.doi.org/10.7717/peerj-cs.93. [86]

Angwin, J. et al. (2016), « Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks », *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [83]

Ashby, M. (2017), « The value of CCTV surveillance cameras as an investigative tool: An empirical analysis », *European Journal on Criminal Policy and Research*, vol. 23/3, pp. 441-459, http://dx.doi.org/10.1007/s10610-017-9341-6. [94]

Barocas, S. et A. Selbst (2016), « Big data's disparate impact », *California Law Review*, vol. 104, pp. 671-729, http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf. [74]

Bavitz, C. et K. Hessekiel (2018), *Algorithms and Justice: Examining the Role of the State in the Development and Deployment of Algorithmic Technologies*, Berkman Klein Center for Internet and Society, https://cyber.harvard.edu/story/2018-07/algorithms-and-justice. [82]

Berg, T. et al. (2018), « On the Rise of FinTechs – Credit Scoring Using Digital Footprints », *NBER Working Paper Series*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3163781. [32]

Bloom, N. et al. (2017), « Are ideas getting harder to find? », *document de travail*, n° 23782, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w23782. [52]

Bösch, P. et al. (2018), « Cost-based analysis of autonomous mobility services », *Transport Policy*, vol. 64, pp. 76-91. [3]

Bose, A. et al. (2016), « The VEICL Act: Safety and security for modern vehicles », *Willamette Law Review*, vol. 53, p. 137. [15]

Brayne, S., A. Rosenblat et D. Boyd (2015), « Predictive Policing, Data & Civil Rights: A New Era of Policing and Justice », *Pennsylvania Law Review*, vol. 163/327, http://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf. [76]

Brodeur, S. et al. (2017), « HoME: A Household Multimodal Environment », *arXiv, Cornell University,* 1107, https://arxiv.org/abs/1711.11017. [106]

Brodmerkel, S. (2017), « Dynamic pricing: Retailers using artificial intelligence to predict top price you'll pay », *ABC News,* 27 juin, http://www.abc.net.au/news/2017-06-27/dynamic-pricing-retailers-using-artificial-intelligence/8638340. [50]

Brundage, M. et al. (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Centre for a New American Security, Electronic Frontier Foundation and Open AI, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf. [108]

Burgess, M. (2018), « UK police are using AI to make custodial decisions but it could be discriminating against the poor », *WIRED* 1 mars, http://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit. [79]

Butler, K. et al. (2018), « Machine learning for molecular and materials science », *Nature*, vol. 559/7715, pp. 547-555, http://dx.doi.org/10.1038/s41586-018-0337-2. [60]

Carey, N. et P. Lienert (2018), « Honda to invest $2.75 billion in GM's self-driving car unit », *Reuters,* 3 octobre, https://www.reuters.com/article/us-gm-autonomous/honda-buys-in-to-gm-cruise-self-driving-unit-idUSKCN1MD1GW. [7]

CFPB (2017), « CFPB explores impact of alternative data on credit access for consumers who are credit invisible », Consumer Financial Protection Bureau, https://www.consumerfinance.gov/about-us/newsroom/cfpb-explores-impact-alternative-data-credit-access-consumers-who-are-credit-invisible/. [37]

Cheng, E. (2017), « Just 10% of trading is regular stock picking, JPMorgan estimates », *CNBC,* 13 June, https://www.cnbc.com/2017/06/13/death-of-the-human-investor-just-10-percent-of-trading-is-regular-stock-picking-jpmorgan-estimates.html. [44]

Chintamaneni, P. (26 juin 2017), *How banks can use AI to reduce regulatory compliance burdens*, digitally.cognizant Blog, https://digitally.cognizant.com/how-banks-can-use-ai-to-reduce-regulatory-compliance-burdens-codex2710/. [41]

Chow, M. (2017), « AI and machine learning get us one step closer to relevance at scale », *Google*, https://www.thinkwithgoogle.com/marketing-resources/ai-personalized-marketing/. [45]

Christin, A., A. Rosenblat et D. Boyd (2015), *Courts and Predictive Algorithms*, Data & Civil Rights, A New Era of Policing and Justice, New York University, http://www.law.nyu.edu/sites/default/files/upload_documents/Angele%20Christin.pdf. [81]

Civardi, C. (2017), *Video Surveillance and Artificial Intelligence: Can A.I. Fill the Growing Gap Between Video Surveillance Usage and Human Resources Availability?*, Balzano Informatik, http://dx.doi.org/10.13140/RG.2.2.13330.66248. [96]

CMU (2015), « Uber, Carnegie Mellon announce strategic partnership and creation of advanced technologies center in Pittsburgh », *Carnegie Mellon University News,* 2 février, https://www.cmu.edu/news/stories/archives/2015/february/uber-partnership.html. [8]

Comfort, S. et al. (2018), « Sorting through the safety data haystack: Using machine learning to identify individual case safety reports in social-digital media », *Drug Safety*, vol. 41/6, pp. 579-590, https://www.ncbi.nlm.nih.gov/pubmed/29446035. [64]

Cooke, A. (2017), *Digital Earth Australia*, exposé présenté à la conférence AI: Intelligent Machines, Smart Policies, Paris, les 26 et 27 octobre 2017, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-cooke.pdf. [26]

CSF (2017), *Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications*, Conseil de stabilité financière, Bâle. [31]

De Jesus, A. (2018), « Augmented reality shopping and artificial intelligence – Near-term applications », *Emerj,* 18 décembre, https://www.techemergence.com/augmented-reality-shopping-and-artificial-intelligence/. [51]

Fagnant, D. et K. Kockelman (2015), « Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations », *Transportation Research A: Policy and Practice*, vol. 77, pp. 167-181, https://www.sciencedirect.com/science/article/pii/S0. [2]

FAO (2017), « Can artificial intelligence help improve agricultural productivity? », *e-agriculture,* 19 décembre, http://www.fao.org/e-agriculture/news/can-artificial-intelligence-help-improve-agricultural-productivity. [20]

FAO (2009), *Comment nourrir le monde en 2050*, Organisation des Nations Unies pour l'alimentation et l'agriculture, Rome, http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf. [21]

FDA (2018), *FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems*, Food and Drug Administration, News Release 11 avril 2018, https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm. [66]

Ferguson, A. (2014), « Big Data and Predictive Reasonable Suspicion », *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.2394683. [75]

FIT (2018), *Safer Roads with Automated Vehicles?*, Forum international des transports, https://www.itf-oecd.org/sites/default/files/docs/safer-roads-automated-vehicles.pdf. [12]

Flores, A., K. Bechtel et C. Lowenkamp (2016), *False positives, false negatives, and false analyses: A rejoinder to "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks«* , Federal probation, 80. [84]

Fridman, L. (8 octobre 2018), « Tesla autopilot miles », MIT Human-Centered AI Blog, https://hcai.mit.edu/tesla-autopilot-miles/. [17]

Fridman, L. et al. (2018), « MIT autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation », *arXiv, Cornell University*, vol. 30/September, https://arxiv.org/pdf/1711.06976.pdf. [18]

FT (2018), « US and China back AI bug-detecting projects », *Financial Times, Cyber Security and Artificial Intelligence,* 28 septembre, https://www.ft.com/content/64fef986-89d0-11e8-affd-da9960227309. [93]

Gartner (2017), « Gartner's worldwide security spending forecast », Gartner, 7 décembre, https://www.gartner.com/newsroom/id/3836563. [90]

Gordon, M. et V. Stewart (2017), « CFPB insights on alternative data use on credit scoring », *Law 360,* 3 May, https://www.law360.com/articles/919094/cfpb-insights-on-alternative-data-use-in-credit-scoring. [36]

Günger, C. et K. Zengin (2017), *A Survey on Augmented Reality Applications using Deep Learning*, https://www.researchgate.net/publication/322332639_A_Survey_On_Augmented_Reality_Applications_Using_Deep_Learning. [99]

Hall, N. (2017), « 8 ways AI makes virtual & augmented reality even more real, », *Topbots,* 13 mai, https://www.topbots.com/8-ways-ai-enables-realistic-virtual-augmented-reality-vr-ar/. [102]

Higgins, T. et C. Dawson (2018), « Waymo orders up to 20,000 Jaguar SUVs for driverless fleet », *Wall Street Journal,* 27 mars, https://www.wsj.com/articles/waymo-orders-up-to-20-000-jaguar-suvs-for-driverless-fleet-1522159944. [6]

Hinds, R. (2018), *How Natural Language Processing is shaping the Future of Communication*, MarTechSeries, Marketing Technology Insights, 5 février, https://martechseries.com/mts-insights/guest-authors/how-natural-language-processing-is-shaping-the-future-of-communication/. [46]

Hong, P. (27 août 2017), « Using machine learning to boost click-through rate for your ads », LinkedIn Blog, https://www.linkedin.com/pulse/using-machine-learning-boost-click-through-rate-your-ads-tay/. [48]

Horgan, J. (2008), « Against prediction: Profiling, policing, and punishing in an actuarial age – by Bernard E. Harcourt », *Review of Policy Research*, vol. 25/3, pp. 281-282, http://dx.doi.org/10.1111/j.1541-1338.2008.00328.x. [73]

Hutson, M. (2017), « Artificial intelligence prevails at predicting Supreme Court decisions », *Science Magazine,* 2 mai, http://www.sciencemag.org/news/2017/05/artificial-intelligence-prevails-predicting-supreme-court-decisions. [87]

Hu, X. (dir. pub.) (2017), « Human-in-the-loop Bayesian optimization of wearable device parameters », *PLOS ONE*, vol. 12/9, p. e0184054, http://dx.doi.org/10.1371/journal.pone.0184054. [61]

IHS (2017), « Global defence spending to hit post-Cold War high in 2018 », *IHS Markit* 18 décembre, https://ihsmarkit.com/research-analysis/global-defence-spending-to-hit-post-cold-war-high-in-2018.html. [89]

Inners, M. et A. Kun (2017), *Beyond Liability: Legal Issues of Human-Machine Interaction for Automated Vehicles*, Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, septembre, pp. 245-253, http://dx.doi.org/10.1145/3122986.3123005. [14]

Institut canadien d'information sur la santé (2013), « Une meilleure information pour une meilleure santé : vision de l'utilisation des données pour les besoins du système de santé au Canada », en collaboration avec Inforoute Santé du Canada, https://www.cihi.ca/sites/default/files/hsu_vision_report_fr_0.pdf. [63]

ISACA (2016), *The State of Cybersecurity: Implications for 2016*, An ISACA and RSA Conference Survey, Cybersecurity Nexus, https://www.isaca.org/cyber/Documents/state-of-cybersecurity_res_eng_0316.pdf. [91]

Jagtiani, J. et C. Lemieux (2019), « The roles of alternative data and machine learning in Fintech lending: Evidence from the LendingClub Consumer Platform », *Document de travail*, n° 18-15, Federal Reserve Bank of Philadelphia, http://dx.doi.org/10.21799/frbp.wp.2018.15. [30]

Jenkins, N. (2015), « 245 million video surveillance cameras installed globally in 2014 », *IHS Markit, Market Insight,* 11 June, https://technology.ihs.com/532501/245-million-video-surveillance-cameras-installed-globally-in-2014. [95]

Jesus, A. (2018), « Augmented reality shopping and artificial intelligence – near-term applications », *Emerj,* 12 décembre, https://www.techemergence.com/augmented-reality-shopping-and-artificial-intelligence/. [100]

Joh, E. (2017), « The undue influence of surveillance technology companies on policing », *New York University Law Review*, vol. 91/101, http://dx.doi.org/10.2139/ssrn.2924620. [77]

Jouanjean, M. (2019), « Digital opportunities for trade in the agriculture and food sectors », *Documents de l'OCDE sur l'alimentation, l'agriculture et les pêcheries*, n° 122, Éditions OCDE, Paris, https://doi.org/10.1787/91c40e07-en. [22]

Kehl, D., P. Guo et S. Kessler (2017), *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing, Responsive Communities Initiative*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. [80]

Kilpatrick, S. (2018), « The rising force of deep learning in VR and AR », *Logik,* 28 mars, https://www.logikk.com/articles/deep-learning-in-vr-and-ar/. [103]

King, R. et al. (2004), « Functional genomic hypothesis generation and experimentation by a robot scientist », *Nature*, vol. 427/6971, pp. 247-252, http://dx.doi.org/10.1038/nature02236. [56]

Kleinberg, J. et al. (2017), « Human decisions and machine predictions », *document de travail*, n° 23180, National Bureau of Economic Research, Cambridge, MA. [85]

Lee, C., D. Baughman et A. Lee (2017), « Deep learning is effective for classifying normal versus age-related macular degeneration OCT images », *Opthamology Retina*, vol. 1/4, pp. 322-327. [67]

Lee, T. (2018), « Fully driverless Waymo taxis are due out this year, alarming critics », *Ars Technica,* 1 octobre, https://arstechnica.com/cars/2018/10/waymo-wont-have-to-prove-its-driverless-taxis-are-safe-before-2018-launch/. [10]

Lindell, T. (2017), « Augmented reality needs AI in order to be effective », *AI Business,* 6 novembre, https://aibusiness.com/holographic-interfaces-augmented-reality/. [101]

Lippert, J. et al. (2018), « Toyota's vision of autonomous cars is not exactly driverless », *Bloomberg Business Week,* 19 septembre, https://www.bloomberg.com/news/features/2018-09-19/toyota-s-vision-of-autonomous-cars-is-not-exactly-driverless. [5]

Marr, B. (2018), « How AI and machine learning are transforming law firms and the legal sector », *Forbes,* 23 May, https://www.forbes.com/sites/bernardmarr/2018/05/23/how-ai-and-machine-learning-are-transforming-law-firms-and-the-legal-sector/#7587475832c3. [88]

Marr, B. (2018), « How the UK government uses artificial intelligence to identify welfare and state benefits fraud », *Forbes,* 29 October, https://www.forbes.com/sites/bernardmarr/2018/10/29/how-the-uk-government-uses-artificial-intelligence-to-identify-welfare-and-state-benefits-fraud/#f5283c940cb9. [98]

Mar, V. et H. Soyer (2018), « Artificial intelligence for melanoma diagnosis: How can we deliver on the promise? », *Annals of Oncology*, vol. 29/8, pp. 1625-1628, http://dx.doi.org/10.1093/annonc/mdy193. [68]

Matchar, E. (2017), « AI plant and animal identification helps us all be citizen scientists », *Smithsonian.com,* 7 June, https://www.smithsonianmag.com/innovation/ai-plant-and-animal-identification-helps-us-all-be-citizen-scientists-180963525/. [55]

Mateescu, A. et al. (2015), *Social Media Surveillance and Law Enforcement, New Era of Criminal Justice and Policing*, Data Civil Rights, http://www.datacivilrights.org/pubs/2015-1027/Social_Media_Surveillance_and_Law_Enforce. [78]

Medhi, S., B. Scholkopf et M. Hirsch (2017), « EnhanceNet: Single image super-resolution through automated texture synthesis », *arrXiv, Cornell University,* 1612.07919, https://arxiv.org/abs/1612.07919. [97]

MIT (2018), « Cybersecurity's insidious new threat: Workforce stress », *MIT Technology Review,* 7 August, https://www.technologyreview.com/s/611727/cybersecuritys-insidious-new-threat-workforce-stress/. [92]

OCDE (2018), « Artificial intelligence and machine learning in science », *OECD Science, Technology and Innovation Outlook 2018*, Éditions OCDE, Paris. [57]

OCDE (2018), *Base de données pour l'analyse structurelle (STAN)*, Rév. 4, divisions 49-53, consulté en janvier 2018, http://www.oecd.org/fr/industrie/ind/stanbasededonneespourlanalysestructurelle.htm. [1]

OCDE (2018), « Personalised pricing in the digital era – Note by the United Kingdom », Key paper for the joint meeting of the OECD Consumer Protection and Competition committees, OCDE, Paris, http://www.oecd.org/daf/competition/personalised-pricing-in-the-digital-era.htm. [109]

OCDE (2018), *Perspectives de l'économie numérique de l'OCDE 2017*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264282483-fr. [24]

OCDE (2018), *Science, technologie et innovation : Perspectives de l'OCDE 2018 (version abrégée) : S'adapter aux bouleversements technologiques et sociétaux*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-fr. [53]

OCDE (2017), *Technology and Innovation in the Insurance Sector*, Éditions OCDE, Paris, https://www.oecd.org/finance/Technology-and-innovation-in-the-insurance-sector.pdf (consulté le 28 août 2018). [39]

OCDE (2016), *Recommandation du Conseil sur la gouvernance des données de santé*, Éditions OCDE, Paris, https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0433. [71]

OCDE (2015), *Data-Driven Innovation : Big Data for Growth and Well-Being*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264229358-en. [62]

OCDE (2014), « Skills and jobs in the Internet economy », *Documents de travail de l'OCDE sur l'économie numérique*, n° 242, Éditions OCDE, Paris. [19]

OCDE (s.d.), *La gestion des nouvelles technologies de santé : Concilier accès, valeur et viabilité*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/g2g73036-fr. [69]

O'Dwyer, R. (2018), *Algorithms are making the same mistakes assessing credit scores that humans did a century ago*, Quartz, 14 mai 2018, https://qz.com/1276781/algorithms-are-making-the-same-mistakes-assessing-credit-scores-that-humans-did-a-century-ago/. [34]

Ohnsman, A. (2018), « Waymo dramatically expanding autonomous taxi fleet, eyes sales to individuals », *Forbes,* 31 May, https://www.forbes.com/sites/alanohnsman/2018/05/31/waymo-adding-up-to-62000-minivans-to-robot-fleet-may-supply-tech-for-fca-models. [4]

ORAD (2016), « Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles », On-Road Automated Driving (ORAD) Committee, SAE International, http://dx.doi.org/10.4271/j3016_201609. [9]

Ornstein, C. et K. Thomas (2018), *Sloan Kettering's Cozy Deal With Start-Up Ignites a New Uproar*, The New York Times, 20 septembre 2018, https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html. [70]

Patton, E. (2018), *Integrating Artificial Intelligence for Scaling Internet of Things in Health Care*, OCDE-GCOA-Cornell-Tech Expert Consultation on Growing and Shaping the Internet of Things Wellness and Care Ecosystem, 4-5 octobre, New York. [65]

Plummer, L. (2017), « This is how Netflix's top-secret recommendation system works », *WIRED,* 22 August, https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like. [47]

Press, G. (2017), « Equifax and SAS leverage AI and deep learning to improve consumer access to credit », *Forbes,* 20 February, https://www.forbes.com/sites/gilpress/2017/02/20/equifax-and-sas-leverage-ai-and-deep-learning-to-improve-consumer-access-to-credit/2/#2ea15ddd7f69. [29]

Rakestraw, R. (2017), « Can artificial intelligence help feed the world? », *Forbes,* 6 September, https://www.forbes.com/sites/themixingbowl/2017/09/05/can-artificial-intelligence-help-feed-the-world/#16bb973646db. [23]

Roeland, C. (2017), *EC Perspectives on the Earth Observation*, exposé présenté à la conférence AI: Intelligent Machines, Smart Policies, Paris, les 26 et 27 octobre 2017, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-roeland.pdf. [25]

Rollet, C. (2018), « The odd reality of life under China's all-seeing credit score system », *WIRED,* 5 June, https://www.wired.co.uk/article/china-blacklist. [35]

Segler, M., M. Preuss et M. Waller (2018), « Planning chemical syntheses with deep neural networks and symbolic AI », *Nature*, vol. 555/7698, pp. 604-610, http://dx.doi.org/10.1038/nature25978. [59]

Simon, M. (2017), « Phone-powered AI spots sick plants with remarkable accuracy », *WIRED,* 2 February, https://www.wired.com/story/plant-ai/. [28]

Slusallek, P. (2018), *Artificial Intelligence and Digital Reality: Do We Need a CERN for AI?*, The Forum Network, OCDE, Paris, https://www.oecd-forum.org/channels/722-digitalisation/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai. [105]

Sohangir, S. et al. (2018), « Big data: Deep learning for financial sentiment analysis », *Journal of Big Data*, vol. 5/1, http://dx.doi.org/10.1186/s40537-017-0111-6. [40]

Sokolin, L. et M. Low (2018), *Machine Intelligence and Augmented Finance: How Artificial Intelligence Creates $1 Trillion Dollar of Change in the Front, Middle and Back Office*, Autonomous Research LLP, https://next.autonomous.com/augmented-finance-machine-intelligence. [38]

Solotko, S. (2017), « Virtual reality is the next training ground for artificial intelligence », *Forbes,* 11 October, https://www.forbes.com/sites/tiriasresearch/2017/10/11/virtual-reality-is-the-next-training-ground-for-artificial-intelligence/#6e0c59cc57a5. [107]

Song, H. (2017), « JPMorgan software does in seconds what took lawyers 360,000 hours », *Bloomberg.com,* 28 February, https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance. [42]

Spangler, S. et al. (2014), *Automated Hypothesis Generation based on Mining Scientific Literature*, ACM Press, New York, http://dx.doi.org/10.1145/2623330.2623667. [54]

Stanford (2016), *Artificial Intelligence and Life in 2030*, AI100 Standing Committee and Study Panel, Stanford University, https://ai100.stanford.edu/2016-report. [104]

Surakitbanharn, C. et al. (2018), *Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles (CATV)*, Purdue University, https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf. [16]

Voegeli, V. (2016), « Credit Suisse, CIA-funded palantir to target rogue bankers », *Bloomberg,* 22 March, https://www.bloomberg.com/news/articles/2016-03-22/credit-suisse-cia-funded-palantir-build-joint-compliance-firm. [43]

Waid, B. (2018), « AI-enabled personalization: The new frontier in dynamic pricing », *Forbes,* 9 July, https://www.forbes.com/sites/forbestechcouncil/2018/07/09/ai-enabled-personalization-the-new-frontier-in-dynamic-pricing/#71e470b86c1b. [49]

Walker-Smith (2013), « Automated vehicles are probably legal in the United States », *Texas A&M Law Review*, vol. 1/411. [11]

Webb, L. (2017), *Machine Learning in Action*, exposé présenté à la conférence AI: Intelligent Machines, Smart Policies, Paris, les 26 et 27 octobre 2017, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-webb.pdf. [27]

Welsch, D. et E. Behrmann (2018), « Who's winning the self-driving car race? », *Bloomberg,* 7 May, https://www.bloomberg.com/news/articles/2018-05-07/our-world-s-fragile-cities-need-a-78-trillion-boost. [13]

Williams, K. et al. (2015), « Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases », *Journal of The Royal Society Interface*, vol. 12/104, pp. 20141289-20141289, http://dx.doi.org/10.1098/rsif.2014.1289. [58]

Wyllie, D. (2013), « How 'big data' is helping law enforcement », *PoliceOne.Com* 20 August, https://www.policeone.com/police-products/software/Data-Information-Sharing-Software/articles/6396543-How-Big-Data-is-helping-law-enforcement/. [72]

Zeng, M. (2018), *Alibaba and the Future of Business*, Harvard Business Review, septembre-octobre 2018, https://hbr.org/2018/09/alibaba-and-the-future-of-business. [33]

Notes

[1] Base de données STAN pour l'analyse structurelle, 2018, valeur ajoutée des services « Transports et entreposage », CITI Rev. 4 Divisions 49 à 53, en pourcentage de la valeur ajoutée totale, moyenne non pondérée de 2016 dans la zone OCDE. La moyenne pondérée de 2016 dans la zone OCDE était de 4.3 %.

[2] D'après https://www.crunchbase.com/.

[3] La notation FICO a été créée en 1989 par l'entreprise Fair, Isaac and Company (FICO). Elle est toujours utilisée par la majorité des banques et des établissements de prêt.

[4] Le Comité de la politique à l'égard des consommateurs (CPC) de l'OCDE a adopté la définition de la personnalisation des prix donnée par l'*Office of Fair Trading* du Royaume-Uni, à savoir : « la pratique consistant pour une entreprise à utiliser l'information fournie volontairement, recueillie par observation ou obtenue par inférence au sujet du comportement des individus ou de leurs caractéristiques personnelles pour différencier les prix entre consommateurs (individuellement ou collectivement sur la base de certaines catégories) en fonction de ce que l'entreprise estime être leur consentement à payer » (OCDE, 2018[109]). Appliquée par les vendeurs, cette pratique pourrait conduire certains clients à payer moins pour un bien ou un service donné, et d'autres à payer plus que ce qu'ils auraient dû payer si le prix proposé avait été le même pour tout le monde.

[5] Cette section s'inspire de travaux du Comité de la politique scientifique et technologique (CPST) de l'OCDE, en particulier le chapitre 5 de la publication OCDE (2018[53]), consacré à l'intelligence artificielle et à l'apprentissage automatique, dont les principaux auteurs sont Ross D. King, de l'Université de Manchester, et Stephen Roberts, de l'Université d'Oxford.

[6] Voir https://iris.ai/.

[7] Voir https://www.atomwise.com/2015/03/24/new-ebola-treatment-using-artificial-intelligence/.

[8] Voir https://www.bbc.com/news/technology-40592520.

[9] Voir https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html.

[10] *State of Wisconsin v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

[11] *Loomis v. Wisconsin*, 137 S.Ct. 2290 (2017).

[12] Les dépenses publiques consacrées aux technologies de l'IA à des fins de défense ont leur importance, mais ce domaine d'étude n'entre pas dans le champ de la présente publication.

[13] Sauf mention contraire, la présente publication utilise le terme « sécurité numérique » pour désigner la gestion des risques économiques et sociaux résultant d'infractions à l'accessibilité, l'intégrité ou la confidentialité des technologies de l'information et de la communication, et des données.

[14] Le hameçonnage est une pratique frauduleuse consistant à tenter d'obtenir d'une cible des informations sensibles par messagerie électronique en lui faisant croire à tort qu'elle est en contact avec un intermédiaire de confiance. Demandant plus de travail, le harponnage est une variante de l'hameçonnage consistant à adapter spécifiquement la méthode à l'individu ou l'organisation visée en collectant et en utilisant des informations sensibles telles que le nom, le sexe, l'affiliation, etc. (Brundage et al., 2018[108]). Le harponnage est le vecteur d'infection le plus courant : il était à l'origine de 71 % des cyberattaques en 2017.

[15] Tel que confirmé par le groupe e-Leaders, rattaché au Comité de la gouvernance publique de l'OCDE. Son groupe thématique sur les technologies émergentes – composé de représentants de 16 pays – s'intéresse principalement à l'IA et à la chaîne de blocs.

[16] BlindTool (https://play.google.com/store/apps/details?id=the.blindtool&hl=en) et Seeing AI (https://www.microsoft.com/en-us/seeing-ai) en sont des exemples.

4. CONSIDÉRATIONS DE POLITIQUE PUBLIQUE

Le présent chapitre examine les considérations de politique publique à prendre en compte pour la mise en place de systèmes d'IA dignes de confiance et centrés sur l'humain. Il y est question des enjeux liés à l'éthique et à l'équité ; du respect des valeurs humaines et démocratiques, notamment celui de la vie privée ; et des risques de transposition des biais existant dans le monde analogique vers le monde numérique, à l'instar des préjugés sexistes et racistes. La nécessité de faire évoluer les systèmes d'IA pour les rendre plus fiables, sûrs, sécurisés et transparents et les doter de mécanismes permettant de déterminer clairement les responsabilités quant aux résultats obtenus y est soulignée.

La promotion de systèmes d'IA dignes de confiance passe, notamment, par des politiques ayant pour effet de favoriser l'investissement dans la recherche et le développement responsables ; de créer un écosystème numérique où la protection de la vie privée n'est pas menacée par un élargissement de l'accès aux données ; de permettre aux petites et moyennes entreprises de prospérer ; d'encourager la concurrence sans porter atteinte à la propriété intellectuelle ; et d'aider les travailleurs à passer d'un emploi à l'autre au gré de l'évolution du monde du travail.

Une IA centrée sur l'humain

L'IA joue un rôle de plus en plus prépondérant. Plus cette technologie se diffuse, plus les répercussions que peuvent avoir, sur la vie des individus, les prévisions qu'elle établit, les recommandations qu'elle formule ou les décisions qu'elle prend deviennent importantes. La communauté technique, les entreprises et les responsables politiques cherchent activement la meilleure façon d'obtenir une IA centrée sur l'humain et digne de confiance, qui présente un maximum d'avantages pour un minimum de risques et recueille l'adhésion de la société.

Encadré 4.1. Les systèmes d'IA fonctionnant comme des « boîtes noires » posent de nouveaux défis par rapport aux progrès technologiques précédents

Les réseaux neuronaux sont souvent qualifiés de « boîtes noires ». S'il est effectivement possible d'observer le comportement de ces systèmes, il existe une différence considérable entre les réseaux neuronaux et les technologies précédentes en matière de possibilité d'observation, d'où l'expression « boîtes noires ». Les réseaux neuronaux fonctionnent par itération des données sur lesquelles ils sont entraînés. Ils établissent des corrélations probabilistes complexes à plusieurs variables qui deviennent constitutives du modèle qu'ils construisent. Toutefois, ils n'indiquent pas comment les données peuvent être liées entre elles (Weinberger, 2018[1]). Les données sont bien trop complexes pour être appréhendées par le cerveau humain. Les caractéristiques qui distinguent l'IA des avancées technologiques antérieures et nuisent à la transparence et à la détermination des responsabilités sont entre autres les suivantes :

- **La possibilité d'exploration** : Les algorithmes basés sur des règles peuvent être lus et vérifiés règle par règle, ce qui permet de trouver relativement facilement certains types d'erreurs. En revanche, certaines catégories de systèmes d'apprentissage automatique, notamment les systèmes neuronaux, consistent uniquement en des relations mathématiques abstraites entre différents facteurs. Ces systèmes peuvent s'avérer extrêmement complexes et difficiles à comprendre, mêmes pour ceux qui les programment et les entraînent (OCDE, 2016).
- **Le caractère évolutif** : Certains systèmes d'apprentissage automatique fonctionnent en boucle et évoluent avec le temps, ils peuvent même modifier leur comportement de manière imprévue.
- **Le faible niveau de reproductibilité** : Il est possible que le système d'apprentissage automatique n'établisse une prévision spécifique ou ne prenne une décision particulière que dans certaines conditions et avec certaines données, lesquelles ne sont pas nécessairement reproductibles.
- **Davantage de tensions dans la protection des données personnelles et sensibles** :
  - **Les inférences** : Même en l'absence de données protégées ou sensibles, les systèmes d'IA peuvent être capables de déduire ces données et d'établir des corrélations à partir de variables indirectes qui ne sont ni personnelles ni sensibles, telles qu'un historique d'achats ou des données de localisation (Kosinski, Stillwell et Graepel, 2013[2]).
  - **Les variables indirectes indésirables** : Les stratégies politiques et techniques de protection de la vie privée et de non-discrimination ont tendance à limiter les données collectées, à interdire l'utilisation de certaines données ou à supprimer certaines données pour en empêcher l'utilisation. Or, un système d'IA peut baser une prévision sur des données indirectes ayant un lien étroit

avec des données interdites et non-collectées. Qui plus est, la seule façon de détecter ces données indirectes est de collecter également les données sensibles ou personnelles telles que la race. Si de telles données sont collectées, alors il devient essentiel de garantir qu'elles seront toujours utilisées de façon appropriée.

- **Le paradoxe données-vie privée** : dans le cas de nombreux systèmes d'IA, augmenter la quantité des données d'entraînement peut améliorer la précision de leurs prévisions et contribuer à réduire les risques de biais dus à des échantillons faussés. Cependant, plus le volume de données collectées est élevé, plus la vie privée des individus concernés est menacée.

Certains types d'IA – qui reçoivent souvent l'appellation de « boîtes noires » – posent de nouveaux défis par rapport aux progrès technologiques précédents (Encadré 4.1). Au vu de ces défis, l'OCDE – s'appuyant sur les travaux du Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO) – a défini les grandes priorités à suivre pour une IA centrée sur l'humain. Premièrement, celle-ci doit contribuer à une croissance inclusive et durable ainsi qu'au bien-être. Deuxièmement, elle doit respecter les valeurs centrées sur l'humain et l'équité. Troisièmement, son utilisation et le fonctionnement de ses systèmes doivent être transparents. Quatrièmement, les systèmes d'IA doivent être fiables et sûrs. Cinquièmement, il convient de déterminer les responsabilités quant aux résultats des prévisions établies par l'IA et des décisions qui en ont découlé. Ces mesures sont considérées comme essentielles pour les prévisions qui emportent des enjeux élevés. Elles sont aussi importantes pour les recommandations commerciales ou pour les utilisations de l'IA de moindre conséquence.

Croissance inclusive et durable et bien-être

L'IA recèle un formidable potentiel à mettre au service des Objectifs de développement durable

L'IA peut contribuer au bien commun et à la réalisation des Objectifs de développement durable (ODD) des Nations Unies dans des domaines tels que l'éducation, la santé, le transport, l'agriculture et les villes durables, entre autres. De nombreuses organisations publiques et privées, dont la Banque mondiale, plusieurs institutions des Nations Unies et l'OCDE, œuvrent pour que l'IA soit mise au service de la réalisation des ODD.

Développer une IA équitable et ouverte à tous devient une priorité croissante

Développer une IA équitable et ouverte à tous devient une priorité croissante. Cela est d'autant plus vrai dans la mesure où l'on redoute que l'IA aggrave les inégalités ou creuse les écarts existant au sein des pays et entre pays développés et pays en développement. Ces écarts résultent de la concentration des ressources en IA – technologies, compétences, ensembles de données et puissance de calcul – dans quelques entreprises et pays. D'autres inquiétudes concernent le fait que l'IA puisse perpétuer des préjugés (Talbot et al., 2017[3]). Certains craignent qu'elle ait un impact différent sur les populations vulnérables et sous-représentées, à savoir, entre autres, les personnes moins instruites, les personnes peu qualifiées, les femmes et les personnes âgées, en particulier dans les pays à revenu faible ou intermédiaire (Smith et Neupane, 2018[4]). Le Centre canadien de recherches pour le développement international a récemment recommandé la mise en place d'un fonds mondial baptisé « l'IA au service du développement ». Ce fonds permettrait la création», dans les pays à revenu faible ou intermédiaire, de « centres d'excellence en IA chargés d'appuyer l'élaboration et

l'exécution de politiques inclusives fondées sur des données probantes (Smith et Neupane, 2018[4]). L'objectif est de veiller à ce que les retombées de l'IA soient réparties de manière équitable et conduisent à des sociétés plus égalitaires. Les initiatives en faveur d'une IA inclusive visent à garantir un large partage des gains économiques générés par l'IA au sein de la société, pour que personne ne soit laissé de côté.

L'IA inclusive et durable intéresse tout particulièrement des pays comme l'Inde (NITI, 2018[5]) ; des entreprises comme Microsoft[1] ; et des groupes d'universitaires comme le Berkman Klein Center à Harvard. Ainsi, Microsoft a lancé, entre autres projets, l'application mobile « Seeing AI ». Cette application à l'usage des personnes malvoyantes scanne et reconnaît tous les éléments de leur environnement direct, et en fournit une audiodescription. Par ailleurs, Microsoft investit actuellement deux millions de dollars des États-Unis dans des initiatives permettant d'utiliser l'IA pour répondre à des enjeux de durabilité, par exemple en matière de préservation de la biodiversité et de lutte contre le changement climatique (Heiner et Nguyen, 2018[6]).

Valeurs centrées sur l'humain et équité

Droits de l'homme et codes d'éthique

Le droit international des droits de l'homme consacre des normes éthiques

Le droit international des droits de l'homme consacre des normes éthiques. L'IA peut favoriser le respect des droits de l'homme tout comme elle peut créer de nouveaux risques de violation, délibérée ou accidentelle, de ces droits. Avec les structures juridiques et autres structures institutionnelles connexes, le droit relatif aux droits de l'homme peut aussi constituer l'un des outils au service d'une IA centrée sur l'humain (Encadré 4.2).

Encadré 4.2. Les droits de l'homme et l'IA

Le droit international des droits de l'homme internationaux renvoie à un corpus juridique international, incluant la Charte internationale des droits de l'homme[1], ainsi qu'aux systèmes régionaux de protection des droits de l'homme élaborés au cours des 70 dernières années à travers le monde. Les droits de l'homme fournissent une série de normes minimales universelles fondées, entre autres, sur les valeurs de dignité humaine, d'autonomie et d'égalité, dans le cadre de l'État de droit. Ces normes et les mécanismes juridiques qui y sont associés font que les pays sont tenus en droit de respecter et protéger les droits de l'homme et d'en garantir la pleine jouissance. Ils exigent en outre que ceux qui ont été privés de leurs droits ou dont les droits ont été violés disposent d'un recours effectif.

Les droits de l'homme incluent notamment le droit à l'égalité, le droit à la non-discrimination, le droit à la liberté d'association, le droit à la vie privée ainsi que divers droits économiques, sociaux et culturels tels que le droit à l'éducation ou le droit à la santé.

Des instruments intergouvernementaux récents, tels que les *Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme* (HCDH, 2011[7]) traite aussi du rôle des acteurs privés dans le contexte des droits de l'homme, les investissant d'une « responsabilité » quant au respect de ces droits. En outre, la version mise à jour en 2011 des *Principes directeurs de l'OCDE à l'intention des entreprises multinationales* (OCDE, 2011[8]), recueil de recommandations adressées par les gouvernements aux entreprises, contient un chapitre consacré aux droits de l'homme.

Les droits de l'homme recoupent des préoccupations éthiques plus larges et d'autres domaines de réglementation en rapport avec l'IA, tels que la protection des données personnelles ou la législation relative à la sécurité des produits. Toutefois, ces autres préoccupations et questions ont souvent une portée différente.

1. La Charte internationale des droits de l'homme comprend la Déclaration universelle des droits de l'homme, le Pacte international relatif aux droits civils et politiques, et le Pacte international relatif aux droits économiques, sociaux et culturels.

L'IA promet de faire grandir le respect des droits de l'homme

Compte tenu de l'ampleur potentielle de ses applications et utilisations, l'IA promet de faire progresser la protection et le respect des droits de l'homme. Elle pourrait ainsi servir à analyser les ressorts des pénuries alimentaires pour mieux lutter contre la faim, à améliorer les diagnostics et les traitements médicaux ainsi qu'à accroître la disponibilité et l'accessibilité des services de santé, et à dévoiler en plein jour les discriminations.

L'IA pourrait aussi desservir la cause des droits de l'homme

L'IA peut aussi poser plusieurs problèmes dans le domaine des droits de l'homme, ces problèmes étant souvent mentionnés lors des débats dont elle fait l'objet et, plus généralement, des débats d'éthique. Certains systèmes d'IA, ou l'usage qui est en fait, pourraient constituer une violation, accidentelle ou non, des droits de l'homme. L'aspect accidentel est particulièrement étudié. Les algorithmes d'apprentissage automatique qui prédisent la récidive, par exemple, peuvent présenter un biais non détecté. Néanmoins, il arrive aussi que des technologies d'IA soient associées à des atteintes délibérées aux droits de l'homme. Ainsi en est-il lorsqu'elles servent, par exemple, à traquer des dissidents politiques, à museler la liberté d'expression ou encore à restreindre la participation des individus à la vie politique. Dans ces cas-là, la violation en elle-même ne réside pas tout entière dans l'utilisation de l'IA. Toutefois, elle pourrait être aggravée par la technicité et l'efficacité de celle-ci.

L'utilisation de l'IA peut également poser des problèmes inédits lorsque ses effets sur les droits de l'homme ne sont pas voulus ou sont difficiles à déceler. Cela peut tenir à l'utilisation de données d'entraînement de mauvaise qualité, à la façon dont le système est conçu ou à la complexité des interactions entre le système d'IA et son environnement. L'exacerbation par les algorithmes des discours haineux ou des incitations à la violence sur l'internet en est un exemple. Un autre exemple est l'amplification non intentionnelle des fausses nouvelles, qui peut avoir des répercussions sur le droit à prendre part à la vie politique et aux affaires publiques. L'ampleur et l'incidence probables du préjudice seront fonction de celles que peuvent avoir les décisions prises par un système d'IA donné. Par exemple, une décision prise par un système d'IA de recommandation d'actualités a une incidence potentielle plus faible que la décision d'un algorithme qui prédit le risque de récidive des détenus en liberté conditionnelle.

Les codes d'éthique de l'IA complètent les cadres relatifs aux droits de l'homme

Les codes d'éthique peuvent parer au risque que l'IA puisse ne pas être centrée sur l'humain ou ne pas être en adéquation avec les valeurs humaines. Les entreprises privées comme les gouvernements ont adopté un grand nombre de codes d'éthique en lien avec l'IA.

Par exemple, l'entreprise DeepMind, qui appartient à Google, a créé en octobre 2017 une unité consacrée à l'éthique (DeepMind Ethics & Society)[2]. L'unité a pour but d'aider les

technologues à comprendre les incidences éthiques de leur travail et d'aider la société à décider en quoi l'IA peut lui être profitable. En outre, l'unité financera des recherches externes sur, entre autres, le biais algorithmique, l'avenir du travail ou les armes létales autonomes. L'entreprise Google a elle aussi annoncé la mise en place d'une série de principes éthiques destinés à guider ses recherches, le développement de ses produits et ses décisions commerciales[3]. Elle a publié un livre blanc sur la gouvernance de l'IA, dans lequel elle met en évidence les points à éclaircir avec les gouvernements et les sociétés civiles[4]. La philosophie de Microsoft en matière d'IA est de « développer l'ingéniosité humaine grâce à une technologie intelligente » (Heiner et Nguyen, 2018[9]). L'entreprise a lancé des projets visant à garantir un développement inclusif et durable.

Avec ses mécanismes institutionnels et son architecture globale, le droit relatif aux droits de l'homme fournit l'orientation et l'assise nécessaires pour garantir un développement et une utilisation de l'IA en société qui soient éthiques et centrés sur l'humain.

Le recours aux cadres relatifs aux droits de l'homme dans le contexte de l'IA offre des avantages

Le recours aux cadres relatifs aux droits de l'homme dans le contexte de l'IA offre différents avantages, notamment de par les institutions en place, la jurisprudence, le langage universel et la reconnaissance internationale qui entourent ces cadres :

- **Les institutions en place** : Une vaste infrastructure internationale, régionale et nationale a été mise en place au fil du temps dans le domaine des droits de l'homme. Elle est composée d'organisations intergouvernementales, de tribunaux, d'organisations non gouvernementales, d'universités, ainsi que d'autres institutions et communautés dans le cadre desquelles il est possible d'invoquer les droits de l'homme et d'exercer un recours.
- **La jurisprudence** : En tant que normes juridiques, les valeurs protégées par les droits de l'homme reçoivent leur traduction concrète, et sont rendues juridiquement contraignantes, dans des situations spécifiques grâce à la jurisprudence et au travail d'interprétation des institutions internationales, régionales et nationales.
- **Un langage universel** : Les droits de l'homme fournissent un langage universel pour une problématique internationale. Associé à l'infrastructure relative aux droits de l'homme, ce langage peut aider à autonomiser un plus large éventail de parties prenantes. Celles-ci peuvent ainsi participer au débat sur la place de l'IA dans la société aux côtés d'acteurs intervenant directement dans le cycle de vie de cette technologie.
- **Une reconnaissance internationale** : Les droits de l'homme bénéficient d'une reconnaissance et d'une légitimité importantes au niveau international. Qu'un acteur passe seulement pour les enfreindre et il y aura probablement de lourdes conséquences, puisque sa réputation en sera sans doute passablement écornée.

Une approche de l'IA basée sur les droits de l'homme peut aider à identifier les risques, les priorités, les groupes vulnérables et à proposer des solutions

- **Identification des risques** : Les cadres relatifs aux droits de l'homme peuvent aider à identifier les risques de préjudice. En particulier, ils peuvent servir à mettre en œuvre des études sur la diligence raisonnable en matière de droits de l'homme , par exemple des études d'impact sur les droits de l'homme (Encadré 4.3).

- **Exigences fondamentales** : En tant que normes minimales, les droits de l'homme définissent des exigences fondamentales inviolables. Par exemple, dans le cadre de la réglementation relative à l'expression sur les réseaux sociaux, la jurisprudence en matière de droits de l'homme aide à faire des discours haineux une limite à ne pas franchir.
- **Identification des situations à haut risque** : Les droits de l'homme peuvent s'avérer utiles pour repérer les situations ou activités à haut risque. Dans de tels cas, il convient de redoubler d'attention à moins que l'on estime qu'il n'est pas approprié de recourir à l'IA.
- **Identification des groupes ou des communautés vulnérables** : Les droits de l'homme peuvent aider à identifier les groupes ou communautés vulnérables ou à risque en lien avec l'IA. Certains individus ou communautés peuvent être sous-représentés en raison, par exemple, d'une utilisation limitée des smartphones.
- **Réparation** : En tant que normes juridiques assorties d'obligations, les droits de l'homme peuvent assurer une réparation à ceux à qui l'on a fait du tort. Ces réparations incluent, par exemple, une cessation d'activité, l'élaboration de nouveaux processus ou politiques, des excuses ou une indemnité pécuniaire.

Encadré 4.3. Les études d'impact sur les droits de l'homme

Les études d'impact sur les droits de l'homme peuvent aider à mettre en évidence des risques que les acteurs intervenant au cours du cycle de vie de l'IA n'auraient pas nécessairement prévus sans cela. À cette fin, elles portent davantage sur les effets connexes sur l'homme que sur l'optimisation de la technologie ou de ses produits. Ces études, ou des processus similaires, pourraient garantir le respect des droits de l'homme dès la conception de la technologie et tout au long de son cycle de vie.

Les études d'impact sur les droits de l'homme mesurent un grand nombre des effets que la technologie peut avoir sur les droits de l'homme, et ce dans le cadre d'une démarche de grande ampleur nécessitant beaucoup de ressources. Il est probablement plus simple de partir du système d'IA en question. De cette façon, l'étude ne porte que sur un nombre limité de domaines où l'on a le plus de chances de constater des problèmes en matière de droits. Les organisations du secteur peuvent contribuer à la réalisation études d'impact pour le compte de petites et moyennes entreprises (PME) ou pour celui d'entreprises non technologiques qui utilisent des systèmes d'IA sans forcément maîtriser la technologie. La *Global Network Initiative* est l'une de ces organisations qui œuvrent au respect de la liberté d'expression et à la protection de la vie privée. Elle aide des entreprises à planifier des études sur les droits de l'homme et à les intégrer dans leurs projets de nouveaux produits (https://globalnetworkinitiative.org/).

Les études d'impact sur les droits de l'homme présentent l'inconvénient d'être généralement exécutées entreprise par entreprise, alors même que les systèmes d'IA peuvent impliquer de nombreux acteurs. De ce fait, il peut s'avérer inefficace de n'étudier qu'une seule composante. Microsoft a été la première grande entreprise technologique à mener à bien une étude d'impact de l'IA en 2018.

D'autre part, la mise en œuvre d'une approche de l'IA basée sur les droits de l'homme se heurte à d'importantes difficultés, lesquelles sont liées au fait que les droits de l'homme s'adressent aux États, que leur garantie dépend des pays et territoires, qu'ils sont mieux adaptés pour remédier à des préjudices majeurs causés à un petit groupe d'individus et qu'ils peuvent coûter cher aux entreprises :

- **Les droits de l'homme s'adressent aux États, non aux acteurs privés**, or les acteurs du secteur privé jouent un rôle clé dans la recherche sur l'IA comme dans le développement et le déploiement de systèmes fondés sur cette technologie. Cette difficulté n'est pas propre à l'IA. Plusieurs initiatives intergouvernementales cherchent à combler le fossé entre les secteurs public et privé. Au-delà de ces efforts, il est de plus en plus généralement admis que les entreprises ont tout intérêt à respecter les droits de l'homme[5].
- **La garantie des droits de l'homme dépend des pays et territoires**. En général, la partie requérante doit démontrer qu'elle a qualité pour agir dans un pays ou sur un territoire donné. Cette démarche n'est peut-être pas optimale lorsque sont mis en causes de grandes entreprises multinationales et des systèmes d'IA couvrant de multiples pays et territoires.
- **Les droits de l'homme sont mieux adaptés pour remédier à des préjudices majeurs causés à un petit groupe d'individus**, qu'à des préjudices moins importants subis par un grand nombre d'individus. En outre, les droits de l'homme et leurs structures peuvent sembler opaques aux non-initiés.
- **Dans certains contextes, les droits de l'homme ont la réputation de coûter cher aux entreprises**. En conséquence, les démarches qui mettent en avant l'éthique, la protection des consommateurs ou la conduite responsable des entreprises, ainsi que les arguments économiques en faveur du respect des droits de l'homme, semblent prometteuses.

Certains des défis généraux posés par l'IA, tels que la transparence et l'explicabilité, concernent aussi le respect des droits de l'homme (voir ci-après la section « Transparence et explicabilité »). Sans transparence, il est difficile de repérer les violations des droits de l'homme ou d'étayer une plainte pour violation. Il en va de même pour ce qui est de demander réparation, de déterminer les liens de causalité et d'établir les responsabilités.

La protection des données personnelles

L'IA défie les notions de « données personnelles » et de consentement

L'IA est de plus en plus capable d'établir des liens entre différents ensembles de données et de faire coïncider différents types d'information, ce qui a de graves conséquences. Les données conservées séparément étaient autrefois considérées comme non personnelles (ou, s'étant vu retirer tout élément d'identification, elles avaient été « anonymisées »). Cependant, l'IA est capable de croiser ces données non personnelles avec d'autres données et de les réattribuer ensuite aux individus concernés, ce qui en fait à nouveau des données personnelles (ou « désanonymisées »). Ainsi, la corrélation algorithmique fragilise la distinction entre les données personnelles et les autres données. Les données non personnelles peuvent de plus en plus servir à ré-identifier des individus ou à déduire des informations sensibles les concernant, au-delà de celles qu'ils avaient divulgués de leur plein gré à l'origine (Cellarius, 2017[10]). Par exemple, en 2007, des chercheurs avaient déjà utilisé des données dites anonymes pour associer la liste des films loués sur Netflix avec des avis publiés sur le site IMDB. Ce faisant, ils ont identifié les personnes ayant loué des films et ont eu accès à l'historique complet de leurs locations. L'augmentation du nombre de données collectées et les progrès technologiques vont de plus en plus permettre d'établir ce type de rapprochements. Il devient difficile de déterminer quelles données peuvent être considérées comme non personnelles et le resteront.

Il est toujours plus difficile de distinguer les données sensibles des données non sensibles, comme l'illustre le règlement général de l'Union européenne sur la protection des données (RGPD). Certains algorithmes parviennent à déduire des informations sensibles à partir de données « non sensibles », ainsi ceux qui déterminent l'état émotionnel d'individus à la manière dont ils tapent sur leur clavier (Privacy International et Article 19, 2018[11]). L'utilisation de l'IA pour identifier ou ré-identifier des données initialement non personnelles ou anonymisées représente aussi un problème sur le plan juridique. Les garde-fous en place, tels que la *Recommandation du Conseil de l'OCDE concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel* (ci-après les « Lignes directrices relatives à la vie privée »), s'appliquent aux données personnelles (Encadré 4.4). En conséquence, il n'est pas clairement établi si, ou à quel moment, ces cadres incluent dans leur périmètre les données qui, dans certaines circonstances, seraient, ou pourraient être, identifiables (Office of the Victorian Information Commissioner, 2018[12]). Une interprétation extrême pourrait élargir considérablement le champ de la protection de la vie privée, mais rendrait du même coup cette protection difficile à assurer dans les faits.

Encadré 4.4. Les Lignes directrices de l'OCDE relatives à la vie privée

La *Recommandation du Conseil concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel* (ci-après dénommée les « Lignes directrices relatives à la vie privée ») a été adoptée en 1980 et actualisée en 2013 (OCDE, 2013[13]). Elle contient des définitions de termes pertinents en ce domaine, et notamment celle des « données de caractère personnel », entendues comme « toute information relative à une personne physique identifiée ou identifiable (personne concernée) ». Elle définit également les principes devant régir le traitement de ces données. Ces principes ont trait à la limitation en matière de collecte (avec, lorsqu'il y a lieu, le consentement des individus comme garantie), à la qualité des données, à la spécification des finalités, à la limitation de l'utilisation, aux garanties de sécurité, à la transparence, à la participation individuelle et à la responsabilité. En outre, la Recommandation dispose que, lors de la mise en œuvre des Lignes directrices relatives à la vie privée, les membres doivent veiller à ce que les personnes concernées ne fassent l'objet d'aucune discrimination déloyale. La mise en œuvre des Lignes directrices relatives à la vie privée devait faire l'objet d'une révision en 2019 pour que soient pris en considération, entre autres, les avancées récentes, notamment celles réalisées dans le domaine de l'IA.

L'IA défie également les principes de protection des données personnelles concernant la limitation en matière de collecte, la limitation de l'utilisation et la spécification des finalités

Pour entraîner et optimiser les systèmes d'IA, les algorithmes d'apprentissage automatique ont besoin d'énormes quantités de données, ce qui incite à en maximiser la collecte plutôt qu'à la freiner. Avec l'utilisation croissante des dispositifs fondés sur l'IA et de l'internet des objets (IdO), cette collecte à la fois est plus abondante, plus fréquente et plus simple. Ces données sont reliées à d'autres données, parfois plus ou moins à l'insu des personnes concernées ou sans leur consentement.

Les tendances identifiées et l'évolution de « l'apprentissage » sont difficiles à anticiper. En conséquence, la collecte et l'utilisation de données peuvent aller au-delà de ce que savait initialement la personne concernée, de ce qui lui avait été communiqué et de ce à quoi elle

avait consenti (Privacy International et Article 19, 2018[11]). Cela est potentiellement incompatible avec les principes de limitation en matière de collecte, de limitation de l'utilisation et de spécification des finalités énoncés dans les Lignes directrices relatives à la vie privée (Cellarius, 2017[10]). Les deux premiers principes reposent en partie sur le consentement de la personne concernée (selon le cas, étant donné qu'il n'est pas possible de recueillir le consentement dans certaines situations). Ce consentement est le point de départ de la collecte de données à caractère personnel ou de leur utilisation à des fins autres que celles initialement indiquées. Les technologies d'IA telles que l'apprentissage profond, qui sont difficiles à comprendre ou à surveiller, sont également difficiles à expliquer aux personnes concernées. Cela constitue un défi pour les entreprises. Elles indiquent qu'il est compliqué de concilier la vitesse à laquelle l'IA accède à des données, les analyse et les utilise, qui augmente de manière exponentielle, avec ces principes de protection des données (OCDE, 2018[14]).

Ces difficultés sont exacerbées par l'association des technologies liées à l'IA avec les progrès de l'IdO, c'est-à-dire la connexion à internet d'un nombre croissant d'appareils et d'objets avec le temps. Le fait que les technologies d'IA et celles de l'IdO soient de plus en plus souvent associées (avec, par exemple, des dispositifs de l'IdO dotés d'IA, ou des algorithmes d'IA utilisés pour analyser les données de l'IdO) entraîne la collecte constante de données toujours plus nombreuses, et notamment de données personnelles. Ces données peuvent être de plus en plus facilement croisées entre elles et analysées. D'une part, les appareils qui recueillent des informations sont toujours plus nombreux (comme les caméras de surveillance ou les véhicules autonomes), d'autre part, la technologie liée à l'IA s'est améliorée (c'est le cas, par exemple, de la reconnaissance faciale). Combinées, ces deux tendances risquent de donner lieu à des résultats plus intrusifs que chaque facteur pris séparément (Office of the Victorian Information Commissioner, 2018[12]).

L'IA peut aussi renforcer la participation et le consentement des individus

L'IA a le potentiel de renforcer les données personnelles. Par exemple, des initiatives visant à élaborer des systèmes d'IA reposant sur les principes de la protection de la vie privée dès la conception et de la protection de la vie privée par défaut sont en cours au sein de plusieurs organismes de normalisation technique. Pour la plupart, ces organismes utilisent et adaptent des lignes directrices relatives à la vie privée, dont celles de l'OCDE. En outre, l'IA est utilisée pour offrir aux individus des services personnalisés adaptés à leurs besoins, basés sur leurs préférences de confidentialité telles qu'acquises au fil du temps (Office of the Victorian Information Commissioner, 2018[12]). Ces services peuvent aider les individus à s'y retrouver parmi les différentes politiques de traitement des données personnelles et à s'assurer que leurs préférences sont prises en considération partout. Dans ce cas, l'IA facilite le consentement éclairé et la participation des individus. À titre d'exemple, une équipe de chercheurs a mis au point Polisis, structure automatisée qui utilise les classificateurs d'un réseau neuronal pour analyser les politiques de confidentialité (Harkous, 2018[15]).

Équité et éthique

Les algorithmes d'apprentissage automatique peuvent refléter les biais implicites de leurs données d'entraînement

À ce jour, les initiatives stratégiques relatives à l'IA donnent une place prépondérante aux questions d'éthique, d'équité et/ou de justice. La propension des algorithmes d'apprentissage automatique à refléter et à reproduire les biais implicites de leurs données d'entraînement, tels que les biais raciaux et les associations stéréotypées, suscitent de nombreuses inquiétudes.

Parce que les artefacts technologiques incarnent souvent des valeurs sociales, les débats sur l'équité doivent établir clairement à quelles sociétés les technologies doivent profiter, qui doit être protégé et grâce à quelles valeurs fondamentales (Flanagan, Howe et Nissenbaum, 2008[16]). Des disciplines telles que la philosophie, le droit et l'économie sont aux prises depuis des décennies avec diverses conceptions de l'équité qui correspondent à autant d'éclairages différents, illustrant toute la diversité des interprétations que l'on peut donner de cette notion et des implications qu'elle peut avoir dans le champ politique.

Les notions philosophiques, juridiques et informatiques de l'équité et d'une IA éthique varient

La philosophie met l'accent sur les concepts de bonne et mauvaise conduites, de bien et de mal, et de morale. Trois grandes théories philosophiques sont dignes d'attention dans le contexte d'une IA éthique (Abrams et al., 2017[17]) :

- **L'approche basée sur les droits fondamentaux de l'homme**, associée à Emmanuel Kant, définit les principes formels de l'éthique, qui sont des droits spécifiques tels que le respect de la vie privée ou la liberté. Elle protège ces principes au moyen de réglementations que les systèmes d'IA doivent respecter.
- **L'approche utilitariste**, suivie par Jeremy Bentham et John Stuart Mill, met l'accent sur les politiques publiques qui maximisent le bien-être humain en se basant sur des analyses de rentabilité économique. S'agissant de l'IA, l'approche utilitariste soulève la question de savoir *qui* doit voir son bien-être maximisé (par exemple, les individus, la famille, la société ou les institutions/gouvernements), la réponse pouvant influer sur la conception des algorithmes.
- **L'approche fondée sur l'éthique de la vertu**, inspirée de la philosophie d'Aristote, est axée sur les valeurs et normes éthiques dont une société a besoin afin d'aider les individus dans leurs efforts quotidiens pour vivre une vie qui vaut la peine d'être vécue. Cette approche soulève la question de savoir quelles sont les valeurs et les normes éthiques qui garantissent une protection.

Dans le droit, les termes « égalité » et « justice » sont souvent utilisés pour désigner les concepts d'équité. Les deux grandes approches juridiques de l'équité sont l'équité individuelle et l'équité de groupe.

- **L'équité individuelle** correspond à la notion d'égalité devant la loi. Elle implique que tous les individus doivent être traités sur un pied d'égalité et ne pas subir de discriminations au regard de leurs spécificités. L'égalité fait partie des droits humains internationaux.
- **L'équité de groupe** privilégie l'équité du résultat. Elle veille à ce que le résultat ne diffère pas de façon systématique pour les personnes qui, sur la base d'une caractéristique protégée (telles que la race ou le genre), appartiennent à des groupes différents. L'équité de groupe soutient que les différences et les contextes historiques peuvent conduire des groupes distincts à réagir diversement face à une situation donnée. L'approche de l'équité de groupe diffère considérablement selon les pays. Certains utilisent, par exemple, la discrimination positive.

Les concepteurs de systèmes d'IA ont réfléchi à la façon de traduire l'équité dans leurs systèmes. Aux différentes définitions de l'équité correspondent différentes approches possibles (Narayanan, 2018[18]) :

- **L'approche basée sur « l'ignorance »**, dans le cadre de laquelle un système d'IA doit ignorer tout facteur identifiable, va de pair avec l'approche juridique de l'équité individuelle. Dans ce cas, le système d'IA ne prend pas en considération les données concernant des caractéristiques sensibles ou interdites de traitement, telles que le sexe, la race et l'orientation sexuelle (Yona, 2017[19]). Toutefois, de nombreux autres facteurs peuvent être en corrélation avec une caractéristique dont le traitement est protégé/interdit (comme le sexe), et les supprimer pourrait réduire la précision d'un système d'IA.
- **L'équité basée sur la connaissance** tient compte des différences entre les groupes et vise à traiter des individus similaires de la même manière. Le défi consiste néanmoins à déterminer qui traiter sur un pied d'égalité avec qui. Pour cerner quels individus devraient être considérés comme similaires aux fins d'une tâche particulière, il faut connaître des caractéristiques sensibles.
- **Les approches basées sur l'équité de groupe** s'attachent à garantir que les résultats ne diffèrent pas systématiquement pour les personnes appartenant à des groupes distincts. On craint en effet que les systèmes d'IA puissent être inéquitables, en perpétuant ou en renforçant les biais traditionnels, car ils reposent souvent sur des ensembles de données qui portent la marque des pratiques passées.

Des notions de l'équité différentes donnent des résultats différents pour les divers groupes de la société et les divers types de parties prenantes. Ils ne peuvent tous être atteints simultanément. Ce sont des considérations, et parfois des choix, politiques qui doivent éclairer les choix en matière de conception technologique susceptibles de nuire à des groupes spécifiques.

L'application de l'IA aux ressources humaines donne une illustration des biais qu'elle peut introduire et des problèmes qui en résultent

Dans le domaine des ressources humaines, l'utilisation de l'IA perpétue les biais de recrutement, ou aide au contraire à mettre au jour et à réduire ceux qui ont des effets préjudiciables. Une étude menée par l'université Carnegie Mellon au sujet des tendances observées en ce qui concerne les offres d'emplois publiées sur l'internet a montré qu'une annonce publiée pour un poste de cadre bien rémunéré était présentée 1 816 fois à des hommes et 311 fois seulement à des femmes (Simonite, 2018[20]). Ainsi, un domaine de collaboration potentiel entre les humains et l'IA est la recherche de la transparence des applications de l'IA utilisées pour le recrutement et l'évaluation. Ces applications ne devraient pas codifier de biais, par exemple en disqualifiant automatiquement les candidatures issues de la diversité lorsqu'il s'agit de pourvoir un emploi dans un domaine jusque-là fermé à celle-ci (OCDE, 2018[21]).

Plusieurs approches peuvent aider à réduire la discrimination dans les systèmes d'IA

Les approches proposées pour réduire la discrimination dans les systèmes d'IA incluent la sensibilisation ; les politiques et pratiques organisationnelles relatives à la diversité ; les normes ; les solutions techniques permettant de détecter et de corriger les biais algorithmiques ; et les approches d'auto-réglementation ou de réglementation. Par exemple, dans le cadre des systèmes de prévision policière, certains proposent le recours à des études ou à des déclarations d'impact algorithmique. Il s'agirait, pour les services de police, d'évaluer l'efficacité, les avantages et les éventuels effets discriminatoires de l'ensemble des options technologiques qui s'offrent à eux (Selbst, 2017[22]). La responsabilité et la transparence sont importantes pour atteindre l'équité. Cependant, même combinées, elles ne la garantissent pas (Weinberger, 2018[23]) ; (Narayanan, 2018[18]).

Les efforts visant à atteindre l'équité dans les systèmes d'IA peuvent nécessiter des compromis

On attend des systèmes d'IA qu'ils soient « équitables ». Cela doit se traduire, par exemple, par le fait que seuls les prévenus les plus dangereux restent en prison ou que seul le prêt le plus approprié au regard de la capacité de remboursement soit proposé. Les **erreurs de type I** (ou faux positifs) signalent la classification erronée d'une personne ou d'un comportement. Par exemple, les systèmes peuvent prédire à tort qu'un prévenu récidivera alors qu'il ne le fera pas. De même, ils peuvent se tromper en prédisant une maladie qui n'a pas lieu d'être. Les **erreurs de type II** (ou faux négatifs) se rencontrent dans les cas où un système d'IA prédit à tort, par exemple, qu'un prévenu ne récidivera pas. Un autre exemple serait un test qui indiquerait, à tort, l'absence d'une maladie.

Les approches de l'équité de groupe tiennent compte de points de départ différents selon les groupes. Elles tentent de rendre compte des différences sur le plan mathématique en garantissant une « précision égale » ou un taux d'erreur identique entre tous les groupes. Par exemple, elles classeraient à tort le même pourcentage d'hommes et de femmes en tant que récidivistes (ou égaliseraient les faux positifs et les faux négatifs).

Égaliser les faux positifs et les faux négatifs entraîne une difficulté. Les faux négatifs sont souvent considérés comme plus indésirables et risqués que les faux positifs parce que plus préjudiciables (Berk et Hyatt, 2015[24]). Par exemple, le coût pour une banque d'un prêt fait à un individu dont un système d'IA a prédit qu'il pourrait rembourser – mais qui ne le peut pas – est supérieur au gain tiré de ce prêt. Un individu qu'un mauvais diagnostic a déclaré indemne d'une maladie alors qu'il en est bel et bien atteint peut endurer de grandes souffrances. Égaliser les faux positifs et les faux négatifs peut aussi entraîner des effets indésirables, tels que le fait d'incarcérer des femmes qui ne présentent aucune menace pour la sécurité pour parvenir à libérer la même proportion d'hommes et de femmes (Berk et Hyatt, 2015[24]). Certaines approches visent, par exemple, à égaliser les faux positifs et les faux négatifs en même temps. Cependant, il est difficile de satisfaire à différentes notions d'équité simultanément (Chouldechova, 2016[25]).

Les responsables politiques pourraient réfléchir à un traitement approprié des données sensibles dans le contexte de l'IA

Il pourrait être opportun de revenir sur la question du traitement à réserver aux données sensibles. Dans certains cas, des organisations peuvent avoir besoin de garder et d'utiliser des données sensibles pour assurer que leurs algorithmes ne reconstruisent pas ces données sans que l'on y prenne garde. Une autre priorité d'action est de surveiller chaînes de réaction imprévues. Ainsi, lorsque la police se rend dans des quartiers identifiés par des algorithmes comme ayant une criminalité élevée, cela pourrait conduire à une collecte de données faussées et introduire, par la suite, un biais dans l'algorithme – et dans la société – contre ces quartiers (O'Neil, 2016[26]).

Transparence et explicabilité

La transparence sur l'utilisation de l'IA et le fonctionnement des systèmes d'IA est essentielle

Le terme « transparence » n'a pas la même signification sur les plans technique et politique. Pour les responsables de l'élaboration des politiques, la transparence concerne traditionnellement la façon dont une décision est prise, les participants au processus et les facteurs entrant dans

la prise de décision (Kosack et Fung, 2014[27]). Sous cet angle, les mesures de transparence pourraient révéler comment l'IA est actuellement utilisée dans le cadre d'une prévision, d'une recommandation ou d'une décision. Elles pourraient en outre consister à informer l'usager que son interlocuteur est un système d'IA lorsque tel est le cas.

Pour les technologues, la transparence d'un système d'IA concerne principalement les questions liées aux processus. Il s'agit de permettre aux individus de comprendre comment un système d'IA est développé, entraîné et mis en place. Il peut s'agir également d'expliciter les facteurs qui influent sur une prévision ou une décision spécifique. En général, cela ne passe pas par le partage de code ou d'ensembles de données précis. Dans de nombreux cas, les systèmes sont trop complexes pour que ces éléments apportent une transparence digne de ce nom (Wachter, Mittelstadt et Russell, 2017[28]). En outre, la divulgation de ces renseignements pourrait entraîner celle de secrets commerciaux ou de données sensibles d'utilisateurs.

Plus généralement, on considère qu'il est important de faire connaître et comprendre les systèmes de raisonnement employés en IA pour que cette technologie soit acceptée de tous et utile à tous.

Les approches de la transparence dans les systèmes d'IA

Des experts du *Berkman Klein Center Working Group on Explanation and the Law* (Groupe de travail du Centre Berkman Klein sur l'explication et la législation), de l'Université de Harvard, définissent des approches visant à améliorer la transparence des systèmes d'IA, et notent que chacune implique des compromis (Doshi-Velez et al., 2017[29]). Une approche supplémentaire réside dans la transparence de l'optimisation, c'est-à-dire la transparence sur les objectifs d'un système d'IA et les résultats obtenus en lien avec ceux-ci. Ces approches sont : i) les garanties théoriques ; ii) les données empiriques ; et iii) l'explication (Tableau 4.1).

Tableau 4.1. Approches visant à améliorer la transparence et la responsabilité dans les systèmes d'IA

Approche	Description	Contextes bien adaptés	Contextes peu adaptés
Garanties théoriques	Dans certaines situations, il est possible de donner des garanties théoriques à propos d'un système d'IA étayées par des preuves.	L'environnement est entièrement observable (ex., le jeu de Go) et le problème comme la solution peuvent être formalisés.	La situation ne peut pas être décrite avec précision (la plupart des situations en conditions réelles).
Preuves statistiques/ probabilité	Les données empiriques mesurent la performance globale d'un système, montrant si le système est bénéfique ou néfaste, sans expliquer les décisions particulières.	Les résultats peuvent être entièrement formalisés ; il est acceptable d'attendre de voir les résultats négatifs pour les mesurer ; les problèmes peuvent n'apparaître que dans les agrégats.	L'objectif ne peut pas être entièrement formalisé ; il est possible d'établir les responsabilités à l'égard d'une décision donnée.
Explication	Les humains peuvent interpréter des informations concernant la logique suivie par un système pour traiter un ensemble particulier d'entrées et atteindre une conclusion spécifique.	Les problèmes ne sont pas intégralement spécifiés, les objectifs ne sont pas clairs et les entrées peuvent être erronées.	D'autres formes de responsabilité sont possibles.

Source : adapté de Doshi-Velez et al. (2017[29]), « Accountability of AI under the law: The role of explanation », https://arxiv.org/pdf/1711.01134.pdf.

Certains systèmes offrent des garanties théoriques sur leurs contraintes d'exploitation

Dans certains cas, il est possible d'apporter des **garanties théoriques**, qui indiqueront que le système fonctionnera de manière visible dans le cadre de contraintes bien précises. Les garanties théoriques s'appliquent aux situations dans lesquelles l'environnement est entièrement observable et le problème comme la solution peuvent être intégralement formalisés, comme dans le jeu de Go. Dans de telles situations, certains types de résultats ne peuvent pas être obtenus, même si un système d'IA traite de nouveaux genres de données. Par exemple, un système pourrait être conçu pour suivre, de manière prouvée, les processus définis d'un commun accord pour un vote et le décompte des voix. Dans ce cas, il n'est pas forcément nécessaire d'apporter des explications ou des preuves : le système n'a pas besoin d'expliquer comment il est parvenu à un résultat parce que les types de résultats qui suscitent l'inquiétude sont mathématiquement impossibles. Il est possible de réaliser une étude dès le départ pour déterminer si ces contraintes sont suffisantes.

Des preuves statistiques de la performance globale peuvent être fournies dans certains cas

Dans certains cas, il peut être suffisant de se baser sur les **preuves statistiques** de la performance globale d'un système. Apporter la preuve qu'un système d'IA donné accroît de manière sensible tel bienfait ou tel préjudice pour la société ou les individus peut constituer une garantie de responsabilité suffisante. Par exemple, un système autonome d'atterrissage pour les aéronefs peut causer moins d'incidents de sécurité que des pilotes humains, ou un outil d'aide au diagnostic clinique réduire la mortalité. Les preuves statistiques pourraient constituer un mécanisme de responsabilité approprié pour de nombreux systèmes d'IA, parce que ce mécanisme protège les secrets commerciaux en plus d'être capable de repérer les préjudices répandus mais à faible risque qui n'apparaissent que dans les agrégats (Barocas et Selbst, 2016[30] ; Crawford, 2016[31]). Les questions de biais ou de discrimination peuvent être vérifiées statistiquement : par exemple, un système d'approbation de prêt présenterait un biais s'il approuvait davantage de prêts pour les hommes que pour les femmes lorsque les autres facteurs sont neutralisés. Le taux d'erreur acceptable et l'incertitude tolérée varient selon l'application. Par exemple, le taux d'erreur jugé acceptable pour un outil de traduction ne le sera peut-être pas pour la conduite autonome ou des examens médicaux.

La transparence de l'optimisation est la transparence des objectifs et des résultats d'un système

Une autre approche de la transparence des systèmes d'IA propose que la gouvernance porte son attention non plus sur les moyens mais sur les finalités d'un système. Il s'agit non plus d'exiger l'explicabilité du fonctionnement interne d'un système mais de mesurer ses résultats, c'est-à-dire ce pour quoi le système est « optimisé ». Cela nécessiterait une déclaration concernant ce pour quoi un système d'IA est optimisé, sachant que les optimisations sont imparfaites, qu'elles entraînent des compromis et doivent être limitées par des « contraintes majeures » telles que la sûreté et l'équité. Cette approche préconise d'utiliser les systèmes d'IA pour faire ce pour quoi ils sont optimisés. Elle s'appuie sur les cadres éthiques et juridiques existants, ainsi que sur des débats sociaux et des processus politiques si besoin est pour fournir des informations sur les domaines pour lesquels les systèmes d'IA devraient être optimisés (Weinberger, 2018[1]).

L'explication concerne un résultat précis d'un système d'IA

L'explication est indispensable pour les situations dans lesquelles une anomalie doit être déterminée dans une instance spécifique – situation qui risque de devenir de plus en plus fréquente à mesure que des systèmes d'IA sont mis en place pour formuler des recommandations ou prendre des décisions actuellement laissées à la discrétion de l'homme (Burgess, 2016[32]). Le RGPD exige que les personnes concernées reçoivent des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues des systèmes de prise de décision automatisée. En général, l'explication n'a pas besoin de présenter le processus de prise de décision du système dans sa totalité. La plupart du temps, il suffit de répondre à l'une des questions ci-dessous (Doshi-Velez et al., 2017[29]) :

1. **Les principaux facteurs d'une décision** : Pour toutes sortes de décisions, concernant, par exemple, les audiences relatives à la garde des enfants, les conditions à remplir pour obtenir un prêt ou une mise en liberté provisoire, divers facteurs doivent être pris en considération (ou au contraire formellement proscrits). Dresser une liste des facteurs qui ont compté dans une prévision établie par l'IA – classés, de préférence, par ordre d'importance – peut aider à garantir que les bons facteurs ont été pris en compte.
2. **Les facteurs déterminants, c'est-à-dire les facteurs qui influent de manière décisive sur le résultat** : Il arrive qu'il soit important de savoir si un facteur donné a orienté un résultat. Le fait de changer un facteur donné, tel que la race dans le cadre d'admissions à l'université, peut montrer si le facteur a été utilisé correctement.
3. **Pourquoi deux cas apparemment similaires donnent-ils des résultats différents, ou inversement** ? Il est possible d'évaluer la cohérence et l'intégrité des prévisions basées sur l'IA. Par exemple, le revenu doit être pris en considération lorsqu'il est décidé d'octroyer ou non un prêt, mais il ne saurait être déterminant dans des situations par ailleurs similaires où il n'a pas lieu d'entrer en ligne de compte.

L'explication fait l'objet de recherches actives mais elle entraîne des coûts, et pourrait même nécessiter des compromis

Des recherches techniques sont actuellement menées par des entreprises, des organismes de normalisation, des organisations à but non lucratif et des institutions publiques en vue de la création de systèmes d'IA capables d'expliquer leurs prévisions. Les entreprises travaillant dans des domaines particulièrement réglementés, tels que la finance, la santé et les ressources humaines, cherchent activement à éliminer les éventuels risques financiers, juridiques et d'atteinte à la réputation liés aux prévisions établies par des systèmes d'IA. Par exemple, en 2016, la banque américaine Capital One a constitué une équipe de recherche chargée de trouver des moyens d'améliorer l'explicabilité des techniques d'IA (Knight, 2017[33]). Des entreprises telles que MondoBrain ont conçu des interfaces utilisateur pour aider à expliquer les facteurs significatifs (Encadré 4.5). Des organisations à but non lucratif, telles qu'OpenAI, cherchent des moyens de mettre au point une IA explicable et de vérifier les décisions prises par l'IA. Des recherches financées par les pouvoirs publics sont par ailleurs en cours. Ainsi, la DARPA finance 13 groupes de recherche différents, qui travaillent sur diverses façons d'améliorer l'explicabilité de l'IA.

Dans de nombreux cas, il est possible de générer au moins un de ces types d'explications concernant les résultats des systèmes d'IA. Toutefois, les explications ont un coût. Concevoir un système destiné à fournir une explication peut s'avérer complexe et onéreux. Exiger des explications pour tous les systèmes d'IA peut ne pas être approprié selon leur finalité et peut désavantager les PME en particulier. Les systèmes d'IA doivent souvent être conçus *ex ante* pour fournir un certain type d'explication. Chercher des explications après coup

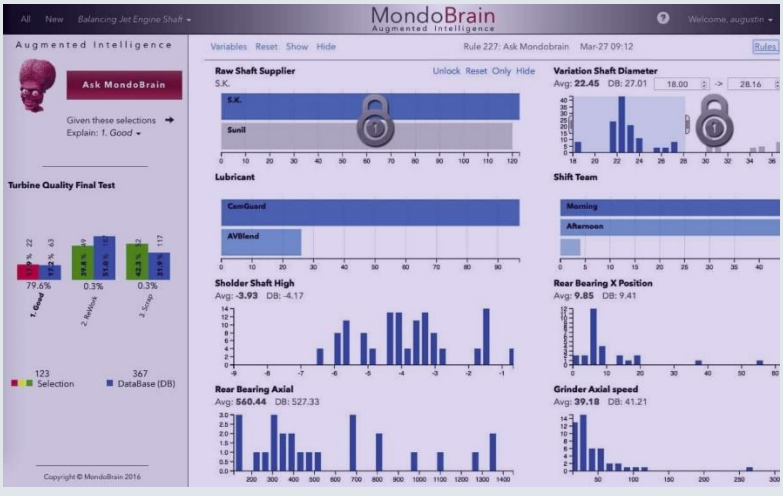
nécessite généralement davantage de travail, potentiellement de recréer l'intégralité du système de décision. Par exemple, un système d'IA ne peut pas fournir une explication pour tous les facteurs majeurs qui ont pesé sur un résultat si sa conception ne lui permet d'en fournir que pour un seul. De même, un système d'IA chargé de dépister les affections cardiaques ne peut être interrogé au sujet des différences de diagnostic entre hommes et femmes si les données ayant servi à son entraînement n'étaient pas ventilées par sexe, et ce même s'il tient bel et bien compte de ce paramètre par le biais de variables indirectes, comme d'autres affections plus fréquentes chez les femmes que chez les hommes.

Encadré 4.5. Régler les problèmes d'explicabilité grâce à des interfaces utilisateur mieux conçues

Certaines entreprises ont commencé à inclure l'explicabilité dans leurs solutions pour que les utilisateurs comprennent mieux les processus d'IA exécutés en arrière-plan. MondoBrain est l'une d'entre elles. Basée en France, elle combine intelligence humaine, collective et artificielle pour fournir une solution de réalité augmentée aux entreprises. Grâce à des tableaux de bord interactifs de visualisation des données, elle évalue l'ensemble des données existantes au sein d'une entreprise (à partir des logiciels de planification des ressources de l'entreprise, de gestion des programmes de l'entreprise ou de gestion des relations avec la clientèle, par exemple) et formule des recommandations normatives sur la base des requêtes des clients (Graphique 4.1). Elle utilise un algorithme d'apprentissage automatique pour éliminer les variables commerciales qui ne présentent pas d'intérêt par rapport à la requête et pour extraire les variables ayant le plus d'impact.

Les couleurs des feux de signalisation guident les utilisateurs à chaque étape de la requête, facilitant leur compréhension du processus de décision. Chaque décision est automatiquement enregistrée et devient vérifiable et traçable. Cela donne un compte rendu complet mais simple de toutes les étapes qui ont conduit à la recommandation commerciale finale.

Graphique 4.1. Illustration des outils de visualisation des données visant à améliorer l'explicabilité



Source : www.mondobrain.com.

Dans certains cas, un compromis doit être trouvé entre explicabilité et précision. Pour être explicables, les variables d'une solution doivent potentiellement être réduites à un ensemble

suffisamment petit pour pouvoir être appréhendé par l'homme. Cela peut s'avérer sous-optimal dans le cadre de problèmes complexes et de grande ampleur. Par exemple, certains modèles d'apprentissage automatique utilisés pour établir un diagnostic médical peuvent prédire avec précision la probabilité d'une maladie, mais sont trop complexes pour être accessibles à l'esprit humain. Dans ce genre de cas, il convient de comparer les préjudices que peut causer un système moins précis qui offre des explications claires avec ceux d'un système plus précis dans lequel les erreurs sont plus difficiles à détecter. Par exemple, la prévision de la récidive peut nécessiter des modèles simples et explicables dans lesquels les erreurs sont décelables (Dressel et Farid, 2018[34]). Dans des domaines tels que les prévisions climatiques, on acceptera plus facilement des modèles plus complexes qui offrent de meilleures prévisions mais sont moins explicables. À plus forte raison s'il existe d'autres mécanismes de responsabilité vis-à-vis des résultats, tels que des données statistiques permettant de détecter un éventuel biais ou une éventuelle erreur.

Robustesse, sûreté et sécurité

Ce qu'il faut entendre par robustesse, sûreté et sécurité

La robustesse peut s'entendre comme la capacité à supporter ou surmonter des conditions défavorables (OCDE, 2019[35]), notamment des risques de sécurité numérique. Les systèmes d'IA pourront être dits « sécurisés » dans la mesure où leur utilisation dans des conditions normales ou prévisibles, y compris si elle est abusive, ne fera jamais peser un risque de sécurité démesuré, quel que soit le stade de leur cycle de vie (OCDE, 2019[36]). Les questions de robustesse et de sécurité de l'IA sont interdépendantes. À titre d'exemple, la sécurité numérique peut avoir une incidence sur la sécurité des produits si les dispositifs connectés, comme les voitures autonomes ou les appareils électroménagers fonctionnant grâce à l'IA ne sont pas suffisamment sécurisés ; des pirates pourraient en prendre le contrôle et en modifier les paramètres à distance.

La gestion des risques dans les systèmes d'IA

Le niveau de protection requis dépend d'une analyse risques-avantages

Il conviendrait de mettre les préjudices susceptibles d'être causés par un système d'IA en regard des coûts qu'il faudrait supporter pour faire la transparence sur ces systèmes et définir les responsabilités y afférentes. Les préjudices potentiels pourraient être des risques pesant sur les droits individuels, la vie privée, l'équité et la robustesse. Toutes les utilisations de l'IA ne s'accompagnent pas des mêmes risques cependant, et exiger une explication, par exemple, génère aussi un certain coût. En matière de gestion des risques, un large consensus semble se dégager autour de l'idée selon laquelle plus les enjeux sont importants plus il faut faire preuve de transparence et de responsabilité, tout particulièrement lorsqu'il y va de la vie ou de la liberté des personnes.

Les stratégies de gestion des risques ont leur place tout au long du cycle de vie des systèmes d'IA

Les organisations ont recours à la gestion des risques pour isoler, évaluer, hiérarchiser et traiter les risques susceptibles d'altérer le comportement d'un système et les résultats attendus de son utilisation. Cette démarche peut également servir à déterminer quels risques pèsent sur les différentes parties prenantes et comment les maîtriser tout au long du cycle de vie du système d'IA considéré (voir au chapitre 1 la section consacrée au cycle de vie des systèmes d'IA).

Les acteurs de l'IA – ceux qui jouent un rôle actif au cours du cycle de vie d'un système d'IA – évaluent et atténuent les risques à l'échelle de ce système pris dans son ensemble, ainsi qu'au cours de chaque phase de son cycle de vie. La gestion des risques dans les systèmes d'IA suit les étapes suivantes, dont l'importance varie selon le stade atteint dans le cycle de vie :

1. **Objectifs** : Définir les objectifs, les fonctions ou les propriétés du système, en contexte. Fonctions et propriétés peuvent changer suivant la phase du cycle de vie.
2. **Parties prenantes et acteurs** : Identifier les parties prenantes et les acteurs concernés, autrement dit ceux qui sont directement ou indirectement intéressés par les fonctions ou les propriétés du système à chaque étape du cycle de vie.
3. **Évaluation des risques** : Évaluer les effets potentiels – avantages et risques – du système pour les parties prenantes et les acteurs. Ces effets varieront en fonction des parties prenantes et des acteurs concernés comme selon la phase de son cycle de vie atteinte par le système d'IA considéré.
4. **Atténuation des risques** : Identifier des stratégies d'atténuation adaptées et proportionnées aux risques. Celles-ci devraient tenir compte de différents paramètres tels que les buts et objectifs de l'entité, les parties prenantes et acteurs concernés, la probabilité d'occurrence du risque et les avantages potentiels.
5. **Mise en œuvre** : Appliquer les stratégies d'atténuation des risques.
6. **Suivi, évaluation et reddition de comptes** : Suivre la mise en œuvre de la stratégie, évaluer ses résultats et en rendre compte.

L'utilisation de la gestion des risques et la consignation des décisions prises à chaque étape du cycle de vie peut contribuer à la transparence d'un système d'IA et à la responsabilisation de l'entité à l'égard de ce système.

Il convient d'apprécier côté à côté l'ampleur du préjudice global et le risque immédiat

Considérées de manière isolée, quelques-unes des utilisations possibles des systèmes d'IA présentent un faible niveau de risque. Elles pourraient cependant nécessiter davantage de robustesse de la part de ces systèmes en raison de leurs effets sur la société. Un système qui, par son fonctionnement, causerait un préjudice mineur à un grand nombre de personnes n'en causerait pas moins un préjudice global significatif pour la collectivité. Supposons, par exemple, qu'un petit nombre d'outils fondés sur l'IA soient intégrés dans une multitude de services et de secteurs et servent pour les demandes de prêt, la souscription de contrats d'assurance ou les enquêtes de moralité. Une seule erreur, un seul biais, introduits dans un système seraient susceptibles d'entraîner une cascade de réponses négatives (Citron et Pasquale, 2014[37]). Ces réponses négatives, prises une à une, ne prêteront probablement pas à conséquence. Leur accumulation, en revanche, pourrait avoir un effet délétère. Il semble par conséquent souhaitable que les décideurs prennent en compte, dans leurs débats, l'ampleur du préjudice global, en plus de considérer le risque immédiat.

La robustesse face aux risques de sécurité numériques associés à l'IA

L'IA permet des attaques plus sophistiquées et d'une envergure potentiellement accrue

L'utilisation de l'IA à des fins malveillantes est appelée à se développer à mesure que celle-ci devient moins onéreuse et plus accessible, et parallèlement à son emploi au service de la sécurité numérique (voir, au chapitre 3, la section sur l'IA dans la sécurité numérique). Les auteurs de cyberattaques s'emploient à renforcer leurs capacités en matière d'IA. La rapidité croissante et la sophistication des attaques ne laissent pas d'inquiéter[6]. Dans ce contexte, on voit se renforcer les menaces existantes tandis qu'il en surgit de nouvelles et que le caractère même des menaces évolue.

Les systèmes d'IA contemporains présentent un certain nombre de vulnérabilités. Des individus malintentionnés peuvent manipuler les données servant à entraîner l'un de ces systèmes (par exemple dans le cas d'un « empoisonnement des données »). Ils peuvent aussi bien découvrir les caractéristiques servant, dans un modèle de sécurité numérique, à détecter les logiciels malveillants et, cette information une fois connue, créer un code malveillant ou causer de manière intentionnelle une classification erronée d'éléments d'information (par exemple en donnant des « exemples contradictoires », Encadré 4.6) (Brundage et al., 2018[38]). Les technologies d'IA devenant de plus en plus accessible, davantage de personnes peuvent les utiliser pour mener des attaques sophistiquées d'une envergure supérieure aux attaques de naguère. La fréquence et l'efficacité des attaques de sécurité numérique nécessitant une préparation méticuleuse, comme c'est le cas avec le harponnage (*spear phishing*), pourraient bien augmenter avec leur automatisation, rendue possible par les algorithmes d'apprentissage automatique.

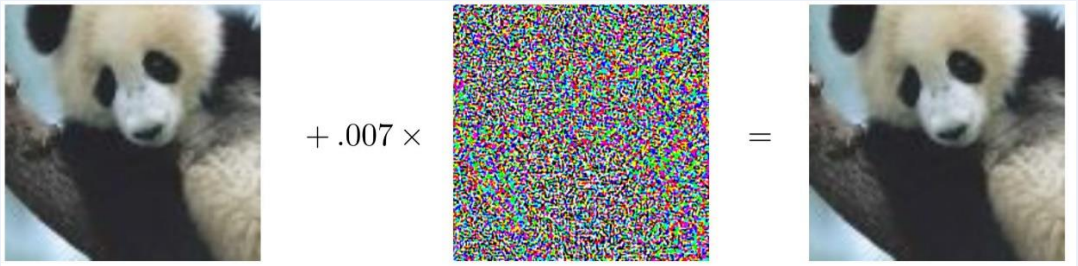
Encadré 4.6. Du danger des exemples contradictoires pour l'apprentissage automatique

On appelle **exemples contradictoires** les éléments introduits à dessein dans des modèles d'apprentissage automatique par des personnes malintentionnées afin d'amener ces modèles à commettre des erreurs présentant toutes les apparences de la fiabilité. Il s'agit d'un réel problème pour la robustesse et la sûreté des systèmes d'IA car plusieurs modèles d'apprentissage automatique, y compris les réseaux neuronaux à la pointe de la technologie, leur sont vulnérables.

Ces exemples contradictoires peuvent être subtils. Ainsi, dans le Graphique 4.2, une modification imperceptible, ou « élément contradictoire », a été ajoutée à l'image d'un panda. Cette modification est destinée à tromper le modèle de classification d'images. Il s'ensuit que l'algorithme confond un panda avec un gibbon avec un niveau de confiance proche de 100 %.

Des recherches récentes ont démontré qui plus est que l'on pouvait créer des exemples contradictoires à partir d'une image imprimée sur papier ordinaire et photographiée à l'aide d'un smartphone avec un niveau de résolution normal. Ces images peuvent être dangereuses : un simple autocollant apposé sur un « stop » pourrait induire une voiture autonome à interpréter ce panneau comme une priorité ou à le confondre avec n'importe quel autre panneau de signalisation.

Graphique 4.2. Trompé par une légère modification, un algorithme confond un panda avec un gibbon



Sources : Goodfellow, Shlens et Szegedy (2015[39]), « Explaining and harnessing adversarial examples », https://arxiv.org/pdf/1412.6572.pdf ; Kurakin, Goodfellow et Bengio (2017[40]), « Adversarial examples in the physical world », https://arxiv.org/abs/1607.02533.

La sûreté

Les systèmes d'apprentissage automatique et les systèmes autonomes bousculent les cadres d'action en place en matière de sûreté

La gamme des équipements dotés d'une intelligence artificielle s'étoffe rapidement – depuis les robots jusqu'aux voitures autonomes en passant par les produits et services grand public comme les appareils et les systèmes d'alarme domestique intelligents. Ces équipements présentent des avantages non négligeables sur le plan de la sûreté mais soulèvent dans le même temps des problèmes d'ordre juridique et pratique en rapport avec les cadres relatifs à la sécurité des produits (OCDE, 2018[21]). Ces cadres tendent à régir des produits « finis » et tangibles davantage que des logiciels, or de nombreux système d'IA continuent d'apprendre et évoluent tout au long de leur cycle de vie[7]. Les produits fondés sur l'IA peuvent aussi être « autonomes » ou « semi-autonomes », autrement dit prendre et appliquer des décisions dans lesquels l'être humain n'intervient pas ou alors seulement de manière limitée.

Aux différents types d'applications de l'IA correspondront vraisemblablement des actions adaptées de la part des pouvoirs publics (Freeman, 2017[41]). D'une manière générale, les systèmes d'IA appellent les décideurs à mener une quadruple réflexion. Il faut premièrement rechercher le meilleur moyen de garantir la sécurité des produits. En d'autres termes, les produits ne doivent pas présenter des risques démesurés pour la sécurité dans des conditions d'utilisation normales ou prévisibles, y compris en cas d'utilisation abusive, et ce tout au long de leur cycle de vie. Cela concerne en particulier les cas où les données disponibles pour entraîner le système d'IA sont peu nombreuses (Encadré 4.7). Deuxièmement, il convient de se demander qui doit être tenu pour responsable, et jusqu'à quel point, des dommages causés par un système d'IA. Il y a lieu en parallèle de se demander quelles sont les parties prenantes qui peuvent concourir à la sûreté des appareils autonomes. Les utilisateurs, les fabricants de produits et de capteurs, les producteurs de logiciels, les concepteurs, les fournisseurs d'infrastructure et les entreprises d'analyse de données pourraient être du nombre. Il importe, troisièmement, de réfléchir au choix du ou des types de responsabilité à appliquer – devrait-il s'agir d'une responsabilité objective ou d'une responsabilité en cas de faute et quel devrait être le rôle dévolu à l'assurance. L'opacité de certaines systèmes d'IA n'aide pas à trancher cette question. Quatrièmement, il y a lieu pour les décideurs de s'interroger sur la manière de donner effet à la législation, sur ce qu'il faut considérer comme un « défaut » dans un produit fondé sur l'IA, sur la définition de la charge de la preuve et sur les voies de recours existantes.

La Directive européenne de 1985 relative à la responsabilité du fait des produits défectueux (Directive 85/374/CEE) pose le principe de « responsabilité sans faute » ou de « responsabilité stricte », en vertu duquel le producteur est responsable des dommages causés au consommateur par un produit défectueux, même en l'absence de négligence ou de faute de sa part. La Commission européenne a entrepris de réviser cette directive. D'après les premières conclusions, le modèle demeure globalement adapté (Ingels, 2017[42]). Il reste que les technologies reposant actuellement sur l'IA et celles que l'on peut attendre dans l'avenir remettent en cause les notions de « produit », de « sûreté », de « défaut » et de « dommage », ce qui tend à rendre plus délicate la question de la charge de la preuve.

Dans le domaine des véhicules autonomes, la sûreté est la préoccupation première des autorités. Des travaux de fond doivent être réalisés au sujet des essais auxquels soumettre ces véhicules en vue de garantir que leur fonctionnement est bien sécurisé. Ces travaux doivent porter notamment sur les régimes de licence qui évaluent la possibilité d'une expérimentation préalable des systèmes installés à bord ou traiter de la nécessité pour ces

systèmes de contrôler la vigilance des conducteurs humains qui peuvent être appelés à reprendre la main en cas de besoin. L'octroi de licences constitue dans certains cas un réel problème pour les entreprises qui souhaitent procéder à des essais de véhicules. Les pouvoirs publics sont d'autre part plus ou moins favorables à la réalisation de tels essais. D'aucuns ont appelé à l'application d'un régime de responsabilité stricte à l'égard des constructeurs de véhicules autonomes, la responsabilité dépendant alors du caractère contrôlable ou non du risque. Il serait reconnu par exemple qu'un simple passager d'une voiture sans chauffeur ne saurait être tenu pour fautif ou manquer à un devoir de vigilance. Les juristes sont d'avis que même le concept de « détenteur déclaré » ne serait pas applicable car le détenteur doit être en mesure de maîtriser le risque (Borges, 2017[43]). L'idée a été émise que les compagnies d'assurance pourraient prendre en charge le risque de dommage du fait des voitures autonomes à partir d'une classification des véhicules déclarés établie sur la base d'évaluations des risques.

Encadré 4.7. Les données synthétiques au service d'une IA plus sûre et plus précise : le cas des véhicules autonomes

L'utilisation des données synthétiques devient de plus en plus courante dans le domaine de l'apprentissage automatique car elle permet de simuler des scénarios difficilement observables ou reproductibles en conditions réelles. Selon les explications de Philipp Slusallek, directeur scientifique du Centre allemand de recherche sur l'IA, il s'agit par exemple de s'assurer par ce moyen qu'une voiture autonome ne percutera pas un enfant qui traverserait la rue en courant.

La « réalité numérique » – un environnement simulé répliquant les caractéristiques pertinentes du monde réel – pourrait avoir quatre effets. Premièrement, elle pourrait fournir les données synthétiques à partir desquelles apprendre aux systèmes d'IA à faire face à des situations complexes. Deuxièmement, elle permettrait la validation des caractéristiques de fonctionnement et le recalibrage des données synthétiques par rapport aux données obtenues en conditions réelles. Troisièmement, elle pourrait servir à l'organisation d'examens, comme celui du permis de conduire pour les conducteurs de véhicule autonome. Elle permettrait en quatrième lieu d'explorer le processus décisionnel suivi par le système et de découvrir les conséquences potentielles des autres choix possibles. Cette méthode a ainsi permis à Google d'entraîner ses voitures autonomes en leur faisant parcourir, en simulation, plus de 4.8 millions de kilomètres par jour (soit plus de 500 allers-retours entre New-York et Los Angeles).

Sources : Golson (2016[44]), « Google's self-driving cars rack up 3 million simulated miles every day », https://www.theverge.com/2016/2/1/10892020/google-self-driving-simulator-3-million-miles ; Slusallek (2018[45]), *Artificial Intelligence and Digital Reality: Do We Need a CERN for AI?*, https://www.oecd-forum.org/channels/722-digitalisation/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai.

Les normes de sécurité au travail demanderont sans doute à être mises à jour

Il est probable qu'entre autres conséquences directes sur les conditions de travail, l'IA va rendre nécessaire l'introduction de nouveaux protocoles de sécurité. L'adoption ou la révision de normes sectorielles et d'accords d'entreprise à portée technologique vont devenir inévitables pour garantir que les conditions de fiabilité et de sûreté nécessaires à la productivité sont bien réunies en milieu de travail. Le Comité économique et social européen (CESE) préconise que les « parties prenantes œuvrent ensemble en faveur de systèmes d'IA complémentaires et de leur mise en place conjointe sur le lieu de travail » (CESE, 2017[46]).

Responsabilité

L'utilisation croissante de l'IA doit s'accompagner d'un effort en matière de responsabilité, garant du bon fonctionnement des systèmes

La notion de **responsabilité** désigne essentiellement le fait de savoir attribuer à chacun, organisation ou individu, la part qui lui revient dans le bon fonctionnement des systèmes d'IA. Les critères sur lesquels elle repose sont le respect des valeurs humaines et de l'équité, de la transparence, de la robustesse et de la sûreté. La responsabilité dépend du rôle de chacun des acteurs de l'IA, du contexte et de l'état de la technologie. Du point de vue des responsables de l'action publique, elle dépend de mécanismes remplissant plusieurs offices. Ces mécanismes identifient la partie qui est comptable de telle recommandation ou de telle décision. Ils corrigent ladite recommandation ou décision avant sa mise à exécution. Il est possible également par ce moyen de contester la décision, ou d'en faire appel, à un stade ultérieur, ou même de récuser le système dont elle est issue (Helgason, 1997[47]).

Dans la pratique, la responsabilité dans les systèmes d'IA dépend souvent du fonctionnement d'un système donné au regard d'indicateurs de précision ou d'efficience, auxquels viennent aujourd'hui s'ajouter de plus en plus fréquemment des indicateurs d'équité, de sûreté et de robustesse. Les seconds cependant restent moins utilisés que les premiers. Comme pour tous les indicateurs, le suivi et l'évaluation peuvent s'avérer coûteux. Aussi le type et la fréquence des relevés doivent-ils être proportionnés aux risques et avantages potentiels.

Le niveau de responsabilité requis dépend du niveau de risque

Les stratégies à suivre sont fonction du contexte et des circonstances. À titre d'exemple, une responsabilité relativement élevée sera sans doute attendue, en matière d'utilisation de l'IA, de la part du secteur public, en particulier dans l'exercice de fonctions régaliennes, telles la sécurité ou l'exécution des lois, d'où peuvent découler des préjudices importants. Des mécanismes de responsabilité formels sont par ailleurs souvent requis à l'égard des applications développées par le secteur privé dans les domaines du transport, de la finance et des soins de santé, qui sont strictement encadrés. Dans d'autres domaines soumis à un contrôle moins sévère, l'utilisation de l'IA s'accompagne plus rarement de semblables mécanismes. Dans ces cas, les stratégies techniques prennent d'autant plus d'importance en matière de transparence et de responsabilité. Elles doivent garantir que les systèmes conçus et exploités par des acteurs du secteur privé respectent un certain nombre de normes sociétales et de contraintes légales.

Certaines applications ou décisions pourraient requérir l'intervention d'un « élément humain » chargé d'apprécier le contexte social dans lequel elles s'inscrivent et leurs potentiels effets indésirables. Lorsqu'une décision emporte des conséquences significatives sur le quotidien des individus, il est communément admis qu'elle ne devrait pas être prise uniquement sur la base d'un résultat fourni par l'IA (par exemple un score). Le Règlement général sur la protection des données préconise ainsi qu'il y ait en pareil cas une intervention humaine. À titre d'exemple, les individus doivent être informés lorsque l'IA est utilisée pour rendre un jugement, accorder ou refuser un prêt, décider de l'orientation d'élèves ou d'étudiants ou sélectionner des candidats à un poste. Lorsque les enjeux sont importants, des mécanismes de responsabilité formels sont souvent exigés. À titre d'exemple, un magistrat qui s'appuie sur l'IA pour prononcer une condamnation constituera l'« élément humain » intervenant dans le processus. Cependant, l'existence d'autres mécanismes de responsabilité – dont la possibilité de faire appel du jugement comme dans une procédure traditionnelle – contribue à garantir que les recommandations formulées par l'IA soient bien prises comme un

élément d'appréciation parmi d'autres (Wachter, Mittelstadt et Floridi, 2017[48]). En l'absence de risque particulier, par exemple lorsqu'il s'agit de recommander un restaurant, la machine seule suffira. Il n'y aura sans doute pas lieu en l'occurrence de multiplier les strates au risque de générer des coûts superflus.

Cadre d'action applicable à l'IA

Des politiques nationales doivent être mises en œuvre pour promouvoir des systèmes d'IA dignes de confiance. Ces mesures peuvent entraîner des effets bénéfiques et équitables pour les individus et pour la planète, en particulier dans des domaines prometteurs actuellement sous-investis par le marché. La création d'un environnement réglementaire propice à l'avènement d'une IA à laquelle on puisse se fier sans crainte suppose, entre autres choses, de favoriser l'investissement public et privé dans les activités de recherche-développement connexes et de faire acquérir aux individus les compétences qui leur permettront de s'adapter avec succès à l'évolution des emplois. Les paragraphes qui suivent sont consacrés à quatre domaines d'action essentiels à la promotion et au développement d'une IA qui mérite toute notre confiance.

Investissement dans la recherche et le développement en matière d'IA

L'investissement à long terme dans la recherche publique peut aider à façonner l'innovation en matière d'IA

L'OCDE s'intéresse au rôle des politiques d'innovation dans la transformation numérique et l'adoption de l'IA (OCDE, 2018[49]). À ce titre, elle étudie notamment le rôle des politiques en faveur de la recherche publique, du transfert de connaissances et de la création conjointe, à l'appui du développement des outils et des infrastructures de recherche pour l'IA. L'intelligence artificielle oblige les décideurs à réévaluer le niveau d'intervention idoine des pouvoirs publics dans la recherche connexe pour relever les défis sociétaux (OCDE, 2018[14]). En outre, les établissements de recherche dans tous les domaines devront se doter de systèmes d'IA robustes pour rester compétitifs, en particulier dans des domaines comme la science biomédicale et la biologie. Des instruments émergents, à l'instar des plateformes de partage des données et des installations de superinformatique, peuvent aider à stimuler la recherche dans l'IA et pourraient appeler de nouveaux investissements. Le Japon, par exemple, consacre plus de 120 millions USD par an à la construction d'une infrastructure de calcul hautes performances pour les universités et les centres publics de recherche.

L'IA est considérée comme une technologie générique susceptible d'avoir des incidences sur de nombreux secteurs (Agrawal, Gans et Goldfarb, 2018[50] ; Brynjolfsson, Rock et Syverson, 2018[51]). Elle est également vue comme l'« invention d'une méthode d'invention » (Cockburn, Henderson et Stern, 2018[52]), déjà largement utilisée par les chercheurs et les inventeurs pour faciliter l'innovation. Sans compter que des secteurs entièrement nouveaux pourraient voir le jour grâce à des percées scientifiques faisant fond sur l'IA. D'où l'importance de la recherche fondamentale et la nécessité d'inscrire les politiques de recherche dans une vision à long terme (OCDE, 2018[53]).

Favoriser l'instauration d'un écosystème numérique propice à l'IA

Technologies et infrastructure d'IA

Des progrès significatifs ont été réalisés ces dernières années dans les technologies liées à l'IA. On les doit à la maturité des techniques de modélisation statistique, comme les réseaux neuronaux, en particulier les réseaux neuronaux profonds (on parle d'« apprentissage

profond »). Nombre des outils employés pour gérer et utiliser l'IA sont des ressources à code source libre, qui relèvent du domaine public. Cela facilite leur adoption et permet de corriger les bogues logiciels à l'aide de solutions issues de contributions participatives. TensorFlow (de Google), Michelangelo (d'Uber) et Cognitive Toolkit (de Microsoft) en sont des exemples. Par ailleurs, des entreprises et des chercheurs partagent publiquement des ensembles de données d'entraînement après curation et des outils d'apprentissage afin de favoriser la diffusion des technologies liées à l'IA.

Une partie des avancées récentes de l'IA s'explique par l'accélération exponentielle des temps de traitement et la Loi de Moore (selon laquelle le nombre de transistors sur un circuit intégré à haute densité double tous les deux ans environ). Grâce à la conjugaison de ces deux phénomènes, les algorithmes d'IA peuvent traiter rapidement des volumes considérables de données. À mesure que les projets d'IA passent du concept à l'application commerciale, les besoins de ressources spécialisées et onéreuses d'infonuagique et de processeurs graphiques vont croissant. L'essor des systèmes d'IA s'accompagne également d'une augmentation fulgurante de la puissance de calcul requise. Selon une estimation, l'expérience la plus importante menée récemment, AlphaGo Zero, a nécessité une puissance de calcul 300 000 fois supérieure à celle utilisée pour mener à bien l'expérience la plus importante six ans plus tôt (OpenAI, 16 mai 2018[54]). De même, les prouesses du programme d'échecs et de Go, AlphaGo Zero, ont mobilisé une puissance de calcul qui dépasserait celle des dix superordinateurs les plus puissants du monde conjugués (OCDE, 2018[53]).

Accessibilité et utilisation des données

L'accessibilité et le partage des données peuvent accélérer ou, selon le cas, freiner les progrès de l'IA

Les technologies actuelles d'apprentissage automatique requièrent, pour s'entraîner et évoluer, des données fiables ayant fait l'objet d'une curation. L'accès à des ensembles de données de qualité s'avère par conséquent essentiel au développement de l'IA. Les facteurs liés à l'accessibilité et au partage des données susceptibles d'accélérer ou, selon le cas, de freiner les progrès de l'IA, sont les suivants (OCDE, 2019[55]) :

- **Normes** *:* Les normes sont nécessaires pour permettre l'interopérabilité et la réutilisation des données d'une application à l'autre, favoriser l'accessibilité et garantir que les données puissent être trouvées, intégrées à des catalogues et/ou interrogées et réutilisées.
- **Risques** : Les risques liés au partage des données, qui pèsent sur les individus, les organisations et les pays, peuvent aller de la violation de la confidentialité et de la vie privée, au non-respect des droits de propriété intellectuelle (DPI), en passant par la menace des intérêts commerciaux ou la compromission de la sécurité nationale et de la sécurité numérique.
- **Coûts des données** : La collecte, l'accès, le partage et la réutilisation nécessitent des investissements en amont et en aval. Outre ceux liés à l'acquisition des données, des investissements supplémentaires doivent être consacrés à leur nettoyage, à leur curation, à la maintenance des métadonnées, au stockage et au traitement des données, et à la sécurisation de l'infrastructure informatique.
- **Incitations** : Les approches fondées sur le marché peuvent encourager à ouvrir l'accès aux données et à les partager avec des marchés et des plateformes qui commercialisent des données et proposent des services à valeur ajoutée, comme les infrastructures de paiement et d'échange de données.

- **Incertitudes quant à la propriété des données** : Les cadres juridiques – régimes de propriété intellectuelle et droit (cyber)criminel, de la concurrence et de la protection de la vie privée –, conjugués à la multiplicité des parties intervenant dans la création des données, créent des incertitudes quant à la « propriété des données ».
- **Autonomisation des utilisateurs, y compris des agents intelligents** : Donner aux utilisateurs les moyens d'agir et faciliter la portabilité des données – tout en mettant en place des mécanismes efficaces de consentement et de choix pour les personnes concernées par les données – peuvent inciter les individus et les entreprises à partager des données personnelles ou professionnelles. D'aucuns soulignent en outre que les agents intelligents qui connaissent les préférences des individus pourraient les aider à négocier des dispositifs complexes de partage des données avec d'autres systèmes d'IA (Neppel, 2017[56]).
- **Tiers de confiance** *:* Les tierces parties peuvent aider à instaurer la confiance et faciliter le partage et la réutilisation des données entre l'ensemble des parties prenantes. Les intermédiaires de données peuvent agir en tant qu'autorités de certification. Les plateformes de confiance spécialisées dans le partage des données, à l'image des fiducies de données, fournissent des données de qualité. Sans oublier les comités d'évaluation éthique, qui veillent au respect des intérêts légitimes des tierces parties.
- **Représentativité des données** : Les systèmes d'IA établissent des prévisions d'après des schémas identifiés dans les ensembles de données d'entraînement. C'est pourquoi, dans une optique à la fois d'exactitude et d'équité, les ensembles de données d'entraînement doivent être inclusifs, divers et représentatifs afin de ne pas sous-représenter ni sur-représenter des groupes particuliers.

Les politiques publiques peuvent favoriser l'accessibilité et le partage des données à l'appui du développement de l'IA

Plusieurs stratégies sont envisageables pour renforcer l'accessibilité et le partage des données (OCDE, 2019[55]) :

- **Favoriser l'accès aux données du secteur public**, qu'il s'agisse de données publiques ouvertes, de données géographiques (des cartes, par exemple) ou de données liées aux transports.
- **Faciliter le partage des données dans le secteur privé**, soit selon un principe de volontariat, soit en application de dispositions obligatoires, auquel cas le partage des données se fait exclusivement avec des utilisateurs de confiance. Certains domaines appellent une attention particulière, à l'instar des « données d'intérêt général », des données relevant d'industries de réseau comme les transports et l'énergie, pour l'interopérabilité des services, et de la portabilité des données à caractère personnel.
- **Développer les capacités statistiques/analytiques**, en mettant en place des centres de technologie qui fournissent un soutien et des conseils en matière d'utilisation et d'analyse des données.
- **Définir des stratégies nationales en matière de données**, afin d'assurer la cohérence des cadres nationaux de gouvernance des données et leur compatibilité avec les stratégies nationales en matière d'IA.

Des approches techniques voient le jour pour remédier aux contraintes liées aux données

Certains algorithmes d'apprentissage automatique, tels que ceux appliqués à la reconnaissance d'images, affichent des performances supérieures aux capacités humaines moyennes. Toutefois, pour y parvenir, ils devaient jusqu'à présent être entraînés à l'aide de bases de données colossales contenant des millions d'images étiquetées. Les besoins en données ont encouragé la recherche active dans des techniques d'apprentissage automatique qui requièrent moins de données pour entraîner les systèmes d'IA. Plusieurs méthodes peuvent aider à parer au manque de données.

- **L'apprentissage profond par renforcement** est une technique d'apprentissage automatique qui allie des réseaux neuronaux profonds et l'apprentissage par renforcement (voir chapitre 1, sous-section « Volet 2 : Techniques d'apprentissage automatique »). Ce faisant, il apprend à privilégier un comportement donné menant au résultat recherché (Mousave, Schukat et Howley, 2018[57]). Des « agents » intelligents rivalisent en exécutant des actions dans un environnement complexe et reçoivent soit une « récompense », soit une « pénalité », selon que l'action a mené au résultat souhaité ou non. Les agents ajustent leurs actions à la lumière de ces « retours d'information »[8].
- **L'apprentissage par transfert ou le pré-entraînement** (Pan et Yang, 2010[58]) réutilise des modèles qui ont été entraînés, en vue d'exécuter des tâches différentes dans le même domaine. Par exemple, certaines couches d'un modèle entraîné à reconnaître des images de chats pourraient être réutilisées pour détecter des images de robes bleues. La taille de l'échantillon d'images s'avérerait alors bien inférieur à ce qu'exigent les algorithmes d'apprentissage automatique traditionnels (Jain, 2017[59]).
- **L'apprentissage fondé sur des données augmentées**, ou synthétisation de données, peut créer artificiellement des données à l'aide de simulations ou d'interpolations à partir de données existantes. Cette technique permet d'accroître le volume de données et, partant, d'améliorer l'apprentissage. Elle s'avère particulièrement intéressante lorsque les contraintes liées au respect de la vie privée limitent l'utilisation des données ou pour simuler des scénarios qui se produisent extrêmement rarement dans la réalité (Encadré 4.7)[9].
- **Les modèles d'apprentissage hybride** peuvent modéliser l'incertitude en alliant différents types de réseaux neuronaux profonds et des approches probabilistes ou bayésiennes. Cette modélisation de l'incertitude a pour objectif d'améliorer les performances et l'explicabilité, et de réduire la probabilité d'obtenir des erreurs de prévisions (Kendall, 23 mai 2017[60]).

Les préoccupations quant à la protection de la vie privée, la confidentialité et la sécurité pourraient avoir pour effet de limiter l'accessibilité et le partage des données. De là peut naître un décalage entre la rapidité d'apprentissage des systèmes d'IA et la disponibilité des ensembles de données utilisés pour les entraîner. Les progrès récents des techniques de cryptographie, comme le calcul multipartite sécurisé (CMS) et le chiffrement homomorphe, pourraient permettre de réaliser des analyses de données tout en garantissant le respect des droits connexes. De fait, les systèmes d'IA pourraient alors opérer sans collecter des données sensibles ni devoir y accéder (Encadré 4.8). Les modèles d'IA sont par ailleurs de plus en plus à même de travailler avec des données chiffrées[10]. Toutefois, dans la mesure où ces solutions nécessitent une puissance de calcul considérable, il peut s'avérer difficile de les déployer à grande échelle (Brundage et al., 2018[38]).

Encadré 4.8. Les nouveaux outils cryptographiques permettent d'exécuter des calculs tout en préservant la vie privée

Les progrès de la cryptographie ouvrent la voie à des applications prometteuses dans le domaine de l'IA. Par exemple, un modèle d'apprentissage automatique pourrait être entraîné à l'aide d'une combinaison de données issues de diverses organisations. Ce faisant, les données de l'ensemble des participants resteraient confidentielles. Une telle solution lèverait les obstacles liés aux problématiques de respect de la vie privée et de confidentialité. Les techniques de chiffrement qui permettent ce type de traitement ne sont pas nouvelles : si le chiffrement homomorphe existe depuis des années, le calcul multipartite sécurisé remonte quant à lui à plusieurs décennies. Pour autant, elles n'étaient pas jusqu'à présent suffisamment efficientes pour une utilisation pratique. Grâce aux progrès récents des algorithmes et de la mise en œuvre, elles deviennent peu à peu des outils fonctionnels capables d'exécuter des analyses sur des ensembles de données réels.

- **Chiffrement homomorphe** : Technique permettant d'exécuter des calculs sur des données chiffrées, sans avoir besoin de disposer des données non chiffrées.
- **Calcul multipartite sécurisé (CMS)** : Technique permettant de calculer une fonction de données collectées à partir de diverses sources sans que les informations relatives aux données de l'une des sources ne soient révélées à aucune des autres sources. Les protocoles de CMS permettent à diverses parties de calculer conjointement des algorithmes dont les entrées restent des données privées.

Sources : Brundage et al. (2018[38]), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf ; Dowlin (2016[61]), *CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy*, https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/CryptonetsTechReport.pdf.

Autre solution : les modèles d'IA pourraient mettre à profit la technologie des chaînes de blocs, qui utilise elle aussi des outils cryptographiques pour sécuriser le stockage des données (Encadré 4.9). Les solutions alliant l'IA et la technologie des chaînes de blocs pourraient contribuer à accroître la disponibilité des données, tout en minimisant les risques d'atteinte à la vie privée et de sécurité liés au traitement de données non chiffrées.

Encadré 4.9. La technologie des chaînes de blocs permet une vérification d'identité respectueuse de la vie privée dans le cadre de l'IA

Kairos, une entreprise qui édite une solution de reconnaissance faciale, a intégré la technologie des chaînes de blocs dans son portefeuille. Elle l'allie à la biométrie faciale pour offrir aux utilisateurs une meilleure protection de leur vie privée. Un algorithme compare l'image d'une personne avec des repères faciaux jusqu'à obtenir une correspondance exacte. Celle-ci est convertie en une chaîne unique et aléatoire de nombres, ce qui permet de ne pas avoir à conserver l'image d'origine. Cette chaîne de blocs biométrique se fonde sur le principe que les entreprises ou les administrations n'ont pas besoin de savoir qui est l'utilisateur pour en vérifier l'identité.

Source : https://kairos.com/.

Concurrence

L'OCDE a étudié l'impact de la transformation numérique sur la concurrence, ainsi que les implications en termes d'action des pouvoirs publics (OCDE, 2019[62]). Cette sous-section expose certaines incidences potentielles propres à l'IA. Elle met en évidence les effets pro-concurrentiels largement reconnus de l'IA, qui facilite l'entrée de nouveaux acteurs. Elle souligne en outre que les politiques de la concurrence tendent à accorder davantage d'attention aux grands acteurs de l'IA, du fait de leur rôle d'opérateurs de plateformes en ligne et de détenteurs de volumes considérables de données. Elles s'intéressent peu en revanche à l'utilisation de l'IA à proprement parler.

Une question liée plus précisément à l'IA se pose : existe-t-il un effet de réseau fondé sur les données ? Si tel est le cas, l'utilité pour chaque utilisateur de recourir à certains types de plateformes augmente dès lors que d'autres l'utilisent également. Par exemple, en recourant à l'une de ces plateformes, ils contribuent à apprendre aux algorithmes qui la sous-tendent à mieux servir les utilisateurs (Heiner et Nguyen, 2018[9]). Par ailleurs, les données se caractérisent par des rendements d'échelle décroissants : l'amélioration des prévisions devient de plus en plus marginale à mesure que les données dépassent un certain seuil. Certains s'interrogent par conséquent sur le risque que l'IA pose des problèmes de concurrence à long terme (Bajari et al., 2018[63] ; OCDE, 2016[64] ; Varian, 2018[65]).

Des économies d'échelle pourraient être dégagées en termes de valeur conférée par des données supplémentaires. Si une entreprise affichant une qualité de données légèrement supérieure à celle de ses concurrents voit sa part de marché bondir, cela pourrait donner lieu à un effet de rétroaction positif. Plus de clients rime avec plus de données, ce qui renforce le cycle et permet à l'entreprise de consolider progressivement sa position sur le marché. Des économies d'échelle peuvent également être réalisées eu égard à l'expertise nécessaire pour bâtir des systèmes d'IA efficaces.

D'un autre côté, on s'inquiète du risque que les algorithmes favorisent la collusion en permettant aux entreprises de surveiller les conditions de marché, les prix et la réaction des concurrents aux variations de prix. Elles pourraient alors disposer d'outils nouveaux ou plus perfectionnés pour coordonner leurs stratégies, fixer les prix et mettre en place des ententes. D'aucuns spéculent en outre sur le risque que des algorithmes d'apprentissage profond plus sophistiqués puissent conduire à des résultats équivalents sans même que soient conclues des ententes formelles entre les concurrents, donc sans intervention humaine. Ce qui ne manquerait pas de poser des difficultés aux autorités de contrôle. Le droit de la concurrence exige en effet que, pour qu'une entente soit constatée et punie, il faut que des éléments probants attestent d'un accord ou d'un consentement entre les parties (OCDE, 2017[66]).

Propriété intellectuelle

Cette sous-section traite de certaines incidences potentielles de l'IA sur la propriété intellectuelle. Elle montre qu'il s'agit là d'un domaine en rapide évolution, qui commence à peine à faire l'objet de travaux analytiques fondés sur des données probantes. Les règles de propriété intellectuelle contribuent généralement à renforcer le degré et le rythme des découvertes, des inventions et de la diffusion des nouvelles technologies liées à l'IA. Elles sont comparables en ce sens aux règles applicables aux autres technologies protégées par des droits de propriété intellectuelle (DPI). Si elles doivent préserver les intérêts des inventeurs, des auteurs, des artistes et des propriétaires de marques, les politiques en matière de propriété intellectuelle doivent également tenir compte du potentiel de l'IA en tant que ressource à l'appui de nouvelles innovations.

La protection de l'IA par des DPI autres que ceux associés aux secrets commerciaux pourrait soulever de nouvelles questions quant aux incitations susceptibles d'encourager les innovateurs à divulguer leurs innovations liées à l'IA, notamment les algorithmes et leur entraînement. Le Bureau du Parlement européen a examiné, lors d'une conférence, trois types de brevets envisageables pour l'IA (OEB, 2018[67]). Le premier type a trait à l'« intelligence artificielle fondamentale » (« *core AI* » en anglais) ; il est souvent lié aux algorithmes qui, en tant que méthodes mathématiques, ne sont pas brevetables. Pour le deuxième type – qui porte sur les modèles entraînés et l'apprentissage automatique –, les demandes portant sur des variations et des fourchettes de paramètres pourraient poser problème. Troisièmement, des brevets pourraient être déposés sur l'IA en tant qu'outil dans un domaine d'application, défini d'après ses effets techniques. D'autres organisations internationales et des pays de l'OCDE étudient eux aussi les incidences de l'IA dans le domaine de la propriété intellectuelle[11].

La diffusion de l'IA soulève une autre problématique : des ajustements doivent-ils être apportés aux systèmes de protection de la propriété intellectuelle, dans un monde où les systèmes d'IA peuvent eux-mêmes créer des inventions (OCDE, 2018[68]) ? De fait, certains systèmes d'IA sont d'ores et déjà en mesure de produire des inventions brevetables, dans des domaines tels que la chimie, les produits pharmaceutiques et les biotechnologies. De nombreuses inventions portent par exemple sur la création de combinaisons originales de molécules pour former de nouveaux composés, ou sur l'identification de nouvelles propriétés de molécules existantes. Par exemple, KnIT, un outil d'apprentissage automatique mis au point par IBM, est parvenu à reconnaître des kinases – des enzymes catalysant le transfert de groupes de phosphates vers des substrats spécifiques. Ces kinases présentaient des propriétés particulières parmi un ensemble de kinases connues, qui ont été testées à titre expérimental. Les propriétés particulières de ces molécules ont été révélées par un logiciel, et des brevets ont été déposés pour protéger cette découverte. Ces aspects (ainsi que d'autres) liés à l'IA et aux DPI sont examinés par des organismes spécialisés de la zone OCDE, à l'instar de l'Office européen des brevets, du United States Patent and Trademark Office, ou de l'Organisation mondiale de la propriété intellectuelle. On pourrait également s'intéresser aux questions liées à la protection des droits d'auteur attachés aux données traitées par l'IA.

Petites et moyennes entreprises

Les politiques et programmes destinés à aider les petites et moyennes entreprises (PME) à opérer une transition vers l'IA revêtent un caractère prioritaire croissant. Il s'agit là d'un domaine en rapide évolution, qui commence à faire l'objet de travaux analytiques fondés sur des données probantes. Plusieurs pistes pourraient favoriser la mise en place d'écosystèmes numériques propices à l'adoption et l'utilisation de l'IA par les PME :

- Renforcer les compétences – un volet essentiel dans la mesure où les PME peinent à rivaliser pour attirer des spécialistes de l'IA par trop rares.
- Encourager la réalisation d'investissements ciblés dans des secteurs verticaux de choix. Les politiques visant à stimuler les investissements dans des applications de l'IA spécifiques dans le secteur français de l'agriculture, par exemple, pourraient bénéficier à l'ensemble des acteurs, y compris aux PME qui n'auraient pas les épaules pour investir seules (OCDE, 2018[14]).
- Aider les PME à accéder aux données, notamment via la création de plateformes d'échange de données.

- Faciliter l'accès des PME aux technologies liées à l'IA, notamment par le biais du transfert de technologies depuis les établissements publics de recherche, ainsi que l'accès à la puissance de calcul et aux plateformes infonuagiques (Allemagne, 2018[69]).
- Améliorer les mécanismes de financement afin d'aider les PME spécialisées dans l'IA à se développer, moyennant par exemple la création d'un fonds d'investissement public, ainsi que l'assouplissement et le relèvement des plafonds de financement des dispositifs en faveur de l'investissement dans les entreprises à forte intensité de savoir (RU, 2017[70]). La Commission européenne s'attache pour sa part à soutenir les PME européennes, notamment dans le cadre du projet AI4EU, une plateforme d'IA à la demande.

Cadre d'action à l'appui de l'innovation dans l'IA

L'OCDE analyse les changements qu'il conviendrait d'opérer dans les politiques d'innovation et d'autres politiques intéressant l'IA, dans le contexte de l'IA et d'autres transformations numériques (OCDE, 2018[49]). L'Organisation s'intéresse notamment aux moyens de renforcer l'adaptabilité, la réactivité et la souplesse des instruments d'action et des expérimentations connexes. Les pouvoirs publics peuvent recourir à l'expérimentation pour fournir des environnements contrôlés afin de tester les systèmes d'IA. De tels environnements pourraient intégrer des bacs à sable réglementaires, des centres d'innovation et des laboratoires des politiques. Les expérimentations de politiques peuvent se faire en « mode startup » : elles peuvent alors être déployées, évaluées et modifiées, transposées à une échelle supérieure ou inférieure, ou abandonnées rapidement, selon le cas.

Une autre solution pour parvenir à une prise de décision plus rapide et efficace consiste à recourir aux outils numériques pour concevoir les politiques (y compris les politiques d'innovation) et suivre la réalisation des objectifs y afférents. Par exemple, certains pays utilisent la modélisation multi-agents pour anticiper l'impact de diverses variantes des politiques sur les différents types d'entreprises.

Les pouvoirs publics peuvent encourager les acteurs de l'IA à mettre au point des mécanismes d'autoréglementation tels que des codes de conduite, des normes volontaires et des pratiques optimales. Ces dispositifs peuvent les guider tout au long du cycle de vie des systèmes d'IA, notamment pour le suivi, la communication, l'évaluation et le traitement des effets néfastes ou de l'utilisation abusive de ces systèmes.

Ils peuvent également instaurer et favoriser les mécanismes de surveillance des systèmes d'IA dans les secteurs public et privé, en tant que de besoin. Ces mécanismes peuvent intégrer des examens de conformité, des audits, des évaluations et des dispositifs de certification. Ils peuvent également être utiles lors de l'examen des besoins particuliers des PME et des contraintes auxquelles elles sont confrontées.

Se préparer à la transformation des emplois et renforcer les compétences

Emplois

L'IA devrait compléter le travail humain dans certaines tâches, le remplacer dans d'autres, et ouvrir la voie à de nouveaux types d'emplois

L'OCDE a réalisé une étude approfondie de l'impact de la transformation numérique sur l'emploi, ainsi que des incidences sur l'action des pouvoirs publics (OCDE, 2019[62]). Si

l'IA est un domaine en rapide évolution dans lequel les travaux analytiques fondés sur des données probantes ne font que commencer, on s'attend néanmoins à ce qu'elle modifie la nature du travail à mesure qu'elle se diffuse dans les différents secteurs. Elle est appelée à compléter le travail humain dans certaines tâches, le remplacer dans d'autres, et ouvrir la voie à de nouveaux types d'emplois. Cette section expose un certain nombre de mutations qui devraient intervenir sur les marchés du travail sous l'effet de l'IA, ainsi que les considérations intéressant l'action des pouvoirs publics, soulevées par la transition vers une économie de l'IA.

L'IA devrait stimuler la productivité

L'IA devrait stimuler la productivité de deux façons. D'une part, certaines activités menées à bien jusqu'à présent par des hommes vont être automatisées. D'autre part, avec l'avènement des machines autonomes, les systèmes fonctionneront et s'adapteront aux conditions moyennant un contrôle humain réduit, voire nul (OCDE, 2018[68] ; Autor et Salomons, 2018[71]). Des travaux de recherche portant sur 12 économies développées ont révélé que l'augmentation de la productivité du travail imputable à l'IA pourrait aller jusqu'à 40 % d'ici à 2035 par rapport aux niveaux de référence attendus (Purdy et Daugherty, 2016[72]). Les exemples sont légion. Le système Watson d'IBM assiste les conseillers des caisses du Crédit Mutuel à répondre aux questions des clients avec une rapidité augmentée de 60 %[12]. Lors de soldes en 2017, l'agent conversationnel d'Alibaba a traité plus de 95 % des demandes des clients. Les chargés de clientèle ont ainsi pu se charger des problématiques plus complexes ou personnelles (Zeng, 2018[73]). En théorie, l'augmentation de la productivité des travailleurs devrait se traduire par des hausses de salaires, puisque chaque employé produit davantage de valeur ajoutée.

La constitution d'équipes mêlant ressources humaines et IA permet de limiter les erreurs et d'ouvrir le champ des possibilités pour les travailleurs. Elles s'avèrent d'ailleurs plus productives que l'IA ou les travailleurs humains pris séparément (Daugherty et Wilson, 2018[74]). Ainsi, dans les usines de BMW, la constitution d'équipes mixtes de ce type a eu pour effet d'accroître la productivité manufacturière de 85 % par rapport à des équipes non intégrées. Les exemples ne se limitent pas aux activités industrielles : les robots de Walmart, par exemple, gèrent les stocks, de sorte que le personnel des magasins peut se concentrer sur l'assistance à la clientèle. Et lorsqu'un radiologue s'appuie sur des modèles d'IA pour réaliser des radiographies pulmonaires en cas de suspicion de tuberculose, la précision des diagnostics est de 100 % – soit un taux supérieur à celui atteint en cas de recours à l'IA ou d'intervention humaine seuls (Lakhani et Sundaram, 2017[75]).

L'IA peut également aider à améliorer et accélérer des tâches déjà automatisées. Elle permet ainsi aux entreprises de produire davantage, à moindre coût. Si la réduction des coûts est répercutée sur les prix aux entreprises ou aux individus, on peut s'attendre à une hausse de la demande de biens. Ce qui stimulerait la demande de main-d'œuvre à la fois au sein de l'entreprise concernée – dans des postes de production, par exemple – et dans les secteurs en aval, dans le cas de biens intermédiaires.

L'IA devrait modifier la physionomie des tâches automatisables – voire accélérer les mutations

L'automatisation n'est pas un phénomène nouveau, mais l'IA devrait modifier la physionomie des tâches susceptibles d'être automatisées, voire accélérer les mutations. Contrairement aux ordinateurs, les technologies liées à l'IA ne sont pas strictement préprogrammées et basées sur des règles. La diffusion des ordinateurs s'est traduite par une réduction des

emplois routiniers nécessitant un niveau de qualification intermédiaire. En revanche, les nouvelles applications faisant appel à l'IA sont de plus en plus à même d'exécuter des tâches relativement complexes impliquant de formuler des prévisions (voir chapitre 3). Ces tâches peuvent aller de la transcription à la traduction, en passant par la conduite de véhicules, l'établissement de diagnostics médicaux, ou le traitement des questions des clients (Graetz et Michaels, 2018[76] ; Michaels, Natraj et Van Reenen, 2014[77] ; Goos, Manning et Salomons, 2014[78])[13].

L'OCDE a réalisé des mesures préliminaires afin d'estimer la capacité des technologies à répondre aux questions de l'Évaluation des compétences des adultes (PIAAC) liées aux compétences à l'écrit et en calcul (Elliott, 2017[79]). Ces travaux ont montré qu'en 2017, les systèmes d'IA étaient à même de répondre aux questions portant sur les compétences à l'écrit à un niveau équivalent à celui de 89 % des adultes des pays de l'OCDE. En d'autres termes, seuls 11 % des adultes affichaient un niveau supérieur à celui que l'IA parvenait à reproduire en termes de maîtrise de la langue. L'étude tablait sur une augmentation de la pression économique en faveur de l'application des capacités informatiques pour certaines compétences à l'écrit et en calcul. Ce qui se traduirait par une baisse de la demande de travailleurs humains pour exécuter des tâches mobilisant des compétences à l'écrit de niveau faible à intermédiaire, à l'inverse des tendances observées récemment. L'étude soulignait en outre la difficulté de concevoir des politiques d'éducation pour les adultes disposant d'un niveau supérieur à celui des capacités informatiques actuelles. Elle proposait par conséquent de nouveaux outils et mesures d'incitation pour promouvoir les compétences des adultes ou associer des politiques de développement des compétences et d'autres interventions, dans des domaines tels que la protection sociale et le dialogue social (OCDE, 2018[14]).

Les incidences de l'IA sur les emplois dépendront de sa rapidité de diffusion dans différents secteurs

Les incidences de l'IA sur les emplois dépendront par ailleurs du rythme de développement et de diffusion des technologies liées à l'IA dans différents secteurs au cours des décennies à venir. Les véhicules autonomes devraient avoir des conséquences sur les emplois liés aux services de transport et de livraison. Des constructeurs de camions bien établis, à l'instar de Volvo et Daimler, par exemple, sont en compétition avec des startups comme Kodiak et Einride pour développer et tester des véhicules sans conducteur (Stewart, 2018[80]). Selon le Forum international des transports, les camions autonomes pourraient se multiplier sur les routes au cours des dix prochaines années. Quelque 50 à 70 % des 6.4 millions d'emplois de chauffeurs routiers professionnels aux États-Unis et en Europe pourraient disparaître d'ici à 2030 (FIT, 2017[81]). En parallèle, de nouveaux emplois seront toutefois créés pour fournir des services de support pour ce parc croissant de camions sans conducteur. De plus, les camions autonomes pourraient contribuer à réduire les frais d'exploitation liés au fret routier d'environ 30 %, notamment du fait de la diminution des coûts de main-d'œuvre. Cela pourrait entraîner la disparition d'entreprises de transport traditionnelles et, par ricochet, une baisse plus rapide encore des emplois de chauffeurs routiers.

Les technologies liées à l'IA devraient avoir des incidences sur les tâches exigeant traditionnellement un niveau de qualification plus élevé

Les technologies liées à l'IA exécutent des tâches de prévision généralement dévolues à des travailleurs très qualifiés – des juristes au personnel médical. Un robot avocat a par exemple aidé des automobilistes à contester des amendes de stationnement d'une valeur totale de 12 millions USD (Dormehl, 2018[82]). En 2016, les systèmes Watson d'IBM et

DeepMind Health ont obtenu de meilleurs résultats que des médecins humains dans le diagnostic de cancers rares (Frey et Osborne, 2017[83]). L'IA a également fait preuve d'une meilleure capacité à prévoir les variations des cours en bourse que les professionnels de la finance (Mims, 2010[84]).

L'IA peut compléter l'homme et créer de nouveaux types de travail

L'IA complète les travailleurs humains et devrait créer des possibilités d'emplois. Les domaines concernés sont ceux qui complètent les prévisions et exploitent les compétences comme la pensée critique, la créativité et l'empathie (EOP, 2016[85] ; OCDE, 2018[21]).

- **Scientifiques des données et experts en apprentissage automatique** : On a besoin de spécialistes pour créer et nettoyer les données, et programmer et développer les applications d'IA. Toutefois, bien que les données et l'apprentissage automatique donnent lieu à l'apparition de certaines tâches nouvelles, celles-ci ne devraient pas être pléthoriques pour les travailleurs.
- **Actions** : Certaines actions revêtent par nature davantage de valeur lorsqu'elles sont exécutées par des hommes (qu'il s'agisse d'athlètes de haut niveau, de professionnels de la petite enfance, ou de commerciaux) plutôt que par des machines. Beaucoup pensent que les humains vont se concentrer sur les emplois qui améliorent la qualité de vie, comme la garde d'enfants, le coaching sportif ou l'accompagnement des malades en fin de vie.
- **Jugement pour déterminer l'objet des prévisions** : Le facteur le plus important est probablement la capacité de jugement – à savoir le processus de détermination de l'intérêt d'une action particulière dans un environnement donné. Lorsque l'on recourt à l'IA pour établir des prévisions, un humain doit décider de ce que l'on va prévoir et de l'usage qui en sera fait. Énoncer des dilemmes, interpréter des situations ou extraire le sens d'un texte nécessitent entre autres des qualités de jugement et d'équité (OCDE, 2018[14]). En science, par exemple, l'IA peut compléter les personnes chargées du raisonnement conceptuel nécessaire pour bâtir les cadres de recherche et définir le contexte d'expériences spécifiques.
- **Jugement pour décider de l'usage à faire des prévisions** : Une décision ne peut être prise uniquement à partir d'une prévision. Par exemple, la décision somme toute banale d'emporter ou non un parapluie avant de sortir sera prise en tenant compte des prévisions sur les risques de précipitations, mais pas seulement : elle dépendra également, pour une large part, de préférences personnelles, selon que la personne déteste être mouillée ou ne souhaite pas s'embarrasser d'un parapluie. Cet exemple vaut pour de nombreuses décisions importantes. En cybersécurité, une prévision quant au caractère hostile d'une nouvelle requête doit être évaluée au regard du risque de refuser une requête amicale ou de laisser une requête hostile obtenir des informations non autorisées.

Les prévisions quant à l'impact net de l'IA sur la quantité de travail varient sensiblement

Au cours des cinq dernières années, des estimations divergentes ont été réalisées sur les conséquences globales de l'automatisation sur les pertes d'emplois (Winick, 2018[86] ; MGI, 2017[87] ; Frey et Osborne, 2017[83]). Par exemple, une étude de Frey et Osborne estimait que 47 % des emplois aux États-Unis étaient menacés de suppression au cours des 10 à 15 prochaines années. Adoptant une approche axée sur les tâches, le McKinsey Global

Institute a pour sa part estimé en 2017 qu'environ un tiers des activités dans 60 % des emplois étaient exposées à un risque d'automatisation. Néanmoins, l'automatisation des emplois identifiés était imputable non pas seulement au développement et au déploiement de l'IA, mais aussi à d'autres évolutions technologiques.

Sans compter qu'il est difficile de prévoir les futures créations d'emplois dans de nouveaux domaines. Selon une étude, l'IA devrait être à l'origine de deux millions de créations nettes d'emplois d'ici à 2025 (Gartner, 2017[88]). Des emplois devraient être créés à la fois dans le sillage de l'émergence de nouveaux métiers et par des canaux plus indirects. Par exemple, l'IA devrait contribuer à réduire les coûts de production des biens et des services et à en accroître la qualité. Ce qui devrait conduire à une hausse de la demande et, par ricochet, des emplois.

Les dernières estimations en date de l'OCDE tiennent compte de l'hétérogénéité des tâches dans des postes très spécifiques, en s'appuyant sur les données du Programme pour l'évaluation internationale des compétences des adultes (PIAAC). Si l'on se fonde sur les technologies existantes, 14 % des emplois sont fortement menacés d'automatisation dans les pays de l'OCDE ; et 32 % des travailleurs devraient voir leurs emplois sensiblement évoluer (Nedelkoska et Quintini, 2018[89]). Les travailleurs les plus jeunes et les plus âgés sont les groupes les plus exposés au risque d'automatisation. Une récente analyse de l'OCDE laisse entrevoir une baisse de l'emploi dans des métiers considérés comme exposés à un risque élevé d'automatisation dans 82 % des régions de 16 pays européens. Elle met en outre en lumière une augmentation plus marquée des emplois faiblement exposés dans 60 % des régions, augmentation qui compense les pertes d'emplois. Ces travaux tendent à confirmer que l'automatisation pourrait entraîner une mutation de la répartition des emplois, sans pour autant provoquer une baisse généralisée du niveau d'emploi (OCDE, 2018[90]).

L'IA est appelée à modifier la nature du travail

L'adoption de l'IA devrait modifier la nature du travail. L'IA pourrait contribuer à rendre le travail plus intéressant en favorisant l'automatisation des tâches répétitives et en ouvrant la voie à un travail plus flexible, voire à un meilleur équilibre entre vie professionnelle et vie privée. La créativité et l'ingéniosité humaines peuvent être conjuguées à l'augmentation des ressources en termes de puissance de calcul, de données et d'algorithmes pour créer de nouvelles tâches et activités faisant appel à leur tour à la créativité (Kasparov, 2018[91]).

Plus généralement, l'IA pourrait accélérer l'évolution du fonctionnement des marchés du travail en stimulant l'efficience. Aujourd'hui, les techniques axées sur l'IA, couplées aux données massives, promettent d'aider les entreprises à définir les rôles des travailleurs – et de contribuer à mettre en correspondance les travailleurs et les emplois. IBM, par exemple, utilise l'IA pour optimiser la formation de ses employés, leur recommandant des modules d'après leurs performances passées, leurs objectifs professionnels et les besoins en compétences de l'entreprise. De même, des sociétés comme KeenCorp et Vibe ont mis au point des techniques d'analyse de texte afin d'aider les entreprises à analyser les communications des employés pour faciliter l'établissement de mesures afférentes par exemple à leur état d'esprit, à leur productivité ou aux effets de réseau (Deloitte, 2017[92]). Grâce à ces informations, l'IA pourrait aider les entreprises à optimiser la productivité des travailleurs.

Les paramètres de changement organisationnel devront être définis

Il devient de plus en plus impératif de mettre en œuvre des normes sectorielles nouvelles ou révisées et des accords technologiques entre les directions et les employés afin de garantir un lieu de travail fiable, sûr et productif. Le Comité économique et social européen

(CESE) « préconise que les parties prenantes œuvrent ensemble en faveur de systèmes d'IA complémentaires et de leur mise en place conjointe sur le lieu de travail » (CESE, 2017[46]). Il importe en outre de favoriser une certaine flexibilité, tout en préservant l'autonomie des travailleurs et la qualité des emplois, y compris en termes de partage des bénéfices. La convention collective conclue récemment entre le syndicat de branche allemand *IG Metall* et les employeurs (*Gesamtmetall)* illustre la faisabilité économique de la mise en place de temps de travail variables. Elle montre en effet que, selon les besoins organisationnels et personnels dans le nouveau monde du travail, les employeurs et les syndicats peuvent parvenir à des accords sans que cela passe par une révision de la législation en matière de protection de l'emploi (Byhovskaya, 2018[93]).

L'utilisation de l'IA pour soutenir les fonctions des marchés du travail – avec des garanties – s'avère également prometteuse

L'IA contribue d'ores et déjà à accroître l'efficience de la mise en correspondance des offres et des demandes d'emploi, ainsi que de la formation. Elle peut aider à orienter les demandeurs d'emploi, y compris ceux dont l'emploi a été supprimé, vers les programmes de valorisation de la main-d'œuvre dont ils ont besoin en vue d'acquérir les qualifications nécessaires pour accéder aux métiers émergents ou en expansion. Dans de nombreux pays de l'OCDE, employeurs et services publics de l'emploi ont déjà recours à des plateformes électroniques pour pourvoir les emplois (OCDE, 2018[90]). À l'avenir, l'IA et d'autres technologies numériques permettront de mettre en place des approches innovantes et personnalisées des processus de recherche d'emploi et de recrutement, et de renforcer l'efficience de l'appariement des offres et des demandes d'emploi. C'est ainsi que la plateforme LinkedIn utilise l'IA pour aider les recruteurs à identifier les bons candidats et aiguiller les candidats vers les emplois les mieux adaptés à leur recherche. Elle s'appuie pour ce faire sur les données relatives au profil et à l'activité de ses 470 millions d'utilisateurs enregistrés (Wong, 2017[94]).

Les technologies liées à l'IA qui exploitent les données massives peuvent également éclairer les pouvoirs publics, les employeurs et les travailleurs sur les conditions des marchés du travail locaux. Ces informations les aident à identifier et prévoir les besoins en compétences, orienter les ressources de formation et guider les travailleurs vers les emplois. Plusieurs pays, tels la Finlande, la République tchèque et la Lettonie, mènent actuellement des projets en vue de développer les informations sur les marchés du travail (OCDE, 2018[90]).

Instaurer une gouvernance de l'utilisation des données des travailleurs

Si l'IA doit s'appuyer sur des ensembles de données volumineux pour être productive, des risques émergent dès lors que ces données concernent les travailleurs individuels, en particulier si les systèmes d'IA qui analysent ces données présentent un fonctionnement opaque. Or la planification des ressources humaines et de la productivité s'appuieront de plus en plus sur les données des salariés et les algorithmes. Les décideurs et les parties prenantes pourraient par conséquent s'intéresser à la manière dont la collecte et le traitement des données influent sur les perspectives et les conditions d'emploi. Les données peuvent être recueillies à partir des applications, des empreintes, des technologies prêt-à-porter et des capteurs en temps réel indiquant la localisation et le lieu de travail des employés. Dans le domaine des services à la clientèle, les logiciels basés sur l'AI analysent l'intonation des employés. Toutefois, selon les témoignages de certains salariés, ils ne tiennent pas compte des modèles de voix et il est difficile d'en contester les résultats (UNI, 2018[95]).

En revanche, les accords sur l'utilisation des données des employés et le droit à la déconnexion font leur apparition dans certains pays. L'opérateur de télécommunications Orange France Télécom et cinq centres syndicaux ont été parmi les premiers à prendre des engagements en faveur de la protection des données des employés. Les dispositions portent notamment sur la transparence quant à l'utilisation des données, la formation et l'introduction de nouveaux équipements. Pour parer aux lacunes réglementaires sur la gestion des données des travailleurs, des mesures pourraient être prises pour mettre en place des organes de gouvernance des données dans les entreprises et établir la responsabilité au regard de l'utilisation des données (personnelles), ainsi que des droits en matière de portabilité, d'explication et de suppression des données (UNI, 2018[95]).

Gérer la transition vers l'IA

Des politiques doivent être mises en place pour gérer la transition vers l'IA, notamment dans le domaine de la protection sociale

Les changements organisationnels n'intervenant pas au même rythme que le développement des technologies, des perturbations et des turbulences pourraient se manifester sur les marchés du travail (OCDE, 2018[14]). Les projections optimistes sur le long terme ne signifient pas pour autant que la transition vers une économie de plus en plus irriguée par l'IA se fera sans heurts : certains secteurs vont vraisemblablement croître, d'autres décliner. Des emplois sont menacés de disparaître, tandis que de nouveaux se créent. Par conséquent, l'enjeu phare de l'action publique sur les questions d'IA et d'emploi sera de gérer la transition, en agissant sur les politiques en matière de protection sociale, d'assurance maladie, d'imposition progressive du travail et du capital, et d'éducation. Les analyses de l'OCDE mettent par ailleurs en évidence la nécessité de prêter attention aux politiques de concurrence et autres politiques susceptibles d'influer sur les phénomènes de concentration, le pouvoir de marché et la répartition des revenus (OCDE, 2019[62]).

Compétences requises pour utiliser l'IA

La mutation des emplois s'accompagne d'une évolution des compétences nécessaires aux travailleurs

La mutation des emplois s'accompagne d'une évolution des compétences nécessaires aux travailleurs (OCDE, 2017[96] ; Acemoglu et Restrepo, 2018[97] ; Brynjolfsson et Mitchell, 2017[98]). Cette sous-section expose quelques-unes des répercussions possibles de l'IA sur les compétences, soulignant qu'il s'agit là d'un domaine en rapide évolution, qui commence à peine à faire l'objet de travaux analytiques fondés sur des données probantes. Des ajustements devront vraisemblablement être apportés aux politiques d'éducation afin d'étendre l'apprentissage tout au long de la vie, la formation et le développement des compétences. Comme pour les autres technologies, l'IA devrait créer une demande de compétences dans trois domaines. Premièrement, on aura besoin de **compétences spécialisées** pour programmer et développer les applications d'IA. Ces compétences sont requises dans différents domaines, depuis la recherche fondamentale, l'ingénierie et le développement d'applications en lien avec l'IA, jusqu'à la science des données et la pensée computationnelle. Deuxièmement, les individus devront disposer de **compétences génériques** pour exploiter l'IA, afin par exemple de pouvoir travailler dans des équipes mixtes IA/travailleurs humains dans les ateliers de fabrication ou pour les activités de contrôle qualité. Troisièmement, l'IA nécessitera des **compétences complémentaires**. Il s'agira notamment de mettre à profit des qualités humaines comme la pensée critique ; la créativité, l'innovation et l'entrepreneuriat ; ou encore l'empathie (EOP, 2016[85] ; OCDE, 2018[21]).

Des initiatives devront être mises en place pour développer et renforcer les compétences en IA nécessaires pour parer à la pénurie actuelle dans ce domaine

La pénurie de compétences en IA devrait s'accentuer et pourrait devenir plus prégnante encore avec l'essor de la demande de spécialistes dans des domaines comme l'apprentissage automatique. Les PME, les universités publiques et les centres de recherche sont d'ores et déjà en concurrence avec les entreprises dominantes pour attirer les talents. Des initiatives visant à développer et renforcer les compétences en IA voient peu à peu le jour dans les secteurs public, privé et universitaire. Par exemple, à Singapour, le gouvernement a mis en place un programme de recherche sur cinq ans sur la gouvernance de l'IA et l'utilisation des données à la Singapore Management University. Son Centre de gouvernance de l'IA et des données (Centre for AI & Data Governance) mène des recherches intéressant le secteur industriel, ciblées sur l'IA et l'industrie, la société et la commercialisation. Côté universitaire, le Massachusetts Institute of Technology (MIT) s'est engagé à consacrer 1 milliard USD à la création du Schwarzman College of Computing. L'objectif est de doter les étudiants et les chercheurs de tous horizons des compétences dont ils ont besoin pour utiliser l'informatique et l'IA en vue de faire progresser leur discipline, et inversement.

La pénurie de compétences en IA a également poussé certains pays à rationaliser les règles d'immigration pour les experts hautement qualifiés. Ainsi, le Royaume-Uni a doublé le nombre de visas *Tier 1* (*Exceptional Talent*), qu'il a porté à 2 000 par an, et rationalisé les règles permettant aux meilleurs étudiants et chercheurs de travailler dans le pays (RU, 2017[99]). Dans la même veine, le Canada a fixé à deux semaines les délais de traitement des demandes de permis de travail émanant de personnes hautement qualifiées et mis en place des dispenses de permis pour les missions de recherche de courte durée. Ces mesures s'inscrivent dans le cadre de la Stratégie en matière de compétences mondiales, adoptée par le Canada en 2017 pour attirer des travailleurs hautement qualifiés et des chercheurs étrangers (Canada, 2017[100]).

Compétences génériques requises pour exploiter l'IA

Tous les pays de l'OCDE évaluent les compétences et tentent d'anticiper les besoins immédiats et à moyen et long termes. La Finlande a ainsi proposé de mettre en place un Programme d'intelligence artificielle qui prévoit un « compte de compétences » ou un programme de bons d'échange pour bénéficier de formations continues afin de stimuler la demande d'éducation et de formation (Finlande, 2017[101]). Le Royaume-Uni promeut pour sa part une main-d'œuvre diversifiée formée à l'IA et investit environ 406 millions GBP (530 millions USD) dans le développement des compétences. Le pays concentre ses efforts sur la science, la technologie, l'ingénierie et les mathématiques, ainsi que sur la formation des enseignants en sciences informatiques (RU, 2017[99]).

Les professionnels doivent désormais disposer d'une double expertise (ce que certains appellent en anglais des « *bilinguals* »), à savoir être spécialisés dans une discipline comme l'économie, la biologie ou le droit, tout en disposant de compétences dans les techniques d'IA telles que l'apprentissage automatique. De la même veine, le MIT a annoncé en octobre 2018 l'évolution la plus importante de sa structure depuis 50 ans : créer une école informatique indépendante de la filière Ingénierie, avec des interconnexions avec tous les autres départements universitaires. On y enseignera aux étudiants l'art d'appliquer les techniques d'IA et d'apprentissage automatique aux défis qui se posent dans leurs propres disciplines. Il s'agit là d'un véritable tournant dans la façon dont le MIT enseigne les sciences informatiques. L'établissement consacre un milliard USD à la création de cette nouvelle école au sein du MIT (MIT, 2018[102]).

Compétences complémentaires

Les compétences non techniques font l'objet d'une attention accrue. Des travaux de recherche ont montré qu'elles peuvent avoir trait au jugement, à l'analyse et à la communication interpersonnelle (Agrawal, Gans et Goldfarb, 2018[103] ; Deming, 2017[104] ; Trajtenberg, 2018[105]). En 2021, l'OCDE intègrera au Programme international pour le suivi des acquis des élèves (PISA) un module destiné à tester les compétences en matière de créativité et de pensée critique. Les résultats aideront à établir une référence pour l'évaluation de la créativité dans les différents pays, afin d'étayer l'action des pouvoirs publics et des partenaires sociaux.

Mesure

La mise en œuvre d'une IA centrée sur l'humain et digne de confiance dépend du contexte. Toutefois, pour tenir leur engagement en ce sens, les décideurs devront définir des objectifs et des mesures permettant d'évaluer les performances des systèmes d'IA, dans des domaines tels que la fiabilité, l'efficience, la réalisation des objectifs sociétaux, l'équité et la robustesse.

Références

Abrams, M. et al. (2017), *Artificial Intelligence, Ethics and Enhanced Data Stewardship*, The Information Accountability Foundation, Plano, Texas. [17]

Acemoglu, D. et P. Restrepo (2018), *Artificial Intelligence, Automation and Work*, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w24196. [97]

Agrawal, A., J. Gans et A. Goldfarb (2018), « Economic Policy for Artificial Intelligence », *National Bureau of Economic Research, Cambridge, MA,* 24690, http://dx.doi.org/10.3386/w24690. [50]

Agrawal, A., J. Gans et A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business School Press. [103]

Agrawal, G. (dir. pub.) (2018), « Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics », *National Bureau of Economic Research,* 24001, https://www.nber.org/papers/w24001. [51]

Allemagne (2018), « Key points for a federal government strategy on artificial intelligence », communiqué de presse, 18 juillet, BMWI, https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2018/20180718-key-points-for-federal-government-strategy-on-artificial-intelligence.html. [69]

Autor, D. et A. Salomons (2018), « Is automation labor-displacing? Productivity growth, employment, and the labor share », *document de travail*, n° 24871, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w24871. [71]

Bajari, P. et al. (2018), « The impact of big data on firm performance: An empirical investigation », *document de travail*, n° 24334, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w24334. [63]

Barocas, S. et A. Selbst (2016), « Big data's disparate impact », *California Law Review*, vol. 104, pp. 671-729, http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf. [30]

Berk, R. et J. Hyatt (2015), « Machine learning forecasts of risk to inform sentencing decisions », *Federal Sentencing Reporter*, vol. 27/4, pp. 222-228, http://dx.doi.org/10.1525/fsr.2015.27.4.222. [24]

Borges, G. (2017), *Liability for Machine-Made Decisions: Gaps and Potential Solutions*, presentation at the "AI: Intelligent Machines, Smart Policies" conference, Paris, 26-27 October, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-borges.pdf. [43]

Brundage, M. et al. (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Centre for a New American Security, Electronic Frontier Foundation and Open AI, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf. [38]

Brynjolfsson, E. et T. Mitchell (2017), « What can machine learning do? Workforce implications », *Science*, vol. 358/6370, pp. 1530-1534, http://dx.doi.org/10.1126/science.aap8062. [98]

Burgess, M. (2016), « Holding AI to account: Will algorithms ever be free of bias if they are created by humans? », *WIRED,* 11 janvier, https://www.wired.co.uk/article/creating-transparent-ai-algorithms-machine-learning. [32]

Byhovskaya, A. (2018), *Overview of the National Strategies on Work 4.0: A Coherent Analysis of the Role of the Social Partners*, Comité économique et social européen, Bruxelles, https://www.eesc.europa.eu/sites/default/files/files/qe-02-18-923-en-n.pdf. [93]

Canada (2017), « Le gouvernement du Canada lance la Stratégie en matière de compétences mondiales », *communiqué de presse*, Immigration, Réfugiés et Citoyenneté Canada, 12 juin, https://www.canada.ca/fr/immigration-refugies-citoyennete/nouvelles/2017/06/le_gouvernement_ducanadalancelastrategieenmatieredecompetencesmo.html. [100]

Cellarius, M. (2017), *Artificial Intelligence and the Right to Informational Self-determination*, Forum de l'OCDE, OCDE, Paris, https://www.oecd-forum.org/users/75927-mathias-cellarius/posts/28608-artificial-intelligence-and-the-right-to-informational-self-determination. [10]

CESE (2017), *L'intelligence artificielle – Les retombées de l'intelligence artificielle pour le marché unique (numérique), la production, la consommation, l'emploi et la société*, Comité économique et social européen, Bruxelles, https://webapi2016.eesc.europa.eu/v1/documents/eesc-2016-05369-00-00-ac-tra-fr.docx/content. [46]

Chouldechova, A. (2016), « Fair prediction with disparate impact: A study of bias in recidivism prediction instruments », *arXiv, Cornell University*, vol. 07524, https://arxiv.org/abs/1610.07524. [25]

Citron, D. et F. Pasquale (2014), « The scored society: Due process for automated predictions », *Washington Law Review*, vol. 89, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209. [37]

Cockburn, I., R. Henderson et S. Stern (2018), « The impact of artificial intelligence on innovation », *document de travail*, n° 24449, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w24449. [52]

Crawford, K. (2016), « Artificial intelligence's white guy problem », *New York Times,* 26 June, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=0. [31]

Daugherty, P. et H. Wilson (2018), *Human Machine: Reimagining Work in the Age of AI*, Harvard Business Review Press, Cambridge, MA. [74]

Deloitte (2017), *HR Technology Disruptions for 2018: Productivity, Design and Intelligence Reign*, Deloitte, http://marketing.bersin.com/rs/976-LMP-699/images/HRTechDisruptions2018-Report-100517.pdf. [92]

Deming, D. (2017), « The growing importance of social skills in the labor market », *The Quarterly Journal of Economics*, vol. 132/4, pp. 1593-1640, http://dx.doi.org/10.1093/qje/qjx022. [104]

Dormehl, L. (2018), « Meet the British whiz kid who fights for justice with robo-lawyer sidekick », *Digital Trends* 3 mars, https://www.digitaltrends.com/cool-tech/robot-lawyer-free-acess-justice/. [82]

Doshi-Velez, F. et al. (2017), « Accountability of AI under the law: The role of explanation », *arXiv, Cornell University,* 21 novembre, https://arxiv.org/pdf/1711.01134.pdf. [29]

Dowlin, N. (2016), *CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy*, Microsoft Research, https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/CryptonetsTechReport.pdf. [61]

Dressel, J. et H. Farid (2018), « The accuracy, fairness and limits of predicting recidivism », *Science Advances*, vol. 4/1, http://advances.sciencemag.org/content/4/1/eaao5580. [34]

Elliott, S. (2017), *Computers and the Future of Skill Demand*, La recherche et l'innovation dans l'enseignement, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264284395-en. [79]

EOP (2016), *Artificial Intelligence, Automation and the Economy*, Executive Office of the President, Gouvernement des États-Unis, https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED AI Economy Report.pdf. [85]

Finlande (2017), *Finland's Age of Artificial Intelligence - Turning Finland into a Leader in the Application of AI*, page web, Ministère finlandais de l'Emploi et de l'Économie, https://tem.fi/en/artificial-intelligence-programme. [101]

FIT (2017), « Driverless trucks: New report maps out global action on driver jobs and legal issues », *International Transport Forum*, https://www.itf-oecd.org/driverless-trucks-new-report-maps-out-global-action-driver-jobs-and-legal-issues. [81]

Flanagan, M., D. Howe et H. Nissenbaum (2008), « Embodying values in technology: Theory and practice », dans van den Hoven, J. et J. Weckert (dir. pub.), *Information Technology and Moral Philosophy*, Cambridge University Press, Cambridge, http://dx.doi.org/10.1017/cbo9780511498725.017. [16]

Freeman, R. (2017), *Evolution or Revolution? The Future of Regulation and Liability for AI*, presentation at the "AI: Intelligent Machines, Smart Policies" conference, Paris, 26-27 October, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-freeman.pdf. [41]

Frey, C. et M. Osborne (2017), « The future of employment: How susceptible are jobs to computerisation? », *Technological Forecasting and Social Change*, vol. 114, pp. 254-280, http://dx.doi.org/10.1016/j.techfore.2016.08.019. [83]

Gartner (2017), « Gartner says by 2020, artificial intelligence will create more jobs than it eliminates », Gartner, communiqué de presse, 13 décembre, https://www.gartner.com/en/newsroom/press-releases/2017-12-13-gartner-says-by-2020-artificial-intelligence-will-create-more-jobs-than-it-eliminates. [88]

Golson, J. (2016), « Google's self-driving cars rack up 3 million simulated miles every day », *The Verge,* 1 février, https://www.theverge.com/2016/2/1/10892020/google-self-driving-simulator-3-million-miles. [44]

Goodfellow, I., J. Shlens et C. Szegedy (2015), « Explaining and harnessing adversarial examples », *arXiv*, vol. 1412.6572, Cornell University, https://arxiv.org/pdf/1412.6572.pdf. [39]

Goos, M., A. Manning et A. Salomons (2014), « Explaining job polarization: Routine-biased technological change and offshoring », *American Economic Review*, vol. 104/8, pp. 2509-2526, http://dx.doi.org/10.1257/aer.104.8.2509. [78]

Graetz, G. et G. Michaels (2018), « Robots at work », *Review of Economics and Statistics*, vol. 100/5, pp. 753-768, http://dx.doi.org/10.1162/rest_a_00754. [76]

Harkous, H. (2018), « Polisis: Automated analysis and presentation of privacy policies using deep learning », *arXiv, Cornell University,* 29 juin, https://arxiv.org/pdf/1802.02561.pdf. [15]

HCDH (2011), *Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme*, Haut-Commissariat des Nations Unies aux droits de l'homme, https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_fr.pdf. [7]

Heiner, D. et C. Nguyen (2018), « Amplify Human Ingenuity with Intelligent Technology », *Shaping human-centered artificial intellilgence, A.Ideas Series*, Réseau du Forum, OCDE, Paris, https://www.oecd-forum.org/users/86008-david-heiner-and-carolyn-nguyen/posts/30653-shaping-human-centered-artificial-intelligence. [6]

Heiner, D. et C. Nguyen (2018), « Amplify Human Ingenuity with Intelligent Technology », *Shaping Human-Centered Artificial Intelligence, A.Ideas Series*, The Forum Network, OCDE, Paris, https://www.oecd-forum.org/users/86008-david-heiner-and-carolyn-nguyen/posts/30653-shaping-human-centered-artificial-intelligence. [9]

Helgason, S. (1997), *Vers un principe de responsabilité fondée sur la performance : éléments de discussion*, Service de la gestion publique, Éditions OCDE, Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=PUMA/PAC(97)8&docLanguage=Fr. [47]

Ingels, H. (2017), *Artificial Intelligence and EU Product Liability Law*, presentation at the "AI: Intelligent Machines, Smart Policies" conference, Paris, 26-27 October, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-ingels.pdf. [42]

Jain, S. (2017), « NanoNets : How to use deep learning when you have limited data, Part 2 : Building object detection models with almost no hardware », *Medium* 30 janvier, https://medium.com/nanonets/nanonets-how-to-use-deep-learning-when-you-have-limited-data-f68c0b512cab. [59]

Kasparov, G. (2018), *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, Public Affairs, New York. [91]

Kendall, A. (23 mai 2017), « Deep learning is not good enough, we need Bayesian deep learning for safe AI », Alex Kendall Blog, https://alexgkendall.com/computer_vision/bayesian_deep_learning_for_safe_ai/. [60]

Knight, W. (2017), « The financial world wants to open AI's black boxes », *MIT Technology Review,* 13 April, https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes/. [33]

Kosack, S. et A. Fung (2014), « Does transparency improve governance? », *Annual Review of Political Science*, vol. 17, pp. 65-87, https://www.annualreviews.org/doi/pdf/10.1146/annurev-polisci-032210-144356. [27]

Kosinski, M., D. Stillwell et T. Graepel (2013), « Private traits and attributes are predictable from digital records of human behavior », *PNAS,* 11 mars, http://www.pnas.org/content/pnas/early/2013/03/06/1218772110.full.pdf. [2]

Kurakin, A., I. Goodfellow et S. Bengio (2017), « Adversarial examples in the physical world, », *arXiv, Cornell University* 02533, https://arxiv.org/abs/1607.02533. [40]

Lakhani, P. et B. Sundaram (2017), « Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks », *Radiology*, vol. 284/2, pp. 574-582, http://dx.doi.org/10.1148/radiol.2017162326. [75]

Matheson, R. (2018), *Artificial intelligence model "learns" from patient data to make cancer treatment less toxic*, MIT News, 9 août 2018, http://news.mit.edu/2018/artificial-intelligence-model-learns-patient-data-cancer-treatment-less-toxic-0810. [107]

MGI (2017), *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*, McKinsey Global Institute, New York. [87]

Michaels, G., A. Natraj et J. Van Reenen (2014), « Has ICT polarized skill demand? Evidence from eleven countries over twenty-five years », *Review of Economics and Statistics*, vol. 96/1, pp. 60-77, http://dx.doi.org/10.1162/rest_a_00366. [77]

Mims, C. (2010), « AI that picks stocks better than the pros », *MIT Technology Review,* 10 June, https://www.technologyreview.com/s/419341/ai-that-picks-stocks-better-than-the-pros/. [84]

MIT (2018), « Cybersecurity's insidious new threat: Workforce stress », *MIT Technology Review,* 7 August, https://www.technologyreview.com/s/611727/cybersecuritys-insidious-new-threat-workforce-stress/. [102]

Mousave, S., M. Schukat et E. Howley (2018), « Deep reinforcement learning: An overview », *arXiv,* 1806.08894, https://arxiv.org/abs/1806.08894. [57]

Narayanan, A. (2018), « Tutorial: 21 fairness definitions and their politics », https://www.youtube.com/watch?v=jIXIuYdnyyk. [18]

Nedelkoska, L. et G. Quintini (2018), « Automation, skills use and training », *Documents de travail de l'OCDE sur les questions sociales, l'emploi et les migrations*, n° 202, Éditions OCDE, Paris, https://dx.doi.org/10.1787/2e2f4eea-en. [89]

Neppel, C. (2017), *AI: Intelligent Machines, Smart Policies*, exposé présenté à la conférence AI: Intelligent Machines, Smart Policies, Paris, les 26-27 octobre 2007, http://oe.cd/ai2017. [56]

NITI (2018), *National Strategy for Artificial Intelligence #AIforall*, NITI Aayog, juin 2018, http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf. [5]

OCDE (2019), *Enhanced Access to Data and Sharing of Data (EASD)*, Groupe de travail sur la sécurité et la vie privée dans l'économie numérique , DSTI/CDEP/SPDE(2017)13/REV3. [55]

OCDE (2019), *Going Digital: Shaping Policies, Improving Lives*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264312012-en. [62]

OCDE (2019), *Recommandation du Conseil sur l'intelligence artificielle*, OCDE, Paris, https://legalinstruments.oecd.org/api/print?ids=648&lang=fr. [36]

OCDE (2019), *Scoping Principles to Foster Trust in and Adoption of AI – Proposal by the Expert Group on Artificial Intelligence at the OECD (AIGO)*, Éditions OCDE, Paris, http://oe.cd/ai. [35]

OCDE (2018), « AI: Intelligent machines, smart policies: Conference summary », *Documents de travail de l'OCDE sur l'économie numérique*, n° 270, Éditions OCDE, Paris, https://dx.doi.org/10.1787/f1a650d9-en. [14]

OCDE (2018), « Approaches to market openness in the digital age », dans *« Perspectives on innovation policies in the digital age », dans OECD Science, Technology and Innovation Outlook 2018 : Adapting to Technological and Societal Disruption*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-8-en. [49]

OCDE (2018), *Job Creation and Local Economic Development 2018: Preparing for the Future of Work*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264305342-en. [90]

OCDE (2018), *La prochaine révolution de la production : Conséquences pour les pouvoirs publics et les entreprises*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264280793-fr. [68]

OCDE (2018), *Perspectives de l'économie numérique de l'OCDE 2017*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264282483-fr. [21]

OCDE (2018), *Science, technologie et innovation : Perspectives de l'OCDE 2018 (version abrégée) : S'adapter aux bouleversements technologiques et sociétaux*, Éditions OCDE, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-fr. [53]

OCDE (2017), *Algorithms and Collusion: Competition Policy in the Digital Age*, Éditions OCDE, Paris, https://www.oecd.org/fr/concurrence/algorithms-collusion-competition-policy-in-the-digital-age.htm. [66]

OCDE (2017), *Getting Skills Right: Skills for Jobs Indicators*, Getting Skills Right, Éditions OCDE, Paris, https://dx.doi.org/10.1787/9789264277878-en. [96]

OCDE (2016), *Données massives : Adapter la politique de la concurrence à l'ère du numérique (Synthèse)*, Comité de la concurrence, https://one.oecd.org/document/DAF/COMP/M(2016)2/ANN4/FINAL/fr/pdf. [64]

OCDE (2013), *Recommandation du Conseil concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel*, OCDE, Paris, https://www.oecd.org/fr/internet/ieconomie/lignesdirectricesregissantlaprotectiondelavieprivee etlesfluxtransfrontieresdedonneesdecaracterepersonnel.htm. [13]

OCDE (2011), *Les principes directeurs de l'OCDE à l'intention des entreprises multinationales*, Editions OCDE, Paris, https://doi.org/10.1787/9789264115439-fr. [8]

OEB (2018), *Patenting Artificial Intelligence - Conference summary*, Office européen des brevets, Munich, 30 mai, http://documents.epo.org/projects/babylon/acad.nsf/0/D9F20464038C0753C125829E0031B814/$FILE/summary_conference_artificial_intelligence_en.pdf. [67]

Office of the Victorian Information Commissioner (2018), « Artificial intelligence and privacy », *Issues Paper*, juin, Office of the Victorian Information Commissioner, https://ovic.vic.gov.au/wp-content/uploads/2018/08/AI-Issues-Paper-V1.1.pdf. [12]

O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Broadway Books. [26]

OpenAI (16 mai 2018), « AI and compute », OpenAI blog, San Francisco, https://blog.openai.com/ai-and-compute/. [54]

Pan, S. et Q. Yang (2010), « A survey on transfer learning », *IEEE Transactions on Knowledge and Data Engineering*, vol. 22/10, pp. 1345-1359. [58]

Patki, N., R. Wedge et K. Veeramachaneni (2016), « The Synthetic Data Vault », *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, http://dx.doi.org/10.1109/dsaa.2016.49. [106]

Privacy International et Article 19 (2018), *Privacy and Freedom of Expression in the Age of Artificial Intelligence*, https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf. [11]

Purdy, M. et P. Daugherty (2016), « Artificial intelligence poised to double annual economic growth rate in 12 developed economies and boost labor productivity by up to 40 percent by 2035, according to new research by Accenture », Accenture, Press Release, 28 septembre, http://www.accenture.com/futureofAI. [72]

RU (2017), *UK Digital Strategy*, Gouvernement du Royaume-Uni, https://www.gov.uk/government/publications/uk-digital-strategy. [70]

RU (2017), *UK Industrial Strategy: A Leading Destination to Invest and Grow*, Royaume-Uni et Irlande du Nord, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/668161/the_labour_market_story-_skills_use_at_work.pdf. [99]

Selbst, A. (2017), « Disparate impact in big data policing », *Georgia Law Review*, vol. 52/109, http://dx.doi.org/10.2139/ssrn.2819182. [22]

Simonite, T. (2018), « Probing the dark side of Google's ad-targeting system », *MIT Technology Review,* 6 July, https://www.technologyreview.com/s/539021/probing-the-dark-side-of-googles-ad-targeting-system/. [20]

Slusallek, P. (2018), *Artificial Intelligence and Digital Reality: Do We Need a CERN for AI?*, The Forum Network, OCDE, Paris, https://www.oecd-forum.org/channels/722-digitalisation/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai. [45]

Smith, M. et S. Neupane (2018), *Artificial Intelligence and Human Development: Toward a Research Agenda*, Centre de recherches pour le développement international, Ottawa, https://idl-bnc-idrc.dspacedirect.org/handle/10625/56949. [4]

Stewart, J. (2018), « As Uber gives up on self-driving trucks, another startup jumps in », *WIRED,* 8 July, https://www.wired.com/story/kodiak-self-driving-semi-trucks/. [80]

Talbot, D. et al. (2017), « Charting a roadmap to ensure AI benefits all », *Medium,* 30 November, https://medium.com/berkman-klein-center/charting-a-roadmap-to-ensure-artificial-intelligence-ai-benefits-all-e322f23f8b59. [3]

Trajtenberg, M. (2018), « AI as the next GPT: A political-economy perspective », *National Bureau of Economic Research*, vol. 24245, Cambridge, MA, http://dx.doi.org/10.3386/w24245. [105]

UNI (2018), *10 Principles for Workers' Data Rights and Privacy*, UNI Global Union, http://www.thefutureworldofwork.org/docs/10-principles-for-workers-data-rights-and-privacy/. [95]

Varian, H. (2018), « Artificial intelligence, economics and industrial organization », n° 24839, National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w24839. [65]

Wachter, S., B. Mittelstadt et L. Floridi (2017), « Transparent, explainable and accountable AI for robotics », *Science Robotics,* 31 May, http://robotics.sciencemag.org/content/2/6/eaan6080. [48]

Wachter, S., B. Mittelstadt et C. Russell (2017), « Counterfactual explanations without opening the black box: Automated decisions and the GDPR », *arXiv, Cornell University,* 00399, https://arxiv.org/pdf/1711.00399.pdf. [28]

Weinberger, D. (2018), « Optimization over explanation - Maximizing the benefits of machine learning without sacrificing its intelligence », *Medium,* 28 January, https://medium.com/@dweinberger/optimization-over-explanation-maximizing-the-benefits-we-want-from-machine-learning-without-347ccd9f3a66. [1]

Weinberger, D. (2018), *Playing with AI Fairness*, Google PAIR, 17 septembre, https://pair-code.github.io/what-if-tool/ai-fairness.html. [23]

Winick, E. (2018), « Every study we could find on what automation will do to jobs, in one chart », *MIT Technology Review,* 25 January, https://www.technologyreview.com/s/610005/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/. [86]

Wong, Q. (2017), « "At LinkedIn, artificial intelligence is like "oxygen"" », *Mercury News,* 1 June, http://www.mercurynews.com/2017/01/06/at-linkedin-artificial-intelligence-is-like-oxygen. [94]

Yona, G. (2017), « A gentle introduction to the discussion on algorithmic fairness », *Toward Data Science,* 5 October, https://towardsdatascience.com/a-gentle-introduction-to-the-discussion-on-algorithmic-fairness-740bbb469b6. [19]

Zeng, M. (2018), *Alibaba and the Future of Business*, Harvard Business Review, septembre-octobre 2018, https://hbr.org/2018/09/alibaba-and-the-future-of-business. [73]

Notes

[1] Pour en savoir plus, consulter la page https://www.microsoft.com/en-us/ai/ai-for-good.

[2] Voir : https://deepmind.com/applied/deepmind-ethics-society/.

[3] Voir : https://www.blog.google/technology/ai/ai-principles/.

[4] Voir : https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf.

[5] Voir l'Organisation internationale du Travail, les Principes directeurs de l'OCDE à l'intention des entreprises multinationales, ou encore les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme.

[6] Pour en savoir plus sur le sujet, voir les pages https://www.dudumimran.com/2018/05/speaking-about-ai-and-cyber-security-at-the-oecd-forum-2018.html et https://maliciousaireport.com/.

[7] Pierre Chalançon, Président du Groupe de réflexion du BIAC sur la protection des consommateurs et Vice-Président des affaires réglementaires, Vorwerk & Co KG, Représentation auprès de l'Union européenne – *Science-Fiction is not a Sound Basis for Legislation.*

[8] Cette technique est notamment utilisée pour entraîner les véhicules autonomes à effectuer des manœuvres complexes, entraîner le programme AlphaGo, ou encore traiter des patients atteints de cancers, en déterminant la dose et la fréquence d'administration minimales efficaces sur la réduction des tumeurs cérébrales (Matheson, 2018[107]).

[9] Une récente étude révèle que dans de nombreux cas, les données synthétiques peuvent remplacer utilement les données réelles et aident les chercheurs à s'affranchir des contraintes liées au respect de la vie privée (Patki, Wedge et Veeramachaneni, 2016[106]). Les auteurs montrent en effet que dans 70 % des cas, les résultats générés à l'aide des données synthétiques ne sont pas foncièrement différents de ceux obtenus avec des données réelles.

[10] Des solutions alliant des mécanismes tels que le chiffrement homomorphe complet et des réseaux neuronaux ont été testées avec succès et utilisées à cet effet (Dowlin, 2016[61]).

[11] Voir https://www.wipo.int/about-ip/fr/artificial_intelligence/index.html et https://www.uspto.gov/about-us/events/artificial-intelligence-intellectual-property-policy-considerations.

[12] Voir https://www.ibm.com/watson/stories/creditmutuel/.

[13] À titre d'exemple, Alibaba n'emploie plus de travailleurs temporaires pour gérer les demandes des clients pendant les périodes de forte activité ou les promotions spéciales. Le jour où Alibaba a réalisé son pic de ventes en 2017, l'agent conversationnel a traité plus de 95 % des questions des clients et répondu à quelque 3.5 millions de consommateurs (Zeng, 2018[73]). À mesure que les agents conversationnels prennent en charge des fonctions de service à la clientèle, le rôle des conseillers humains évolue vers le traitement de questions plus complexes ou personnelles.

5. Politiques et initiatives dans le domaine de l'IA

Les politiques et initiatives liées à l'intelligence artificielle (IA) se multiplient, que ce soit au niveau des pouvoirs publics, des entreprises, des organismes techniques, de la société civile et des syndicats. Des initiatives intergouvernementales en matière d'IA sont aussi en cours de développement. Ce chapitre recense les politiques, initiatives et stratégies d'IA élaborées par diverses parties prenantes dans le monde entier, aux niveaux tant national qu'international. Il observe que, d'une manière générale, les initiatives adoptées par les gouvernements nationaux envisagent d'abord l'IA comme un moyen d'améliorer la productivité et la compétitivité et s'appuient sur des plans d'action visant à renforcer : i) les facteurs, tels que les capacités de recherche en IA ; ii) la demande ; iii) les industries amont et apparentées ; iv) la stratégie, la structure et la rivalité des entreprises ; et v) la gouvernance et la coordination au plan national. Les initiatives internationales incluent notamment la Recommandation du Conseil de l'OCDE sur l'intelligence artificielle*, qui constitue le premier instrument d'orientation intergouvernemental sur l'IA et énonce un certain nombre de principes et de priorités d'action pour une approche responsable à l'appui d'une IA digne de confiance.*

Intelligence artificielle et compétitivité économique : stratégies et plans d'action

L'intelligence artificielle (IA) apparaît comme une priorité croissante des institutions gouvernementales, aux niveaux tant national qu'international. Nombre des initiatives lancées à ce jour par des gouvernements nationaux envisagent le recours à l'IA comme un moyen d'améliorer la productivité et la compétitivité. Les priorités énoncées dans les plans d'action nationaux relatifs à l'IA peuvent être regroupées en cinq grandes catégories, qui recoupent en partie celles du cadre de compétitivité économique de Porter. Ces priorités concernent : i) les facteurs, tels que les capacités de recherche en IA, y compris les compétences ; ii) la demande ; iii) les industries amont et apparentées ; iv) la stratégie, la structure et la rivalité des entreprises ; et v) les questions de gouvernance et de coordination au plan national (Encadré 5.1). Par ailleurs, parmi les considérations politiques propres à l'IA, les pouvoirs publics portent une attention croissante à la transparence, au respect des droits de l'homme et à l'éthique.

Plusieurs pays et économies partenaires de l'OCDE, notamment l'Allemagne, le Canada, la République populaire de Chine (ci-après « la Chine »), les États-Unis, la France, l'Inde, le Royaume-Uni et la Suède, se sont dotés de stratégies d'IA dédiées. Certains pays, comme la Corée, le Danemark et le Japon, ont inscrit des mesures en matière d'IA dans des programmes de portée plus générale. De nombreux autres – dont l'Australie, l'Espagne, l'Estonie, la Finlande, Israël et l'Italie – travaillent actuellement à l'élaboration d'une stratégie. L'ensemble des stratégies gouvernementales visent à : accroître le nombre de chercheurs et de diplômés qualifiés en IA ; renforcer les capacités nationales de recherche en IA ; et faire en sorte que les résultats de la recherche en IA débouchent sur des applications dans les secteurs public et privé. Dans leur manière d'aborder les conséquences économiques, sociales, éthiques, politiques et juridiques des progrès de l'IA, ces initiatives nationales reflètent les différences de cultures, de systèmes juridiques, de taille de pays et de degré d'adoption de l'IA, bien que la mise en œuvre des politiques en soit encore à un stade précoce. Le présent chapitre examine par ailleurs l'évolution récente de la réglementation et des politiques en matière d'IA ; il s'abstient cependant d'analyser ou d'évaluer la réalisation des objectifs des initiatives nationales ou le succès des différentes approches adoptées à cet égard.

La question de l'IA a également été abordée au niveau d'instances internationales telles que le Groupe des sept (G7), le Groupe des vingt (G20), l'OCDE, l'Union européenne et les Nations Unies. La Commission européenne met l'accent sur le rôle de l'IA pour promouvoir l'efficience et la flexibilité, les échanges et la coopération, la productivité, la compétitivité et la croissance, ainsi que la qualité de vie des citoyens. Suite à la réunion des ministres des TIC du G7 qui s'était tenue au Japon en avril 2016, la réunion des ministres des TIC et de l'industrie du G7 organisée à Turin (Italie) en septembre 2017 a mis en avant une vision de l'IA « centrée sur l'humain ». À cette occasion, les ministres ont décidé d'encourager la collaboration internationale et le dialogue multipartite autour de l'IA, et de promouvoir une meilleure compréhension de la coopération dans ce domaine avec l'appui de l'OCDE. On notera également l'attention continue que le G20 porte à l'IA, en particulier avec la proposition du Japon de mettre l'intelligence artificielle à l'ordre du jour de sa présidence du G20 en 2019 (G20, 2018[1]).

Principes concernant l'utilisation de l'intelligence artificielle dans la société

Plusieurs groupes de parties prenantes réfléchissent activement aux moyens de guider le développement et le déploiement de l'IA, afin que celle-ci bénéficie à l'ensemble de la société. En avril 2016, par exemple, l'Institute of Electrical and Electronics Engineers (IEEE)

a lancé une Initiative mondiale sur l'éthique dans la conception des systèmes autonomes et intelligents. Il a également publié en décembre 2017 la deuxième version de ses Principes d'intégration de l'éthique dès la conception (*Ethically Aligned Design*). La version finale était prévue pour le début de l'année 2019. Le Partenariat pour l'intelligence artificielle au service des citoyens et de la société (*Partnership on Artificial Intelligence to Benefit People and Society*), lancé en septembre 2016 sur la base d'une série de principes généraux, a commencé à travailler à l'élaboration de principes portant sur des questions spécifiques comme celle de la sécurité. Les Principes d'Asilomar formulent une série d'orientations de recherche, de normes éthiques et de valeurs pour assurer un développement sûr et socialement bénéfique de l'IA à court et à long terme. L'Initiative sur l'intelligence artificielle (*AI Initiative*), qui regroupe des experts, des praticiens et des citoyens du monde entier, s'est fixée pour objectif l'élaboration d'une définition commune de notions comme celle d'explicabilité de l'IA.

Plusieurs initiatives ont abouti à l'élaboration de séries de principes utiles pour guider le développement de l'IA (Tableau 5.1). Nombre de ces principes s'adressent aux communautés techniques qui mènent des activités de recherche et développement (R-D) ayant trait aux systèmes d'IA. Bien qu'ayant été développés pour une grande part dans le cadre de processus multipartites, ils ciblent cinq grandes communautés d'acteurs : les experts techniques, le secteur privé, les administrations, les universités et les syndicats. La communauté des experts techniques comprend le Future of Life Institute, l'IEEE, l'association JSAI (Japanese Society for Artificial Intelligence), la conférence FATML (Fairness, Accountability and Transparency in Machine Learning) et l'ACM (Association for Computing Machinery). Pour le secteur privé, on citera par exemple le Partenariat sur l'IA, le Conseil de l'industrie des technologies de l'information et Satya Nadella, président-directeur général de Microsoft. Du côté des administrations, on mentionnera le ministère des Affaires intérieures et des Communications du Japon, la Commission mondiale d'éthique des connaissances scientifiques et des technologies, et le Conseil britannique de la recherche en ingénierie et en sciences physiques. Pour les universités, on évoquera l'Université de Montréal et Nicolas Economou, président-directeur général de H5 et conseiller spécial pour l'AI Initiative lancée par l'incubateur The Future Society, à la Harvard Kennedy School. Enfin, du côté des syndicats, la fédération syndicale internationale UNI Global Union a elle aussi élaboré des principes éthiques sur l'intelligence artificielle.

Des thèmes communs se dégagent de ces initiatives. De fait, les différentes parties prenantes ont élaboré des lignes directrices sur le respect des valeurs humaines et des droits de l'homme, la non-discrimination, l'information et le contrôle, l'accès aux données, la protection de la vie privée et le contrôle des informations à caractère personnel, la sûreté et la sécurité, les compétences, la transparence et l'explicabilité, la responsabilité et la redevabilité, le dialogue à l'échelle de l'ensemble de la société, et la mesure de l'IA.

En mai 2018, le Comité de la politique de l'économie numérique de l'OCDE a créé le Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO) chargé d'élaborer des principes à mettre en œuvre dans le cadre des politiques publiques et de la coopération internationale et ayant vocation à promouvoir la confiance dans l'IA et son adoption (OCDE, 2019[2]). Les travaux du groupe d'experts ont nourri l'élaboration de la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle* (OCDE, 2019[3]), à laquelle 42 pays ont adhéré le 22 mai 2019.

Tableau 5.1. Liste non exhaustive de lignes directrices, principes ou déclarations sur l'IA émanant de diverses parties prenantes

Désignation	Principes et lignes directrices sur l'IA élaborés par diverses parties prenantes
ACM	ACM (2017), « 2018 ACM Code of Ethics and Professional Conduct: Draft 3 », Association for Computing Machinery Committee on Professional Ethics, https://ethics.acm.org/2018-code-draft-3/ USACM (2017), « Statement on Algorithmic Transparency and Accountability », Association for Computing Machinery US Public Policy Council, www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
AI safety	Amodei D. et al. (2016), « Concrete Problems in AI Safety », 25 juillet, https://arxiv.org/pdf/1606.06565.pdf
Asilomar	FLI (2017), « Asilomar AI Principles », Future of Life Institute, https://futureoflife.org/ai-principles/
COMEST	COMEST (2017), « Rapport de la COMEST sur l'éthique de la robotique », Commission mondiale d'éthique des connaissances scientifiques et des technologies, https://unesdoc.unesco.org/ark:/48223/pf0000253952_fre
Economou	Economou N. (2017), « A "Principled" Artificial Intelligence Could Improve Justice », 3 octobre, *Aba Journal*, www.abajournal.com/legalrebels/article/a_principled_artificial_intelligence_could_improve_justice
EPSRC	EPSRC (2010), « Principles of Robotics » , Engineering and Physical Sciences Research Council (Royaume-Uni), https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/
FATML	FATML (2016), « Principles for Accountable Algorithms and a Social Impact Statement for Algorithms », Fairness, Accountability and Transparency in Machine Learning, www.fatml.org/resources/principles-for-accountable-algorithms
FPF	FPF (2018), « Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models », The Future of Privacy Forum, https://fpf.org/wp-content/uploads/2018/06/Beyond-Explainability.pdf
GEE	GEE (2018), « Déclaration sur l'intelligence artificielle, la robotique et les systèmes "autonomes" », Groupe européen d'éthique des sciences et des nouvelles technologies, https://publications.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-fr/format-PDF
Google	Google (2018), « AI at Google: Our Principles », https://www.blog.google/technology/ai/ai-principles/
HLEG AI (CE)	Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle (AI HLEG) de la CE (2019), « Lignes directrices en matière d'éthique pour une IA digne de confiance », https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60427
IEEE	IEEE (2017), « Ethically Aligned Design: Version 2 », IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Institute of Electrical and Electronics Engineers, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf
Intel	Intel (2017), « AI - The Public Policy Opportunity », https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf
ITI	ITI (2017), « AI Policy Principles », Information Technology Industry Council, www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf
JSAI	JSAI (2017), « The Japanese Society for Artificial Intelligence Ethical Guidelines », The Japanese Society for Artificial Intelligence, http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf
MIC	MIC (2017), « Draft AI R&D Guidelines for International Discussions », Ministère des Affaires intérieures et des Communications du Japon, www.soumu.go.jp/main_content/000507517.pdf
MIC	MIC (2018), « Draft AI Utilization Principles », Ministère des Affaires intérieures et des Communications du Japon, www.soumu.go.jp/main_content/000581310.pdf
Montréal	UdeM (2017), « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, https://www.declarationmontreal-iaresponsable.com/
Nadella	Nadella S. (2016), « The Partnership of the Future », 28 juin, *Slate*, www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_and_a_i_can_work_together_to_solve_society.html
PAI	PAI (2016), « TENETS », Partenariat sur l'IA, www.partnershiponai.org/tenets/
Polonski	Polonski V. (2018), « The Hard Problem of AI Ethics: Three Guidelines for Building Morality into Machines », 28 février, Forum Network on Digitalisation and Trust, www.oecd-forum.org/users/80891-dr-vyacheslav-polonski/posts/30743-the-hard-problem-of-ai-ethics-three-guidelines-for-building-morality-into-machines
Taddeo et Floridi	Taddeo M. et L. Floridi (2018), « How IA Can Be a Force for Good », *Science*, 24 août, vol. 61/6404, p. 751-752, http://science.sciencemag.org/content/361/6404/751
The Public Voice Coalition	UGAI (2018), « Universal Guidelines on Artificial Intelligence », The Public Voice Coalition, https://thepublicvoice.org/ai-universal-guidelines/
Déclaration de Tokyo	Next Generation Artificial Intelligence Research Center (2017), « The Tokyo Statement – Co-operation for Beneficial AI », www.ai.u-tokyo.ac.jp/tokyo-statement.html
Twomey	Twomey P. (2018), « Toward a G20 Framework for Artificial Intelligence in the Workplace », *CIGI Papers,* n° 178, Centre pour l'innovation dans la gouvernance internationale, www.cigionline.org/sites/default/files/documents/Paper%20No.178.pdf
UNI	UNI Global Union (2017), « Top 10 Principles for Ethical Artificial Intelligence », www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf

Initiatives nationales

Tour d'horizon des politiques nationales en matière d'IA

De nombreux pays ont annoncé l'adoption de stratégies et de politiques nationales en matière d'IA, qui visent généralement à leur assurer un rôle de leadership dans le domaine de l'intelligence artificielle. Ces initiatives ont conduit à l'établissement de cibles et d'objectifs qui requièrent une action concertée de l'ensemble des parties prenantes. Les gouvernements servent fréquemment de coordinateurs et de facilitateurs à cet égard. L'Encadré 5.1 décrit les éléments qui apparaissent souvent dans les politiques et les mesures adoptées par les pouvoirs publics en vue de promouvoir la compétitivité nationale dans le domaine de l'IA. En outre, certains pays ont créé une entité publique spécifiquement chargée de s'occuper des questions d'éthique liées à l'IA et aux données, ou confié cette responsabilité à une entité existante.

Encadré 5.1. Comment les pays s'y prennent-ils pour développer un avantage compétitif dans le domaine de l'intelligence artificielle ?

Porter a identifié quatre éléments déterminants dans l'obtention d'un avantage compétitif dans un secteur d'activité particulier : i) les facteurs ; ii) la demande ; iii) les industries amont et apparentées ; et iv) la stratégie, la structure et la rivalité des entreprises. Il reconnaît que les acteurs à l'origine d'un avantage concurrentiel dans un secteur d'activité donné sont les entreprises. Cependant, il souligne le rôle essentiel des pouvoirs publics en ce qu'ils soutiennent ces quatre éléments au niveau des processus de développement industriel national.

- **Facteurs** : ce premier élément dépend de la localisation géographique, de l'existence d'une main-d'œuvre qualifiée, du niveau d'instruction et des capacités de recherche. Pour renforcer leurs capacités de recherche en IA, les pays recourent à différents types de mesures, par exemple : i) l'établissement d'instituts de recherche en IA ; ii) la création de nouveaux diplômes universitaires de troisième cycle et de doctorat dans le domaine de l'IA, et l'adaptation des cursus existants de manière à y inclure un module d'IA, en particulier dans les disciplines scientifiques ; et iii) les mesures visant à attirer des talents nationaux et étrangers, notamment en augmentant le nombre de visas accordés à des spécialistes de l'IA.
- **Demande** : plusieurs pays ont identifié des secteurs stratégiques dans l'optique du développement de l'IA, notamment les transports, la santé et les services publics. Ils mettent alors en place des mesures pour stimuler la demande intérieure de services d'IA dans ces secteurs. Pour ce qui est des services publics, certains pays s'assurent, par le biais des marchés publics, que les systèmes d'IA utilisés respectent des normes particulières, notamment en termes de précision ou de robustesse.
- **Industries amont et apparentées** : pour être concurrentiels, les systèmes d'IA doivent avoir accès aux infrastructures et services numériques, aux données, à des capacités informatiques et au haut débit. Un certain nombre de pays travaillent par conséquent au développement de pôles technologiques autour de l'IA et de structures de soutien pour les petites et moyennes entreprises (PME).
- **Stratégie, structure et rivalité des entreprises** : les pays recourent à diverses approches pour stimuler l'investissement privé et la concurrence dans le domaine de l'IA ; ils s'attachent notamment à : i) établir des feuilles de route pour le développement de l'IA, en vue de promouvoir l'investissement privé ; ii) encourager les entreprises internationales spécialisées dans l'IA à investir sur le territoire national, par exemple

via la création de laboratoires de recherche en IA ; et iii) tester des mesures expérimentales comme les « bacs à sable réglementaires » pour les applications de l'IA, afin d'inciter les entreprises à innover.

En outre, pour assurer une mise en œuvre effective de leurs initiatives nationales dans le domaine de l'IA, de nombreux pays réfléchissent à des mécanismes de gouvernance adéquats, qui garantiraient une approche coordonnée au niveau de l'ensemble de l'administration. La France, par exemple, a mis en place une fonction de coordination de l'IA au sein du cabinet du Premier ministre pour veiller à la mise en œuvre de la stratégie nationale en matière d'IA.

Source : Porter (1990[4]), « The competitive advantage of nations », https://hbr.org/1990/03/the-competitive-advantage-of-nations.

Allemagne

Le gouvernement fédéral allemand a lancé sa stratégie pour l'IA en décembre 2018 (Allemagne, 2018[5]). L'Allemagne entend devenir un centre majeur de l'IA en favorisant une transposition rapide et systématique des résultats de la recherche en applications, et en faisant de l'« IA made in Germany » une marque d'exportation forte et un gage de qualité reconnu au niveau mondial. Les mesures envisagées pour atteindre cet objectif comprennent : la création de nouveaux centres de recherche, le développement de la coopération franco-allemande dans le domaine de la recherche, le financement de pôles d'activité, et le soutien en faveur des PME. La stratégie nationale aborde également les besoins infrastructurels, l'amélioration de l'accès aux données, le développement des compétences, les normes de sécurité nécessaires pour prévenir l'utilisation abusive de l'IA et les questions éthiques.

En juin 2017, le ministère fédéral des Transports et des Infrastructures numériques a publié des directives éthiques sur les véhicules autonomes. Ces directives, élaborées par la commission d'éthique du ministère, prévoient une série de quinze règles applicables aux décisions programmées qui sont intégrées aux voitures sans conducteur. La commission a examiné de manière approfondie les questions éthiques en jeu, y compris celle des vies à préserver en priorité (le « dilemme du tramway »). Les directives stipulent ainsi que les véhicules autonomes doivent être programmés de manière à considérer toutes les vies humaines sur un pied d'égalité. Lorsqu'un choix s'impose, la décision du véhicule autonome doit porter sur le choc le moins violent, indépendamment de l'âge, de l'origine raciale ou du genre de la personne heurtée. En outre, la commission a clairement énoncé qu'un individu ne doit en aucun cas être contraint de se sacrifier pour sauver d'autres vies (Allemagne, 2017[6]).

Argentine

Le gouvernement argentin prévoit de lancer en juillet 2019 une Stratégie nationale en matière d'IA, qui s'étendra sur une période de dix ans. Cette décision fait suite à une phase d'évaluation menée en 2018 par le ministère de la Science, de la Technologie et de l'Innovation productive dans le cadre de deux initiatives nationales, baptisées *Agenda Digital Argentina 2030* et *Plan Argentina Innovadora 2030*. La Stratégie nationale pour l'IA porte notamment sur les domaines prioritaires suivants : l'éducation et le développement des compétences, les données, la R-D et l'innovation, l'infrastructure de superinformatique, les mesures visant à faciliter la mobilité professionnelle et le soutien de la coopération public-privé en matière d'utilisation des données. Les services publics et le secteur manufacturier sont également considérés comme des secteurs cibles prioritaires pour le développement de l'IA. La stratégie est organisée autour de plusieurs thèmes transversaux : i) l'investissement, l'éthique et la réglementation ; ii) la communication et la sensibilisation ; et iii) la coopération internationale.

La stratégie, qui fait intervenir sept ministères, prévoit la création d'une plateforme nationale d'innovation en matière d'IA pour la mise en œuvre de projets dans chacun des domaines thématiques. Dans chaque domaine prioritaire, un groupe de pilotage composé d'experts sera chargé de fixer des objectifs et de définir des indicateurs afin de mesurer les progrès accomplis.

Arabie saoudite

L'Arabie saoudite a annoncé en 2016 un plan de réforme économique à l'horizon 2030 (*Vision 2030*). Ce plan vise à stimuler le développement de nouveaux secteurs d'activité et la diversification de l'économie, faciliter l'émergence de modèles économiques mixtes (public-privé) et réduire la dépendance du pays à l'égard des revenus pétroliers. Il envisage la transformation numérique comme un levier essentiel de développement de l'économie reposant à la fois sur l'exploitation des données, l'IA et l'automatisation industrielle. Les secteurs prioritaires, notamment dans la perspective de la création de centres d'innovation, sont : la santé, les services publics, l'énergie durable et l'eau, le secteur manufacturier, ainsi que celui de la mobilité et des transports. Le gouvernement prépare une stratégie nationale en matière d'IA ayant pour but de construire un écosystème innovant et éthique de l'IA en Arabie saoudite d'ici à 2030.

L'Arabie saoudite s'efforce de mettre en place des conditions propices au développement de l'IA, en misant en particulier sur le déploiement du très haut débit et des réseaux 5G, l'accès aux données et la sécurité. Le pays encourage également l'adoption précoce des concepts et méthodes d'IA par le biais de plusieurs projets de villes intelligentes devant servir de catalyseurs à la mise en place de solutions nouvelles. Ces initiatives s'inscrivent dans le sillage du mégaprojet de ville intelligente NEOM, lancé en 2017 avec un investissement important de 1 800 milliards SAR (500 milliards USD). L'Arabie saoudite participe aussi activement aux débats internationaux sur les cadres de gouvernance de l'IA.

Australie

Le gouvernement australien a alloué, au titre du budget 2018-19, plus de 28 millions AUD (21 millions USD) au développement des capacités d'IA et à l'aide au développement responsable de l'IA en Australie. Ces fonds permettront de financer : i) les projets du Centre de recherche coopérative (Co-operative Research Centre) portant spécifiquement sur l'IA (18 millions USD) ; ii) des bourses doctorales dans le domaine de l'IA (1 million USD) ; iii) la création de ressources en ligne pour l'enseignement de l'IA dans les établissements scolaires (1.1 million USD) ; iv) l'établissement d'une feuille de route sur les technologies liées à l'IA en vue d'examiner les incidences de l'IA sur les différents secteurs, les opportunités et les défis pour la main-d'œuvre, et les implications en termes d'éducation et de formation (250 000 USD) ; v) l'élaboration d'un cadre éthique à partir d'études de cas (367 000 USD) ; et vi) l'établissement d'une feuille de route sur les normes dans le domaine de l'intelligence artificielle (72 000 USD, avec un apport équivalent de l'industrie).

Le ministère de l'Industrie, de l'Innovation et de la Science mène également des projets ayant trait à l'IA. Il a en particulier chargé le Conseil australien des sociétés savantes (Australian Council of Learned Academies) d'examiner les opportunités, les risques et les conséquences que pourrait entraîner l'introduction à grande échelle de l'IA en Australie pendant la prochaine décennie. La Commission australienne des droits de l'homme (Australian Human Rights Commission) a pour sa part lancé en juillet 2018 un projet de grande ampleur pour étudier la relation entre droits de l'homme et technologies. La publication d'un document de réflexion et l'organisation d'une conférence internationale sont notamment prévues dans ce contexte ; le rapport final est prévu pour 2019-20.

Brésil

La stratégie de transformation numérique du Brésil *E-Digital*, lancée en mars 2018, a pour but d'harmoniser et de coordonner les différentes initiatives gouvernementales introduites dans le domaine du numérique en vue de progresser dans la réalisation des Objectifs de développement durable au Brésil. Pour ce qui concerne spécifiquement l'IA, elle prévoit des « actions stratégiques » visant à « évaluer l'impact économique et social potentiel de (…) l'intelligence artificielle et des données massives, et proposer des mesures pour en limiter les effets négatifs et optimiser les résultats positifs » (Brésil, 2018[7]). La stratégie *E-Digital* donne par ailleurs la priorité à l'allocation de ressources à la recherche, au développement et à l'innovation dans le domaine de l'intelligence artificielle, ainsi qu'au renforcement des capacités d'IA. Le Brésil prévoit de lancer en 2019 une stratégie dédiée à l'IA. Il participe activement aux débats internationaux sur les normes techniques et les politiques d'IA.

Entre 2014 et début 2019, le ministère de la Science, de la Technologie, de l'Innovation et des Communications a soutenu au moyen de mesures d'incitation et d'aides financières 16 projets d'IA et 59 startups spécialisées dans l'intelligence artificielle. En outre, 39 initiatives visent à utiliser l'IA pour les fonctions d'administration électronique au niveau fédéral. L'objectif est notamment d'améliorer les procédures administratives et d'évaluation, en particulier dans le domaine de l'état-civil, des services sociaux et de la publication des offres d'emploi. Un nouvel institut de recherche en IA – *Instituto Avançado para Inteligência Artificial* – a été créé en 2019. Il a pour mission de promouvoir les partenariats entre universités et entreprises autour de projets conjoints de R-D et d'innovation en matière d'IA. Il s'occupera spécifiquement de domaines comme l'agriculture, les villes intelligentes, la gouvernance du numérique, l'infrastructure, l'environnement, les ressources naturelles, et la sécurité et la défense.

Canada

Le Canada cherche à se positionner comme un pays leader dans le domaine de l'IA, avec notamment la Stratégie pancanadienne en matière d'intelligence artificielle, lancée en mars 2017 (CIFAR, 2017[8]). Cette stratégie, qui est placée sous la direction de l'Institut canadien de recherches avancées (CIFAR), un organisme à but non lucratif, a bénéficié d'un financement public de 125 millions CAD (100 millions USD). Ces fonds sont destinés à soutenir, sur une période de cinq ans, des projets visant à : développer le capital humain du Canada, appuyer la recherche en IA au Canada et faire en sorte que la recherche en IA aboutisse à des applications pour les secteurs public et privé. La Stratégie pancanadienne en matière d'intelligence artificielle sert quatre grands objectifs :

1. Accroître le nombre de chercheurs et diplômés qualifiés dans le domaine de l'intelligence artificielle au Canada.
2. Établir des centres d'excellence scientifique interconnectés dans les trois grands instituts canadiens spécialisés dans l'intelligence artificielle, situés à Edmonton (Alberta Machine Intelligence Institute, Amii), Montréal (Institut des Algorithmes d'Apprentissage de Montréal, MILA) et Toronto (Institut Vecteur pour l'intelligence artificielle).
3. Développer un programme mondial sur l'intelligence artificielle dans la société et mener la réflexion, à l'échelle internationale, sur les répercussions économiques, sociales, éthiques, politiques et juridiques des progrès de l'intelligence artificielle.
4. Soutenir une communauté de recherche nationale en intelligence artificielle.

Le gouvernement fédéral, via le Conseil national de recherches Canada (CNRC), prévoit d'investir en tout 50 millions CAD (40 millions USD) sur une période de sept ans dans la recherche sur l'application de l'IA dans un certain nombre de domaines programmatiques clés, notamment l'analytique des données, l'IA et le design, la cybersécurité, les langues autochtones du Canada, le soutien aux « supergrappes » fédérales et aux centres de collaboration avec les universités canadiennes, ainsi que les partenariats stratégiques avec des acteurs internationaux.

Outre ce financement fédéral, le gouvernement du Québec a alloué 100 millions CAD (80 millions USD) au secteur de l'IA à Montréal, et l'Ontario, 50 millions CAD (40 millions USD) à l'Institut Vecteur pour l'intelligence artificielle. En 2016, le Fonds d'excellence en recherche Apogée Canada a alloué 93.6 millions CAD (75 millions USD) à trois universités – l'Université de Montréal, Polytechnique Montréal et HEC Montréal – pour la recherche de pointe sur l'apprentissage profond. Facebook et d'autres entreprises privées dynamiques comme Element AI sont actives au Canada.

Le gouvernement du Québec envisage de mettre sur pied un observatoire mondial sur les impacts sociaux de l'IA et des technologies numériques (Fonds de recherche du Québec, 2018[9]). Un atelier a été organisé en mars 2018 pour commencer à réfléchir à la mission et à l'organisation de cet observatoire, à son mode de gouvernance et à son financement, ainsi qu'aux modalités de la coopération internationale et aux secteurs et domaines d'intérêt à prendre en compte. Le gouvernement du Québec a alloué une enveloppe de 5 millions CAD (3.7 millions USD) pour soutenir la mise en œuvre de cet observatoire.

Le Canada collabore également avec divers partenaires au niveau international pour mener à bien des initiatives dans le domaine de l'IA. En juillet 2018, par exemple, les gouvernements du Canada et de la France ont annoncé leur intention de travailler ensemble à la création d'un nouveau Groupe international d'experts en intelligence artificielle (G2IA). Ce groupe aura pour mission de soutenir et de guider l'adoption responsable d'une IA centrée sur l'humain et axée sur les droits de l'homme, l'inclusion, la diversité, l'innovation et la croissance économique.

Chine

En mai 2016, le gouvernement chinois a rendu public un plan national pour l'IA d'une durée de trois ans, établi conjointement par la Commission nationale pour le développement et la réforme, le ministère de la Science et de la Technologie, le ministère de l'Industrie et des Technologies de l'information et l'Administration du cyberespace de Chine. L'initiative *Internet Plus*, lancée en 2015 en tant que stratégie nationale pour stimuler la croissance économique au moyen des technologies innovantes liées à l'internet pendant les années 2016-18, englobe l'IA (Jing et Dai, 2018[10]). Cette initiative porte principalement sur : i) le renforcement des capacités matérielles d'IA ; ii) la consolidation des écosystèmes de plateformes de réseau ; iii) les applications de l'IA dans les domaines socio-économiques importants ; et iv) l'impact de l'IA sur la société. Le gouvernement chinois prévoyait dans ce cadre la création, d'ici à 2018, d'un marché de 15 milliards USD par le biais de la R-D pour le secteur de l'IA en Chine (Chine, 2016[11]).

Mi-2017, le Conseil des affaires d'État de la Chine a publié des Lignes directrices relatives au Plan de développement de l'IA de nouvelle génération, qui définissent les perspectives à long terme de la Chine en matière d'IA, avec des objectifs industriels échelonnés en trois temps, selon les modalités suivantes : i) parvenir, d'ici à 2020, à une croissance économique tirée par l'IA ; ii) réaliser, d'ici à 2025, des progrès majeurs au niveau des théories fondamentales, ainsi que dans le développement d'une société intelligente ; et iii) faire du

pays, à l'horizon 2030, un centre d'innovation mondial en IA et bâtir un secteur de l'IA de 1 000 milliards CNY (150 milliards USD) (Chine, 2017[12]). La mise en œuvre du plan semble progresser dans l'ensemble des institutions gouvernementales et la Chine a acquis un certain leadership dans le domaine de l'IA grâce au soutien de l'État et au dynamisme des entreprises privées. Les objectifs fixés par le Conseil des affaires d'État prévoient que les « technologies d'information de nouvelle génération » devront représenter, en tant que secteur stratégique, 15 % du produit intérieur brut d'ici à 2020.

Le 13e plan quinquennal (2016-20) affirme l'ambition de la Chine de devenir un leader dans le domaine de la science et de la technologie et inclut seize « mégaprojets d'innovation scientifique et technologique à l'horizon 2030 ». L'un d'eux, dénommé *IA 2.0*, a donné un élan à l'action dans le secteur public (Kania, 2018[13]). Il incite également les entreprises à intensifier leurs activités de R-D concernant le matériel et les logiciels d'IA, en particulier dans les domaines de la reconnaissance visuelle, vocale et biométrique, des interfaces homme-machine et des systèmes de contrôle intelligents.

Le 18 janvier 2018, la Chine a créé un groupe national pour la normalisation de l'IA et un groupe consultatif national d'experts en IA. Au même moment, le comité national de normalisation auprès du ministère de l'Industrie a publié un livre blanc sur la normalisation dans le domaine de l'intelligence artificielle. Ce document a été élaboré avec le soutien de l'Institut national de normalisation électronique (une division du ministère de l'Industrie et des Technologies de l'information) (Chine, 2018[14]).

Les entreprises privées chinoises avaient commencé à manifester un intérêt pour l'IA avant les mesures d'aide adoptées il y a peu au niveau gouvernemental. Des entreprises chinoises comme Baidu, Alibaba et Tencent ont engagé des efforts et des investissements importants dans le domaine de l'IA. L'industrie chinoise s'intéresse tout particulièrement aux applications et à l'intégration des données, tandis que le gouvernement central focalise son attention sur les algorithmes de base, les données ouvertes et le travail conceptuel. Les municipalités s'intéressent quant à elles à l'utilisation qui peut être faite des applications et des données ouvertes dans le contexte urbain.

Corée

Le gouvernement coréen a publié en mars 2016 une Stratégie pour le développement du secteur des systèmes d'information intelligents. Il a annoncé des investissements publics d'une enveloppe de 1 000 milliards KRW (940 millions USD) d'ici à 2020 dans le domaine de l'IA et des technologies de l'information connexes comme l'internet des objets et l'infonuagique. Le but de la stratégie est de créer un nouvel écosystème à l'échelle du secteur des systèmes d'information intelligents et d'encourager l'investissement privé à hauteur de 2 500 milliards KRW (2.3 milliards USD) d'ici à 2020. La stratégie du gouvernement coréen sert trois objectifs : premièrement, le lancement de projets phares pour le développement de l'IA, consacrés par exemple aux technologies intelligentes dans le domaine langagier, visuel, spatial ou émotionnel ; deuxièmement, le renforcement des compétences de la main-d'œuvre liées à l'IA ; troisièmement, la promotion de l'accès aux données et de leur utilisation par les pouvoirs publics, les entreprises et les instituts de recherche (Corée, 2016[15]).

En décembre 2016, le gouvernement coréen a publié un « Plan d'action national pour une société de l'intelligence et de l'information à moyen et long terme ». Ce plan décrit les politiques nationales à mettre en place pour se préparer aux changements et aux enjeux induits par la quatrième révolution industrielle. Il jette aussi les bases du développement de technologies informatiques intelligentes de niveau mondial dans la perspective d'une « société intelligente centrée sur l'humain ». Ces technologies pourraient être introduites

dans tous les secteurs d'activité et être utilisées pour moderniser les politiques sociales. Pour mettre en œuvre ce plan d'action, le gouvernement s'appuie sur des bancs d'essai à grande échelle qui aideront au développement de nouveaux produits et services et, en particulier, à l'amélioration des services publics (Corée, 2016[16]).

En mai 2018, le gouvernement coréen a lancé un plan national d'investissement de 2 200 milliards KRW (2 milliards USD) d'ici à 2022 en vue de renforcer les capacités de R-D dans le domaine de l'intelligence artificielle. Ce plan prévoit la création de six instituts de recherche en IA, le développement des talents au moyen de 4 500 bourses en IA et de cours de formation intensive à court terme, et l'accélération du développement des puces d'IA (Peng, 2018[17]).

Danemark

Le Danemark a publié en janvier 2018 une Stratégie nationale pour la croissance numérique. Cette stratégie vise à faire du pays un chef de file du numérique, en permettant à l'ensemble des Danois de tirer parti de la transformation. Elle prévoit un certain nombre d'initiatives pour aider à saisir les opportunités qu'offrent l'IA, les données massives et l'internet des objets (IdO) en termes de croissance. Les principales orientations stratégiques définies dans ce cadre sont les suivantes : i) créer une plateforme numérique pour les partenariats public-privé ; ii) aider les PME à avancer sur la voie de la numérisation et à développer des activités commerciales reposant sur l'exploitation des données ; iii) établir des établissements d'enseignement dans le cadre d'un « pacte technologique » visant à favoriser le développement des compétences techniques et numériques ; iv) renforcer la cybersécurité dans les entreprises ; et v) mettre en place une réglementation souple afin de faciliter l'expérimentation et le développement de nouveaux modèles économiques. Le gouvernement danois s'est engagé à consacrer 1 milliard DKK (160 millions USD) jusqu'en 2025 à la mise en œuvre de cette stratégie. Ces fonds ont été répartis comme suit : 75 millions DKK (12 millions USD) en 2018 et 125 millions DKK (20 millions USD) pendant la période 2019-25. La plus grande part de ce budget sera affectée à des initiatives de développement des compétences, puis à la création de la plateforme numérique et à l'aide aux PME (Danemark, 2018[18]).

Estonie

L'Estonie prépare une nouvelle étape de son système d'administration électronique, qui reposera sur l'intelligence artificielle, afin de réduire les coûts et d'améliorer l'efficience. Elle teste également des services de santé en ligne et de conscience situationnelle. Le but est d'améliorer le quotidien des citoyens et la vie urbaine, et de promouvoir les valeurs humaines. Sur le plan du contrôle de la mise en œuvre, l'Estonie privilégie des valeurs fondamentales telles que l'éthique, la responsabilité, l'intégrité et la redevabilité, plutôt que de se concentrer sur des technologies en rapide évolution, et la mise en place d'un système de contrôle fondé sur la technologie des « chaînes de blocs » afin d'atténuer les risques en matière d'intégrité et de redevabilité. Un projet pilote est prévu en 2018.

Des véhicules autonomes sont testés sur les axes routiers publics de l'Estonie depuis mars 2017 dans le cadre du projet *StreetLEGAL*. L'Estonie est aussi le premier pays à réfléchir à la possibilité de conférer un statut juridique à certains systèmes reposant sur l'intelligence artificielle, par exemple en accordant à des algorithmes des droits de représentation – et des responsabilités – pour l'achat et la vente de services au nom de leurs propriétaires. En 2016, le gouvernement estonien a créé un groupe de travail pour examiner les problèmes de redevabilité associés aux algorithmes d'apprentissage automatique et les besoins de réglementation de l'IA, en collaboration avec le ministère des Affaires économiques et des Communications et le Bureau du gouvernement (Kaevats, 2017[19] ; Kaevats, 25 septembre 2017[20]).

États-Unis

Le président Trump a signé le 11 février 2019 le décret présidentiel n° 13859 sur « le maintien du leadership des États-Unis dans le domaine de l'intelligence artificielle », qui institue une nouvelle stratégie fédérale, l'*American AI Initiative*. Cette stratégie est organisée autour de cinq axes clés : i) l'investissement dans la R-D en IA ; ii) l'élargissement de l'accès aux ressources informatiques aux fins de la R-D en IA ; iii) la définition d'orientations en vue de la réglementation et de l'établissement de normes techniques pour les applications de l'IA : iv) le développement d'une main-d'œuvre qualifiée en IA ; et v) l'action internationale à l'appui de la recherche et de l'innovation américaines en IA et de l'ouverture de marchés aux entreprises américaines d'IA.

La nouvelle stratégie vient couronner toute une série d'initiatives engagées par l'administration fédérale pour accroître le leadership des États-Unis dans le domaine de l'IA. La Maison Blanche a accueilli en mai 2018 le premier Sommet sur l'intelligence artificielle, qui a réuni des acteurs de l'industrie, des universitaires et des dirigeants gouvernementaux. Les participants à ce sommet ont débattu de l'importance de lever les obstacles à l'innovation en IA dans le pays et de promouvoir la collaboration en matière de R-D dans ce domaine avec les partenaires des États-Unis. Ils ont évoqué la nécessité de sensibiliser le public à l'émergence de l'IA, afin que celui-ci soit à même de mieux comprendre le fonctionnement de ces technologies et les avantages qu'elles offrent au quotidien. Le même mois, l'exécutif américain a publié sous le titre *Artificial intelligence for the American people* (EOP, 2018[21]) une fiche d'information qui énumère les mesures et politiques adoptées par l'administration actuelle dans le domaine de l'IA, notamment : l'augmentation du financement public de la R-D en IA ; la réforme réglementaire pour faciliter le développement et l'utilisation des drones et des véhicules autonomes ; la priorité donnée à l'enseignement des sciences, des technologies, de l'ingénierie et des mathématiques (STIM), avec un accent particulier sur l'informatique ; et l'amélioration de l'accès aux données fédérales aux fins de la recherche et des applications de l'IA.

Dans les budgets fédéraux de R-D pour les exercices 2019 et 2020, l'intelligence artificielle et l'apprentissage automatique sont définis comme des priorités essentielles. La recherche fondamentale en IA menée à la Fondation nationale des sciences et les activités de R-D appliquée poursuivies au ministère des Transports y sont spécifiquement incluses. Les priorités de recherche englobent également le développement de méthodes analytiques de pointe en matière de santé aux National Institutes of Health et d'une infrastructure informatique pour l'IA au ministère de l'Énergie. Globalement, les investissements du gouvernement fédéral dans la R-D non classifiée liée à l'IA et aux technologies connexes ont augmenté de plus de 40 % depuis 2015.

En septembre 2018, la commission spéciale sur l'intelligence artificielle du Conseil national de la science et de la technologie a commencé à actualiser le Plan stratégique national pour la recherche et le développement en matière d'intelligence artificielle. En effet, depuis la publication de ce plan en 2016, la technologie sous-jacente, les cas d'usage et le déploiement commercial de l'IA ont connu un développement rapide. La commission spéciale sollicite l'avis du public sur les améliorations à apporter au plan, en particulier celui des acteurs directement impliqués dans la réalisation d'activités de R-D en IA ou concernés par les résultats qui en découlent.

Le gouvernement américain accorde aussi la priorité à la formation de la future main-d'œuvre des États-Unis. Le président Trump a signé un décret présidentiel sur l'établissement de programmes d'apprentissage reconnus par les entreprises, qui prévoit la création d'un groupe de travail ministériel sur le développement de l'apprentissage (*Task Force on*

Apprenticeship Expansion). Dans le droit fil des politiques décrites dans la fiche d'information mentionnée plus haut, une note présidentielle affirme également la priorité d'assurer un enseignement de qualité dans les STIM, et tout particulièrement en informatique. Cette note annonce l'affectation de 200 millions USD sous forme de subventions, auxquels est venu s'ajouter un engagement du secteur privé à hauteur de 300 millions USD.

Le Congrès américain a créé en mai 2017 un groupe parlementaire mixte sur l'intelligence artificielle, qui est coprésidé par deux parlementaires, John K. Delaney et Pete Olson (US, 2017[22]). Ce groupe organise des réunions avec des experts des milieux universitaires, du gouvernement et du secteur privé pour débattre des incidences des technologies liées à l'IA. Le Congrès envisage d'adopter un texte de loi qui établira un comité consultatif fédéral sur l'IA et instituera des normes fédérales de sécurité pour les véhicules autonomes.

Fédération de Russie

Le gouvernement russe a lancé en juillet 2017 une stratégie numérique intitulée « L'économie numérique de la Fédération de Russie ». Cette stratégie donne la priorité à l'exploitation du développement de l'IA, notamment en instaurant des conditions juridiques propices pour faciliter les activités de R-D. Elle prévoit aussi l'adoption de mesures pour inciter les entreprises d'État à participer aux communautés de recherche existant à l'échelon national (centres de compétences). Elle promeut en outre l'établissement de normes nationales pour les technologies d'IA (Russie, 2017[23]). Avant la publication de cette stratégie numérique, le gouvernement russe avait investi dans divers projets d'IA et mis en place des dispositifs en vue de la création de partenariats public-privé. L'association russe des concepteurs de systèmes d'intelligence artificielle encourage la coopération entre universités et entreprises, afin de faciliter les transferts de technologie vers les entreprises.

Finlande

La Finlande entend développer une société sûre et démocratique s'appuyant sur l'IA, offrir les meilleurs services publics du monde, et faire de l'IA un levier de prospérité, de croissance et de productivité renouvelées pour ses citoyens. La stratégie nationale d'IA publiée en octobre 2017 sous le titre « L'ère de l'intelligence artificielle en Finlande » est une feuille de route décrivant comment mettre à profit à la fois le niveau d'instruction de la population, la transformation numérique déjà bien avancée du pays et les ressources que constituent les données du secteur public. Cette stratégie prévoit aussi le développement de liens internationaux dans le domaine de la recherche et de l'investissement, et l'adoption de mesures pour encourager l'investissement privé. La Finlande espère doubler, d'ici à 2035, la croissance économique nationale grâce à l'IA. Huit mesures clés ont pour but d'améliorer la croissance, la productivité et le bien-être en s'appuyant sur l'IA : i) accroître la compétitivité des entreprises ; ii) favoriser l'utilisation des données dans tous les secteurs ; iii) accélérer et simplifier l'adoption de l'IA ; iv) développer une expertise de haut niveau ; v) prendre des décisions audacieuses, notamment en matière d'investissement ; vi) développer les meilleurs services publics du monde ; vii) établir de nouveaux modèles de coopération ; et viii) faire de la Finlande un pays pionnier à l'ère de l'IA. La stratégie nationale met particulièrement l'accent sur l'utilisation de l'IA pour améliorer les services publics. Le service de l'immigration finlandais utilise par exemple le réseau robotique national de service à la clientèle dénommé Aurora pour fournir des services de communication multilingues (Finlande, 2017[24]).

Le gouvernement finlandais a également créé en février 2018 un organisme de financement de la recherche en intelligence artificielle et des projets commerciaux faisant appel à l'IA. Cet organisme sera chargé d'allouer 200 millions EUR (235 millions USD) au secteur

privé, dont les PME, sous forme d'incitations et de subventions. Selon les autorités nationales, environ 250 entreprises travaillent au développement de l'IA en Finlande. Dans le secteur de la santé, notamment, le déploiement de l'IA changera la donne pour les patients et les professionnels, ainsi que pour les organisations du secteur et le système de santé, et s'accompagnera de réformes approfondies (Sivonen, 2017[25]). Par ailleurs, il est prévu d'étendre le rôle du Centre national de recherche technique, financé par l'État, et de l'Agence finlandaise de financement de la technologie et de l'innovation.

France

Le président Emmanuel Macron a annoncé la stratégie française pour l'IA le 29 mars 2018. Cette stratégie prévoit d'allouer au secteur de l'IA un financement public de 1.5 milliard EUR d'ici à 2022 pour aider la France à devenir un leader de la recherche et de l'innovation dans ce domaine. Les mesures envisagées s'appuient en grande partie sur les recommandations formulées dans le rapport du député Cédric Villani (Villani, 2018[26]). La stratégie appelle à investir dans la recherche et l'enseignement publics, à développer des plateformes de recherche de niveau mondial liées à l'industrie au moyen de partenariats public-privés et d'augmenter l'attractivité de la France pour les chercheurs talentueux en IA expatriés ou étrangers. Pour consolider l'écosystème français de l'IA, la stratégie envisage de « moderniser » les secteurs d'activité existants. Il s'agira ainsi, en introduisant d'abord des applications dans les domaines de la santé, de l'environnement, des transports et de la défense, de renouveler grâce à l'IA l'ensemble des secteurs d'activité. La stratégie propose également d'améliorer en priorité l'accès aux données en créant des « communs de données » entre acteurs privés et publics, de réformer le droit d'auteur de manière à faciliter les pratiques d'exploration de données, et d'ouvrir l'accès aux données du secteur public, notamment les données de santé, à des partenaires privés.

La stratégie française décrit également les mesures initiales à prendre à l'égard des perturbations induites par l'intelligence artificielle, en prenant fermement position sur la question des transferts de données hors d'Europe (Thompson, 2018[27]). Elle envisage la création d'un pôle central des données, doté d'une équipe d'une trentaine d'experts qui rempliraient un rôle de conseil en IA auprès de l'ensemble des administrations publiques. Sur le plan éthique et philosophique, la stratégie met en avant un principe fondamental, celui de la transparence des algorithmes. Les algorithmes développés par des administrations publiques françaises ou financés sur des fonds publics, par exemple, devront être ouverts. Le respect de la vie privée et des droits de l'homme devra être intégré par défaut, dès la conception des systèmes d'IA. La stratégie envisage aussi le développement de la formation professionnelle dans les métiers menacés par l'IA. Elle appelle à expérimenter des mesures sur le marché du travail et à engager le dialogue sur les moyens de répartir la valeur ajoutée générée par l'IA sur l'ensemble de la chaîne de valeur. En outre, un rapport a été publié fin mars 2018, intitulé *Intelligence artificielle et travail* (Benhamou et Janin, 2018[28]).

Hongrie

La Hongrie a lancé en octobre 2018 une « alliance pour l'intelligence artificielle », sous la forme d'un partenariat entre administrations publiques, entreprises informatiques de pointe et universités. Cette alliance est chargée de définir une stratégie visant à faire de la Hongrie un pays innovant dans le domaine de l'IA et d'étudier les incidences économiques et sociales de l'IA sur la société. Elle regroupe environ 70 instituts de recherche universitaires, entreprises et organismes publics et est appelée à devenir un forum de coopération transversale dans le domaine de la R-D en IA. L'Université polytechnique et économique de Budapest et l'Université Loránd Eötvös font partie du consortium prévoyant d'investir 20 millions EUR

(23.5 millions USD) dans le projet AI4EU, qui vise à créer une plateforme à la demande pour l'intelligence artificielle en Europe.

Inde

L'Inde a rendu publique sa stratégie d'IA en juin 2018. Cette stratégie énonce un certain nombre de recommandations visant à faire de l'Inde un pays de pointe dans ce domaine grâce à la mobilisation des ressources humaines et à l'impulsion d'une croissance inclusive bénéficiant à l'ensemble de la société. Cette approche inclusive a été baptisée ♯AIFORALL (« L'IA POUR TOUS »). Les domaines suivants sont désignés comme prioritaires pour l'introduction d'applications reposant sur l'intelligence artificielle : la santé, l'agriculture, l'éducation, les villes intelligentes et les transports. La stratégie souligne que le faible niveau des capacités de recherche et l'absence d'écosystèmes de données en Inde constituent des obstacles à la réalisation du plein potentiel de l'IA. Elle recommande, en conséquence, la création en Inde d'instituts de recherche à deux niveaux (pour la recherche fondamentale et la recherche appliquée), la mise en place de pôles de formation pour la main-d'œuvre actuelle, et le développement d'ensembles de données ciblés et de pépinières de startups. Enfin, elle juge indispensable l'établissement d'un cadre réglementaire sur la protection des données et la cybersécurité (Inde, 2018[29]).

Italie

L'Italie a publié en mars 2018 un livre blanc sur « L'intelligence artificielle au service des citoyens », qui a été rédigé par un groupe de travail de l'agence italienne pour le numérique. Ce document examine les moyens pour l'administration publique de mettre les technologies d'IA au service des citoyens et des entreprises, et de l'amélioration de l'efficience des services publics et du niveau de satisfaction des usagers. Il recense les obstacles à surmonter pour introduire l'IA dans les services publics, que ce soit en termes d'éthique, de technologie, de disponibilité des données et d'évaluation d'impact. Le livre blanc formule également des recommandations en vue : de l'établissement d'une plateforme nationale pour les données étiquetées, les algorithmes et les modèles d'apprentissage ; du développement des compétences ; et de la création d'un centre national de compétences et d'un centre transdisciplinaire sur l'IA. Il appelle en outre à élaborer des lignes directrices et des processus pour parvenir à un niveau de contrôle plus élevé et faciliter l'échange de données entre pays européens sur les cyberattaques exploitant les systèmes d'intelligence artificielle ou les visant (Italie, 2018[30]).

Japon

Le Bureau du Premier ministre japonais a créé en avril 2016 un Conseil de la stratégie des technologies liées à l'IA, afin de promouvoir la R-D et les applications commerciales dans le domaine de l'intelligence artificielle. Ce conseil a publié en mars 2017 une Stratégie pour les technologies d'intelligence artificielle, qui recensait un certain nombre d'enjeux décisifs pour le Japon, en particulier la nécessité d'accroître l'investissement, de faciliter l'accès aux données et leur utilisation et d'augmenter le nombre de chercheurs et d'ingénieurs en IA. Ce document identifiait également les domaines stratégiques où l'IA pourrait avoir d'importantes retombées positives, à savoir : la productivité ; la santé, les soins médicaux et les services sociaux ; la mobilité ; et la sécurité de l'information (Japon, 2017[31]).

La Stratégie intégrée pour l'innovation du Japon, publiée par le Bureau du Premier ministre en juin 2018, est à l'origine d'un certain nombre d'initiatives publiques dans le domaine de l'IA (Japon, 2018[32]). Elle prévoit notamment l'organisation de discussions multipartites sur les questions éthiques, juridiques et sociétales liées à l'IA. Ces discussions ont abouti à

la publication par le Bureau du Premier ministre en avril 2019 d'un document intitulé « Principes sociaux pour une intelligence artificielle centrée sur l'humain » (Japon, 2019[33]).

Lors de la réunion des ministres des TIC du G7 à Takamatsu en avril 2016, le Japon a proposé d'élaborer des principes communs sur la R-D dans le domaine de l'intelligence artificielle. Un groupe d'experts réuni sous le nom de *Conference toward IA Network Society* a produit un « Projet de lignes directrices sur la R-D dans le domaine de l'IA en vue des discussions internationales », qui a été publié par le ministère japonais des Affaires intérieures et des Communications en juillet 2017. Ces lignes directrices ont essentiellement vocation à mettre en balance les avantages et les risques qui sont associés aux réseaux d'IA, tout en garantissant la neutralité technologique et en évitant d'imposer des contraintes excessives aux développeurs de ces réseaux. Elles définissent neuf principes dont doivent tenir compte les chercheurs et développeurs des systèmes d'IA (Japon, 2017[34]). Le Tableau 5.2 ci-dessous donne un aperçu succinct des dispositions de ces lignes directrices. À l'issue de ses discussions, le groupe d'experts a également publié un « Projet de principes sur l'utilisation de l'IA » en juillet 2018 (Japon, 2018[35]).

Tableau 5.2. Principes énoncés dans le projet de « Lignes directrices sur la R-D dans le domaine de l'IA en vue des discussions internationales »

Principes	Les développeurs devraient :
I. Collaboration	Veiller à l'interconnectivité et l'interopérabilité des systèmes d'IA.
II. Transparence	Garantir la vérifiabilité des entrées/sorties des systèmes d'IA et l'explicabilité de leurs décisions.
III. Contrôlabilité	Assurer la contrôlabilité des systèmes d'IA.
IV. Sûreté	Veiller, au moyen d'actionneurs ou d'autres dispositifs, à ce que les systèmes d'IA ne puissent porter atteinte à la vie, aux personnes ou aux biens des utilisateurs ou de tiers.
V. Sécurité	Garantir la sécurité des systèmes d'IA.
VI. Respect de la vie privée	Prendre des dispositions pour éviter que les systèmes d'IA ne violent la vie privée des utilisateurs ou de tiers.
VII. Éthique	Respecter la dignité humaine et l'autonomie individuelle dans la R-D sur les systèmes d'IA.
VIII. Assistance aux utilisateurs	Ne pas perdre de vue que les systèmes d'IA sont au service des utilisateurs et donner à ces derniers la possibilité de prendre des décisions appropriées.
IX. Responsabilité	Assumer leurs responsabilités à l'égard des différentes parties prenantes, y compris les utilisateurs des systèmes d'IA.

Source : Japon (2017[34]), *Draft IA R-D guidelines for international discussions*, www.soumu.go.jp/main_content/000507517.pdf.

Mexique

Le Conseil national de la science et de la technologie du Mexique a créé en 2004 un Centre de recherche sur l'intelligence artificielle, qui dirige le développement des systèmes intelligents.

Un livre blanc a été publié en juin 2018 sous le titre « Vers une stratégie de l'IA au Mexique : exploiter la révolution de l'IA »[1]. Ce rapport note que le Mexique occupe la 22e place parmi les 35 pays de l'OCDE dans le classement établi par Oxford Insight sur la base d'un indice de préparation à l'intelligence artificielle (*AI Readiness Index*). Cet indice composite est en fait la moyenne des notes obtenues par les pays au regard de neuf indicateurs allant du niveau des compétences numériques à celui de l'innovation dans le secteur public. Le Mexique obtient de bonnes notes pour ses politiques en matière de données ouvertes et son infrastructure numérique, mais des notes plus médiocres s'agissant des compétences techniques, de la progression du numérique et de l'innovation dans le secteur public. Le rapport formule à l'intention des pouvoirs publics plusieurs recommandations pour la poursuite du développement et du déploiement de l'IA au

Mexique. Ces recommandations couvrent les cinq domaines suivants : l'administration et les services publics ; la R-D ; les capacités, les compétences et l'enseignement ; les données et l'infrastructure numérique ; et l'éthique et la réglementation (Mexico, 2018[36]).

Norvège

La Norvège a pris un certain nombre de mesures en faveur de l'IA dans le cadre de son programme numérique national et du plan à long terme pour la recherche et l'enseignement supérieur, notamment :

- La création de plusieurs laboratoires d'intelligence artificielle, comme l'Open AI Lab de l'Université de science et de technologie. L'Open AI Lab, qui est parrainé par plusieurs entreprises, travaille dans les domaines de l'énergie, du transport maritime, de l'aquaculture, des télécommunications, des services bancaires en ligne, de la santé et de la biomédecine, où la Norvège occupe une forte position au niveau international.
- Le lancement d'un programme de réforme des compétences dénommé « Apprendre pour la vie », avec une proposition de budget de 130 millions NOK (16 millions USD) en 2019, afin d'aider la main-d'œuvre à acquérir ou améliorer ses compétences dans des domaines tels que l'intelligence artificielle ou les soins de santé.
- Le développement d'une stratégie d'ouverture des données en vertu de laquelle les administrations publiques sont tenues de permettre l'accès à leurs données dans un format exploitable par la machine via des API, et d'enregistrer dans un catalogue commun les ensembles de données disponibles.
- La mise en place d'une plateforme pour la formulation de lignes directrices et de principes éthiques régissant l'utilisation de l'IA.
- La réforme de la réglementation afin d'autoriser les tests de véhicules autonomes sur le réseau routier, y compris les tests de voitures sans conducteur.

Pays-Bas

Avec l'adoption en 2018 d'une Stratégie nationale de transformation numérique, le gouvernement des Pays-Bas a pris deux engagements : premièrement, celui d'optimiser les opportunités économiques et sociales qu'offre le numérique ; deuxièmement, celui de renforcer les conditions indispensables au développement du numérique comme le développement des compétences, la politique en matière de données, la confiance et la résilience, le respect des droits fondamentaux et l'éthique (en veillant par exemple à ce que les algorithmes n'influent pas sur l'autonomie des individus ni sur l'égalité de traitement), ainsi que la recherche et l'innovation en IA. En octobre 2018, l'AINED (« L'intelligence artificielle pour les Pays-Bas »), un partenariat regroupant des entreprises et des universitaires, a publié un document recensant les objectifs et les actions destinés à nourrir un plan national pour l'IA, à savoir par exemple promouvoir l'accès aux talents et compétences en IA, ainsi qu'aux données publiques à forte valeur ajoutée. L'AINED cherche également à faciliter le développement d'entreprises fondées sur l'IA et à promouvoir l'utilisation à grande échelle de l'IA dans les administrations publiques. Il envisage en outre l'établissement d'un cadre socio-économique et d'un cadre éthique pour l'IA, en encourageant la constitution de partenariats public-privé dans certains secteurs clés et au niveau de certaines chaînes de valeur essentielles, et en faisant des Pays-Bas un centre mondial de la recherche en IA. Le gouvernement néerlandais entend en outre finaliser un plan d'action stratégique pour l'IA à l'échelle de l'ensemble de l'administration avant mi-2019. Ce plan d'action doit tenir

compte du rapport de l'AINED, du plan coordonné de l'UE et des débats menés au sein du Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle (AI HLEG) de la Commission européenne.

République tchèque

Le gouvernement tchèque a appelé en 2018 à la réalisation d'une étude sur la mise en œuvre de l'IA, afin de définir des objectifs stratégiques et de soutenir les négociations aux niveaux européen et international. Une équipe d'universitaires du Centre technologique de l'Académie des sciences, de l'Université technique de Prague et de l'Institut d'administration publique et de droit de l'Académie des sciences a produit un rapport intitulé « Analyse du développement potentiel de l'intelligence artificielle en République tchèque » (OGCR, 2018[37]). Ce rapport examine : i) l'état actuel de la mise en œuvre de l'IA en République tchèque ; ii) l'impact potentiel de l'IA sur le marché du travail national ; et iii) les aspects éthiques, juridiques et réglementaires du développement de l'IA en République tchèque.

Royaume-Uni

La stratégie numérique du Royaume-Uni (*UK Digital Strategy*), publiée en mars 2017, reconnaît l'importance de l'IA pour la croissance de l'économie numérique britannique (RU, 2017[38]). Cette stratégie prévoit d'allouer un financement de 17.3 millions GBP (22.3 millions USD) aux universités britanniques pour le développement de l'IA et des technologies robotiques. Le gouvernement a décidé d'augmenter l'investissement dans la R-D en IA de 4.7 milliards GBP (6.6 milliards USD) sur les quatre prochaines années, en partie par l'intermédiaire de l'ISCF (Industrial Strategy Challenge Fund).

En octobre 2017, le gouvernement a publié un examen indépendant du secteur de l'IA au Royaume-Uni. Ce rapport présente le Royaume-Uni comme un centre international d'expertise en IA, en partie grâce aux travaux de pionniers de la recherche en informatique, comme Alan Turing. Le gouvernement britannique estime que l'apport de l'IA à l'économie nationale pourrait atteindre 579.7 milliards GBP (814 milliards USD). Parmi les outils d'IA déjà utilisés au Royaume-Uni, on peut citer le guide personnel de santé (Your.MD), un agent conversationnel (*chatbot*) pour les clients des services bancaires et une plateforme éducative destinée à aider les enseignants à proposer des programmes d'enseignement personnalisés et accompagner les enfants dans leur apprentissage. Le rapport énonce 18 recommandations portant notamment sur l'amélioration de l'accès aux données et le partage des données grâce à la création de fiducies données, ou encore l'amélioration de l'offre de compétences en IA en introduisant des masters en intelligence artificielle qui seraient parrainés par des entreprises. Les autres mesures décrites comme prioritaires visent notamment à : optimiser la recherche en IA via la coordination de la demande de capacités informatiques pour ce type de recherche entre les institutions concernées ; soutenir l'adoption de l'IA au moyen d'un conseil national de l'IA ; et élaborer un cadre pour renforcer la transparence des décisions prises par des systèmes d'IA et la redevabilité y afférente (Hall et Pesenti, 2017[39]).

Le gouvernement britannique a rendu publique sa stratégie industrielle en novembre 2017. Cette stratégie définit l'IA comme l'un des « quatre grands enjeux » pour faire du Royaume-Uni un acteur de premier plan des industries du futur et veiller à ce que le pays tire parti des changements majeurs en cours au niveau mondial (RU, 2017[40]). En avril 2018, le Royaume-Uni a annoncé le *AI Sector Deal*, un programme d'investissements de 950 millions GBP (1.2 milliard USD) qui vise, en s'appuyant sur les atouts actuels du pays, à maintenir un écosystème de classe mondiale dans le domaine de l'intelligence artificielle. Ce programme comporte trois volets principaux : les talents et les compétences ; la promotion de l'adoption de l'IA ; et les données et l'infrastructure (RU, 2018[41]).

Le gouvernement a créé un Bureau de l'intelligence artificielle (Office for Artificial Intelligence, ou OAI) chargé de mettre en œuvre le programme d'investissements décrit ci-dessus et de promouvoir plus généralement l'adoption de l'IA. Il a également mis sur pied un Centre pour l'éthique des données et l'innovation. Ce centre vise à renforcer le cadre de gouvernance pour favoriser l'innovation, tout en veillant à préserver la confiance du public. Il fournira au gouvernement une expertise et des avis indépendants sur les mesures nécessaires pour permettre des innovations majeures, respectueuses de l'éthique et sûres dans le domaine des technologies fondées sur les données et l'intelligence artificielle. Il prévoit pour ce faire de lancer une fiducie de données pilote d'ici à fin 2019. Un conseil de l'IA s'appuyant sur l'expertise des entreprises travaille en collaboration étroite avec l'OAI (RU, 2018[42]).

Singapour

L'Autorité nationale des communications et des médias (Infocomm Media Development Authority, ou IMDA) a publié en mai 2018 un « Cadre d'action pour l'économie numérique ». Ce document trace les grandes lignes de l'action à mener pour faire de Singapour une économie numérique de premier ordre et aborde l'intelligence artificielle comme une technologie de pointe à même de tirer sa transformation numérique (Singapour, 2018[43]).

La Commission de protection des données à caractère personnel a publié en janvier 2019 un cadre de gouvernance de l'IA pour promouvoir l'adoption responsable de l'IA à Singapour. Ce cadre-type fournit des orientations concrètes pour mettre en pratique les principes éthiques. Il s'appuie sur un document de réflexion, ainsi que sur les discussions engagées au niveau national par la Table ronde des régulateurs, une communauté de pratique regroupant des représentants des instances de réglementation sectorielles et des organismes publics. Le cadre-type, que les entités ou organisations peuvent adopter sur une base volontaire, sert également de point de départ à l'élaboration de cadres sectoriels sur la gouvernance de l'IA.

Un Conseil consultatif sur l'utilisation éthique de l'IA et des données a été créé en juin 2018. Cet organe multipartite conseille le gouvernement de Singapour sur les questions éthiques, juridiques, réglementaires et politiques soulevées par le déploiement commercial de l'IA. Tout en soutenant l'adoption de l'IA dans l'industrie, il veille au respect des principes de responsabilité et de redevabilité dans la mise à disposition des produits et des services reposant sur l'IA.

Un programme de recherche sur la gouvernance de l'intelligence artificielle et l'utilisation des données, d'une durée de cinq ans, a été lancé en septembre 2018 en vue de faire de Singapour un centre de connaissances de premier rang, doté d'une expertise internationale dans le domaine des politiques et réglementations de l'IA. Ce programme est basé dans le Centre pour la gouvernance de l'IA et des données de la Faculté de droit de l'Université de management de Singapour, qui mène des travaux de recherche sur l'intelligence artificielle dans ses relations avec l'industrie, la société et les entreprises.

Suède

Le gouvernement suédois a publié en mai 2018 un rapport intitulé « L'intelligence artificielle dans les entreprises et la société suédoises ». Ce rapport, qui a pour but de stimuler la recherche et l'innovation en IA en Suède, définit six domaines stratégiques prioritaires : i) le développement industriel, notamment dans le secteur manufacturier ; ii) le secteur des voyages et des transports ; iii) les villes intelligentes et durables ; iv) les soins de santé ; v) les services financiers ; et vi) la sécurité, en y incluant la police et les douanes. Il souligne la nécessité d'atteindre une masse critique dans la recherche, l'éducation et l'innovation. Il appelle également à une plus grande coopération, notamment en vue de l'investissement

dans la recherche et l'infrastructure, l'éducation et la formation, la réglementation et la mobilité de la main-d'œuvre (Vinnova, 2018[44]).

Turquie

Le Conseil de la recherche scientifique et technologique, principal organisme de gestion et de financement de la recherche en Turquie, a financé de nombreux projets de R-D dans le domaine de l'intelligence artificielle. Il prévoit de lancer un appel multilatéral à projets de recherche en IA via le réseau intergouvernemental pour l'innovation EUREKA. Le ministère des Sciences et de la Technologie a établi une feuille de route nationale sur le numérique, dans le cadre de la plateforme pour la transformation numérique de l'industrie turque. Cette feuille de route porte en partie sur les progrès des technologies numériques émergentes comme l'IA.

Initiatives intergouvernementales

G7 et G20

Lors de la réunion des ministres des TIC du G7 à Takamatsu (Japon) en avril 2016, le ministre japonais des Affaires intérieures et des Communications a présenté, pour examen, une série de principes sur la R-D dans le domaine de l'intelligence artificielle (G7, 2016[45]).

La réunion des ministres des TIC et de l'industrie du G7, qui s'est tenue à Turin en septembre 2017 sous la présidence italienne, a publié une déclaration ministérielle dans laquelle les pays du G7 ont reconnu les avantages potentiels considérables qu'offre l'IA pour la société et l'économie et sont convenus de la nécessité d'une approche centrée sur l'humain en matière d'IA (G7, 2017[46]).

Les ministres de l'innovation du G7 réunis à Montréal en mars 2018 dans le cadre de la présidence canadienne ont exprimé une vision de l'IA centrée sur l'humain et mis l'accent sur l'interdépendance entre la croissance économique suscitée par l'innovation en IA, l'augmentation de la confiance envers l'IA et de l'adoption de l'AI, et la promotion de l'inclusivité dans le développement et le déploiement de l'IA. Les membres du G7 sont convenus d'agir dans un certain nombre de domaines pertinents, et notamment de s'attacher à :

- Investir dans la R-D fondamentale et la R-D appliquée précoce en vue de produire des innovations en IA, et soutenir l'entrepreneuriat en IA et la préparation de la population active à l'automatisation.
- Continuer d'encourager la recherche, y compris pour relever les défis sociétaux, stimuler la croissance économique et examiner les aspects éthiques de l'IA, ainsi que des questions plus vastes comme celles liées à la prise de décision automatisée.
- Appuyer les efforts voués à la sensibilisation du public pour faire connaître les avantages avérés et potentiels ainsi que les implications plus vastes de l'IA.
- Continuer de promouvoir des démarches neutres sur le plan technologique et appropriées sur les plans technique et éthique.
- Soutenir la libre circulation de l'information grâce à la mise en commun de pratiques exemplaires et de cas d'usage sur la fourniture de données gouvernementales ouvertes, interopérables et accessibles par des moyens sécurisés pour la programmation en IA.
- Diffuser la déclaration du G7 à l'échelle mondiale pour promouvoir le développement de l'IA et la collaboration sur la scène internationale (G7, 2018[47]).

À Charlevoix en juin 2018, le G7 a publié un communiqué qui promeut une approche de l'intelligence artificielle centrée sur l'humain et l'adoption commerciale de l'IA. Les membres du G7 sont également convenus à cette occasion de continuer à promouvoir des démarches neutres sur le plan technologique et appropriées sur les plans technique et éthique.

Les ministres de l'innovation du G7 ont décidé d'organiser une conférence multipartite sur l'IA au Canada en décembre 2018. Les discussions ont porté en particulier sur les moyens de concrétiser le potentiel transformationnel positif de l'IA pour favoriser une croissance économique inclusive et durable. La France devait également proposer des initiatives concernant l'IA dans le cadre de sa présidence du G7 en 2019.

On notera par ailleurs l'attention que le G20 porte à l'IA, le Japon ayant en particulier proposé de tenir des discussions à ce sujet dans le cadre de sa présidence en 2019 (G20, 2018[1]). Les ministres de l'économie numérique du G20, réunis à Salta (Argentine) en 2018, avaient encouragé les pays à permettre aux individus et aux entreprises de tirer avantage de la transformation numérique et des technologies émergentes comme les réseaux 5G, l'internet des objets et l'intelligence artificielle. Ils avaient invité le Japon à poursuivre, pendant sa présidence du G20 en 2019, le travail accompli en 2018 au sein du G20 dans plusieurs domaines prioritaires, parmi lesquels l'IA.

OCDE

Principes de l'OCDE sur la confiance dans l'IA et son adoption

En mai 2018, le Comité de la politique de l'économie numérique de l'OCDE a créé le Groupe d'experts sur l'intelligence artificielle à l'OCDE (AIGO) chargé d'élaborer des principes à mettre en œuvre dans le cadre des politiques publiques et de la coopération internationale et ayant vocation à promouvoir la confiance dans l'IA et son adoption. Ces principes ont nourri l'élaboration de la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle* (OCDE, 2019[3]), à laquelle 42 pays ont adhéré le 22 mai 2019. Dans le même esprit, la présidence de la Réunion du Conseil au niveau des Ministres (RCM) de 2018 a exhorté « l'OCDE à poursuivre les discussions avec les diverses parties prenantes sur l'élaboration possible de principes devant guider le développement et l'application éthique de l'intelligence artificielle (IA) au profit des personnes ».

Le Groupe d'experts rassemble plus de 50 experts de disciplines et de secteurs différents ; les administrations, les entreprises, la communauté technique, les syndicats et la société civile, ainsi que la Commission européenne et l'UNESCO, y sont représentés. Il a tenu quatre réunions : deux au siège de l'OCDE, à Paris, les 24 et 25 septembre, et le 12 novembre 2018, une au Massachusetts Institute of Technology (MIT) à Cambridge les 16 et 17 janvier 2019, et la dernière à Dubaï, les 8 et 9 février 2019, en marge du World Government Summit. Le groupe d'experts a défini les principes d'une approche responsable à l'appui d'une IA digne de confiance, valant pour l'ensemble des parties prenantes. Parmi ces principes, citons notamment le respect des droits de l'homme, l'équité, la transparence et l'explicabilité, la robustesse, la sûreté et la sécurité, ou encore la responsabilité. Il a également proposé des recommandations spécifiques pour la mise en œuvre de ces principes dans le cadre des politiques nationales. Ces travaux ont étayé l'élaboration de la *Recommandation du Conseil de l'OCDE sur l'intelligence artificielle,* au premier semestre de 2019 (OCDE, 2019[3]).

Observatoire des politiques relatives à l'IA

L'OCDE prévoit la création en 2019 d'un Observatoire des politiques relatives à l'IA chargé d'examiner les évolutions actuelles et à venir de l'IA et leurs implications en termes de

politiques publiques. Le but est de soutenir la mise en œuvre des principes susmentionnés en travaillant avec un large éventail d'acteurs extérieurs (administrations, entreprises, universitaires, experts techniques et grand public). L'Observatoire a vocation à devenir un centre pluridisciplinaire et multipartite ayant pour mission d'aider à la collecte de données probantes, d'éclairer les débats et de guider les pouvoirs publics aux fins de l'élaboration des politiques. Il offrira en outre aux partenaires extérieurs un point d'accès unique aux activités d'IA et aux conclusions des travaux connexes intéressant l'action publique dans l'ensemble des pays de l'OCDE.

Commission européenne et autres institutions européennes

La Commission européenne a publié en avril 2018 une communication sur « L'intelligence artificielle pour l'Europe », qui énonce trois grandes priorités : premièrement, renforcer la capacité technologique et industrielle de l'Union européenne et intensifier le recours à l'IA dans tous les secteurs de l'économie ; deuxièmement, se préparer aux changements socio-économiques induits par l'IA ; troisièmement, garantir l'existence d'un cadre éthique et juridique approprié. La Commission a également présenté en décembre 2018 un plan coordonné sur le développement de l'IA en Europe. Ce plan a principalement pour but de maximiser l'impact des investissements et de définir collectivement la marche à suivre dans le domaine de l'intelligence artificielle. Il sera appliqué jusqu'en 2027 et contient environ 70 mesures individuelles dans les domaines suivants :

- **Actions stratégiques et coordination** : Inciter les États membres à mettre en place des stratégies nationales d'IA indiquant les niveaux d'investissement et les mesures de mise en œuvre prévus.
- **Maximisation des investissements par le biais de partenariats** : Promouvoir l'investissement dans la recherche et l'innovation stratégiques en IA au moyen de partenariats public-privé, d'un groupe de leaders et d'un fonds spécifique pour soutenir les jeunes pousses et PME innovantes du secteur de l'IA.
- **Du laboratoire au marché** : Renforcer les centres d'excellence dans la recherche en IA et les pôles d'innovation numériques, et mettre en place des installations d'essai et, éventuellement des « bacs à sable réglementaires ».
- **Compétences et apprentissage tout au long de la vie** : Favoriser le talent, les compétences et l'apprentissage tout au long de la vie.
- **Données** : Appeler à la création d'un Espace européen des données pour faciliter l'accès aux données d'intérêt public, et de plateformes de données industrielles pour l'IA, y compris pour les données de santé.
- **Intégration de l'éthique dès la conception et cadre réglementaire** : Souligner la nécessité d'une éthique de l'IA et d'un cadre réglementaire adapté à cette fin (couvrant notamment les questions de sécurité et de responsabilité). Ce cadre éthique s'appuiera sur les lignes directrices en matière d'éthique pour le développement et l'utilisation de l'intelligence artificielle élaborées par un groupe d'experts indépendant (AI HLEG). La Commission européenne s'engage aussi à exiger dans ses politiques d'approvisionnement la prise en compte des considérations éthiques dès le stade de la conception des systèmes d'IA (*ethics by design*).
- **IA dans le secteur public** : Définir des mesures régissant l'adoption de l'IA dans le secteur public, notamment en matière d'achats conjoints et de traduction.

- **Coopération internationale** : Faire valoir l'importance de l'action au niveau international, intégrer l'IA dans les politiques de développement et organiser une réunion mondiale au niveau ministériel en 2019.

Dans le cadre de sa stratégie pour l'IA, la Commission a également créé en juin 2018 le Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle (AI HLEG). Ce groupe d'experts, qui rassemble des représentants des milieux universitaires, de la société civile et des entreprises, s'est vu confier deux missions : premièrement, élaborer à l'intention des développeurs, responsables de déploiement et utilisateurs des systèmes d'IA un projet de « Lignes directrices en matière d'éthique pour une IA digne de confiance » ; deuxièmement, préparer pour la Commission européenne et les États membres des recommandations, destinées à étayer les politiques d'IA et les investissements connexes, sur les évolutions à moyen et long termes de l'IA à même de stimuler la compétitivité mondiale de l'Europe. La Commission a créé parallèlement un forum multipartite, l'Alliance européenne pour l'IA, pour permettre de larges discussions sur les politiques d'IA en Europe. Toute personne peut contribuer, via une plateforme en ligne, aux travaux du groupe d'experts et participer ainsi à l'élaboration des politiques de l'UE.

Le Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle a publié un premier projet de lignes directrices en matière d'éthique pour commentaires en décembre 2018. Ce document établit un cadre, fondé sur la Charte des droits fondamentaux de l'UE, en vue de tendre vers une IA digne de confiance, c'est-à-dire légale, éthique et robuste d'un point de vue sociotechnique. Il énumère une série de principes éthiques applicables à l'IA. Il énonce également un certain nombre de critères clés pour garantir la fiabilité de l'IA, ainsi que les méthodes à utiliser pour les mettre en œuvre. Enfin, ce projet de lignes directrices contient une liste d'évaluation non exhaustive qui, sous forme de questions pratiques au regard de chaque critère, aidera les acteurs concernés à contrôler la bonne application des principes éthiques. À la date de rédaction de ce chapitre, le groupe d'experts travaillait encore à la révision des lignes directrices, sur la base des commentaires recueillis, afin de pouvoir les soumettre officiellement à la Commission européenne le 9 avril 2019. Cette dernière devait alors préciser les étapes à suivre en vue de l'adoption des lignes directrices et d'un cadre éthique mondial pour l'IA. Les recommandations, deuxième volet de travail du groupe d'experts, doivent être prêtes à l'été 2019.

En 2017, l'Assemblée parlementaire du Conseil de l'Europe a publié une Recommandation sur « La convergence technologique, l'intelligence artificielle et les droits de l'homme ». Cette recommandation invite instamment le Comité des Ministres à charger les organes compétents du Conseil de l'Europe d'examiner la manière dont les technologies émergentes comme l'IA remettent en question les droits de l'homme. Elle propose également que des lignes directrices soient élaborées sur des questions comme la transparence, la responsabilité et le profilage. En février 2019, le Comité des Ministres du Conseil de l'Europe a adopté une Déclaration sur les capacités de manipulation des processus algorithmiques. Cette déclaration reconnaît « les dangers qui menacent les sociétés démocratiques » liés à la capacité des « outils d'apprentissage automatique (…) d'influencer les émotions et les pensées » et encourage les États membres à combattre ces dangers. En février 2019, le Conseil de l'Europe a organisé une conférence de haut niveau sur le thème « Maîtriser les règles du jeu – L'impact du développement de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'état de droit ».

En outre, la Commission européenne pour l'efficacité de la justice du Conseil de l'Europe a adopté en décembre 2018 la première Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement. Cette Charte définit cinq

principes devant guider le développement d'outils d'IA dans les systèmes judiciaires européens. En 2019, la Commission des questions juridiques et des droits de l'homme de l'Assemblée parlementaire du Conseil de l'Europe a décidé de créer une sous-commission sur l'intelligence artificielle et les droits de l'homme.

En mai 2017, le Comité économique et social européen (CESE) a adopté un avis sur l'impact sociétal de l'IA. Cet avis appelle les parties prenantes dans l'UE à veiller à ce que le développement, le déploiement et l'utilisation de l'IA soient bénéfiques à la société et contribuent au bien-être social. Le CESE insiste sur la nécessité que les humains conservent le contrôle du moment et des modalités d'utilisation de l'IA dans la vie quotidienne ; il pointe les domaines dans lesquels l'IA soulève des enjeux de société, tels que l'éthique ; la sécurité ; la transparence ; la vie privée ; les normes ; le travail ; l'éducation ; l'égalité d'accès ; la législation et la réglementation ; la gouvernance et la démocratie ; la guerre ; et la superintelligence. Le CESE préconise également l'instauration d'un code de déontologie de l'IA au niveau paneuropéen et l'adaptation des stratégies en matière d'emploi. Enfin, il plaide pour une infrastructure d'IA européenne, composée d'environnements d'apprentissage à code source libre (Muller, 2017[48]). Le CESE a créé un groupe de travail temporaire sur l'IA pour examiner plus en détail ces questions.

Région nordique-balte

En mai 2018, les ministres des pays nordiques et baltes ont signé une déclaration commune sur « L'intelligence artificielle dans la région nordique-balte ». Les pays de la région comprennent le Danemark, l'Estonie, la Finlande, les Îles Féroé, l'Islande, la Lettonie, la Lituanie, la Norvège, la Suède et les Îles d'Åland. Ils ont décidé de renforcer leur coopération en matière d'IA, tout en maintenant leur position de première région d'Europe en termes de progression du numérique (Nordic, 2018[49]). La déclaration mentionne sept axes d'intervention essentiels pour le développement et la promotion de l'utilisation de l'IA au service des individus : i) accroître les possibilités de développement des compétences, afin de permettre à un plus grand nombre d'administrations publiques, d'entreprises et d'organisations d'utiliser l'IA ; ii) améliorer l'accès aux données exploitables par l'IA, afin de fournir des services de meilleure qualité aux citoyens et aux entreprises de la région ; iii) élaborer des lignes directrices, normes, valeurs et principes éthiques clairs indiquant quand et comment utiliser les applications d'IA ; iv) veiller à ce que l'infrastructure, le matériel, les logiciels et les données, qui jouent un rôle clé dans l'utilisation de l'IA, reposent sur des normes garantissant l'interopérabilité, le respect de la vie privée, la sécurité, la confiance, la maniabilité et la portabilité ; v) faire en sorte que l'IA figure en bonne place dans les débats menés au niveau européen et les initiatives mises en œuvre dans le cadre du marché unique numérique ; vi) éviter toute réglementation inutile dans ce domaine, qui se développe rapidement ; vii) faciliter la collaboration dans les domaines pertinents de l'action publique via le Conseil nordique des ministres.

Nations Unies

En septembre 2017, l'Institut interrégional de recherche des Nations Unies sur la criminalité et la justice a signé un accord avec les Pays-Bas, le pays hôte, en vue de l'ouverture, dans le cadre du système de l'ONU, d'un Centre pour l'intelligence artificielle et la robotique à La Haye[2].

L'Union internationale des télécommunications (UIT) a travaillé de concert avec plus de 25 autres organes de l'ONU pour organiser le Sommet mondial *« AI for Good »*. Elle a également noué des partenariats avec des organisations comme la Fondation XPRIZE et

l'ACM (Association for Computing Machinery). Suite au premier sommet de juin 2017, l'UIT a organisé un deuxième sommet, qui s'est tenu à Genève en mai 2018[3].

L'UNESCO est à l'initiative d'un dialogue mondial sur l'éthique de l'intelligence artificielle, afin de mener une réflexion sur sa complexité et ses incidences sur la société et l'humanité en général. Elle a organisé en septembre 2018 une table ronde publique avec des experts puis, en mars 2019, une conférence mondiale sur le thème « Principes pour l'IA : vers une approche humaniste ? ». Ces deux événements visaient à sensibiliser le public et encourager la réflexion sur les opportunités et les défis associés à l'IA et aux technologies apparentées. En novembre 2019, la 40e Conférence générale de l'UNESCO pourrait examiner la possibilité d'élaborer une recommandation sur l'IA en 2020-21, sous réserve de l'approbation du Conseil exécutif de l'UNESCO en avril 2019.

Organisation internationale de normalisation

L'Organisation internationale de normalisation (ISO) et la Commission électrotechnique internationale (IEC) ont créé en 1987 un comité technique commun ISO/IEC JTC 1 chargé de définir des normes pour les applications professionnelles et grand public des technologies de l'information. En octobre 2017, le sous-comité 42 (SC 42) a été créé sous l'égide du comité technique JTC 1 afin d'élaborer des normes en matière d'IA. Ce sous-comité produit des directives à l'intention des comités de l'ISO et de l'IEC qui travaillent sur les applications de l'IA. Il se charge notamment d'établir un cadre et une terminologie communs, d'identifier les approches computationnelles et architectures des systèmes d'IA, et d'évaluer les risques et les menaces qui leur sont associés (Price, 2018[50]).

Initiatives d'acteurs privés

Un grand nombre de partenariats et d'initiatives ont été lancés par des acteurs non gouvernementaux pour réfléchir aux questions soulevées par l'intelligence artificielle. Bien que beaucoup de ces initiatives présentent un caractère multipartite, la présente section mentionne avant tout celles qui émanent de la communauté technique, du secteur privé, des syndicats ou des milieux universitaires. La liste ci-dessous n'est donc pas exhaustive.

Communauté technique et milieux universitaires

L'IEEE a lancé en avril 2016 une Initiative mondiale sur l'éthique dans la conception des systèmes autonomes et intelligents. Cette initiative a pour but de faire progresser le débat public sur la mise en œuvre des technologies d'IA et de définir des normes éthiques et valeurs prioritaires à cet égard. L'IEEE a également publié en décembre 2017 la deuxième version de ses Principes d'intégration de l'éthique dès la conception (*Ethically Aligned Design*), en sollicitant les commentaires du public. La version finale devrait être publiée en 2019 (Tableau 5.3) (IEEE, 2017[51]). Par ailleurs, l'IEEE a annoncé en juin 2018 la création avec le Media Lab du MIT d'un « conseil pour une approche élargie de l'intelligence » (*Council for Extended Intelligence*). Cet organe cherche à promouvoir le développement responsable de systèmes intelligents, la reprise du contrôle sur les données à caractère personnel et l'élaboration d'indicateurs de la prospérité économique autres que le produit intérieur brut (Pretz, 2018[52]).

Les Principes d'Asilomar, qui comprennent 23 principes pour un développement sûr et socialement bénéfique de l'IA à court et plus long termes, sont issus d'une conférence organisée par le Future of Life Institute en janvier 2017. Ils ont été élaborés à partir de débats, de réflexions et de documents émanant de l'IEEE, du monde universitaire et d'organisations à but non lucratif.

Tableau 5.3. Principes généraux de l'IEEE pour une intégration de l'éthique dès la conception (*Ethically Aligned Design,* deuxième version)

Principes	Objectifs
Droits de l'homme	Veiller à ce que les systèmes autonomes et intelligents (SAI) ne portent pas atteinte aux droits fondamentaux reconnus au niveau international
Priorité au bien-être	Donner la priorité aux indicateurs de bien-être dans la conception et l'utilisation des SAI, les indicateurs traditionnels de la prospérité ne prenant pas pleinement en compte les effets des technologies liées aux systèmes d'intelligence artificielle sur le bien-être des individus
Responsabilité/redevabilité	Veiller à ce que les concepteurs et opérateurs soient responsables et redevables du fonctionnement des SAI
Transparence	Assurer la transparence du fonctionnement des SAI
Conscience des risques d'utilisation abusive	Réduire au minimum les risques d'utilisation abusive des SAI

Source : IEEE (2017[51]), *Ethically Aligned Design (Version 2)*, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf.

Ces principes sont classés en trois groupes. Le premier, qui porte sur la recherche, appelle à : financer la recherche sur une utilisation bénéfique de l'IA, qui prenne en compte des questions épineuses en matière d'informatique, d'économie, de droit, d'éthique et de sciences sociales ; développer des relations constructives entre scientifiques et décideurs politiques ; et insuffler une culture de la recherche technique fondée sur la coopération, la confiance et la transparence. Le second, qui concerne l'éthique et les valeurs, appelle à prendre en compte un certain nombre d'exigences dans la conception et le fonctionnement des systèmes d'IA : la sûreté et la sécurité ; la transparence et la responsabilité ; le respect de la liberté individuelle et de la vie privée, de la dignité humaine, des droits fondamentaux et de la diversité culturelle ; et l'autonomisation des individus et le partage des avantages. Le troisième recommande notamment, à plus long terme, d'éviter les hypothèses fortes au sujet des capacités maximum des futurs systèmes d'IA et de planifier soigneusement le développement éventuel de l'intelligence artificielle générale (IAG) (FLI, 2017[53]). Le Tableau 5.4 présente la liste des Principes d'Asilomar les plus importants.

OpenAI, une organisation de recherche en IA à but non lucratif, a été créée fin 2015. Elle emploie 60 chercheurs à plein temps dont la mission est de « développer des systèmes d'IAG sûrs, en veillant à que les avantages qui en résultent soient répartis de la façon la plus large et la plus équitable possible »[4].

The Future Society a lancé en 2015 une initiative sur l'intelligence artificielle (*AI Initiative*) afin de contribuer à l'élaboration d'un cadre mondial pour les politiques d'IA. L'organisation héberge une plateforme en ligne permettant un débat civique et des échanges pluridisciplinaires. Cette plateforme a pour but d'aider à comprendre la dynamique des technologies d'IA, ainsi que les avantages et les risques qui leur sont associés, en vue d'étayer la formulation de recommandations d'action[5].

De multiples initiatives universitaires sont également en cours dans tous les pays de l'OCDE et de nombreuses économies partenaires. L'initiative de recherche sur les politiques de l'internet (*Internet Policy Research Initiative*) du MIT, par exemple, vise à jeter un pont entre la communauté technique et celle des décideurs. En 2017, le Centre Berkman Klein de l'Université d'Harvard a lancé une initiative sur l'éthique et la gouvernance de l'intelligence artificielle (*Ethics and Governance of Artificial Intelligence Initiative*). De son côté, le Media Lab du MIT travaille plus spécialement sur des questions comme : les algorithmes et l'équité, les véhicules autonomes et la transparence, et l'explicabilité des systèmes d'IA.

Tableau 5.4. Principes d'Asilomar sur l'intelligence artificielle (intitulés des principes essentiels)

	Recherche	Éthique et valeurs	Perspectives à plus long terme
Titres des principes	– Finalité de la recherche – Financement de la recherche – Relations entre scientifiques et décideurs – Culture de la recherche – Refus de la course à tout prix	– Sécurité – Transparence en cas de défaillance – Transparence judiciaire – Responsabilité – Concordance des valeurs – Valeurs humaines – Données personnelles – Liberté et respect de la vie privée – Bénéfice collectif – Partage de la prospérité – Contrôle humain	– Prudence au sujet des capacités futures – Ampleur des transformations à venir – Risques – Récursivité et réplicabilité des systèmes d'IA – Bien commun

Source : FLI (2017[53]), *Asilomar IA Principles*, https://futureoflife.org/ai-principles/.

Initiatives du secteur privé

En septembre 2016, Amazon, DeepMindGoogle, Facebook, IBM et Microsoft ont lancé le Partenariat pour l'intelligence artificielle au service des citoyens et de la société (*Partnership on Artificial Intelligence to Benefit People and Society*, ou PAI). Cette initiative a pour but d'étudier et de promouvoir les bonnes pratiques relatives aux technologies d'IA, d'aider le public à mieux comprendre l'IA et d'offrir une plateforme ouverte de discussion et d'engagement autour de l'IA et de ses implications pour les individus et la société. Depuis sa création, ce partenariat est devenu une communauté pluridisciplinaire de parties prenantes et compte aujourd'hui plus de 80 membres comprenant aussi bien des entreprises de technologie à but lucratif que des représentants de la société civile, des universités et instituts de recherche, ou des startups.

Le Conseil de l'industrie des technologies de l'information (*Information Technology Industry Council*, ou ITI) est une association d'entreprises de technologie basée à Washington, qui compte plus d'une soixantaine de membres. L'ITI a publié en octobre 2017 des principes applicables aux politiques en matière d'IA (*AI Policy Principles*) (Tableau 5.5). Ce document précise les responsabilités du secteur dans certains domaines, invite les gouvernements à soutenir la recherche en IA et appelle au développement des partenariats public-privé (ITI, 2017[54]). Les entreprises ont aussi pris individuellement certaines initiatives.

Tableau 5.5. Principes de l'ITI applicables aux politiques en matière d'intelligence artificielle

Responsabilité des entreprises : promouvoir le développement et l'utilisation responsables de l'IA	Opportunité pour les gouvernements : investir dans l'écosystème de l'IA et faciliter sa croissance	Opportunité pour les partenariats public-privé : promouvoir l'éducation tout au long de la vie et la diversité
– Conception et déploiement responsables – Sûreté et contrôlabilité – Robustesse et représentativité des données – Interprétabilité – Responsabilité des systèmes d'IA autonomes	– Investissement dans la recherche et le développement en matière d'IA – Démarche réglementaire flexible – Promotion de l'innovation et de la sécurité de l'internet – Cybersécurité et protection de la vie privée – Élaboration de normes mondiales et promotion des meilleures pratiques	– Démocratisation de l'accès et égalité des chances – Enseignement des sciences, des technologies, de l'ingénierie et des mathématiques – Préparation et adaptation de la main-d'œuvre – Partenariats public-privé

Source : ITI (2017[54]), *AI Policy Principles*, https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf.

Société civile

The Public Voice, un regroupement d'associations créé par l'EPIC (Electronic Privacy Information Center), a publié en octobre 2018 des lignes directrices universelles sur l'intelligence artificielle (*Universal Guidelines on Artificial Intelligence*, UGAI) (The Public Voice, 2018[55]). Ces lignes directrices attirent l'attention sur les enjeux croissants associés aux systèmes informatiques intelligents et formulent des recommandations concrètes pour améliorer et étayer leur conception. Elles s'attachent essentiellement à promouvoir la transparence et la responsabilité des systèmes d'IA, et à garantir que les humains conservent le contrôle des systèmes qu'ils créent[6]. Les douze principes énumérés dans ces lignes directrices énoncent toute une série de droits et d'obligations, en particulier : le droit à la transparence ; le droit à la détermination humaine ; l'obligation d'identification ; l'obligation d'équité ; l'obligation d'évaluation préalable et de responsabilité ; l'obligation de précision, de fiabilité et de validité ; l'obligation de qualité des données ; l'obligation de sécurité publique ; l'obligation de cybersécurité ; et l'obligation de supprimer un système d'IA dans l'éventualité où ce système ne pourrait plus être contrôlé par un humain. À cela s'ajoutent l'interdiction du profilage secret et l'interdiction de tout système de notation centralisé reposant sur l'intelligence artificielle.

Syndicats

UNI Global Union est un syndicat qui représente plus de 20 millions de travailleurs qualifiés et de travailleurs du secteur des services répartis dans plus de 150 pays. Préparer un avenir garantissant l'autonomie des travailleurs et l'accès à un emploi décent est une priorité phare d'UNI Global Union. C'est pourquoi le syndicat a formulé « **Dix principes majeurs pour une intelligence artificielle éthique » (*Top 10 Principles for Ethical AI*). Le but est d'assurer le** respect des droits des travailleurs dans l**es conventions** collectives et les accords-cadres mondiaux, ainsi qu'au niveau des organisations plurinationales de syndicats et de délégués syndicaux et des organisations mondiales (Tableau 5.1) (Colclough, 2018[56]).

Tableau 5.6. Dix principes majeurs pour une intelligence artificielle éthique (UNI Global Union)

1. Réclamer que les systèmes d'IA soient transparents	Les travailleurs doivent avoir le droit de réclamer la transparence des décisions/résultats des systèmes d'IA et des algorithmes sous-jacents. Ils doivent aussi être consultés sur la mise en œuvre, le développement et le déploiement des systèmes d'IA.
2. Doter les systèmes d'IA d'une « boîte noire éthique »	La « boîte noire éthique » doit contenir non seulement les données pertinentes pour garantir la transparence et la redevabilité d'un système, mais aussi des données et des informations claires sur les considérations éthiques intégrées à ce système.
3. Faire en sorte que l'IA serve les individus et la planète	La mise en place de codes d'éthique pour le développement, l'application et l'utilisation de l'IA est nécessaire pour que tout au long de leur processus opérationnel, les systèmes d'IA restent compatibles avec, et renforcent, les principes de dignité humaine, d'intégrité, de liberté, de respect de la sphère privée et de diversité culturelle et de genre, ainsi que les droits fondamentaux.
4. Adopter une approche donnant les commandes à l'être humain	Un préalable absolu est que le développement de l'IA soit responsable, sûr et utile, que les machines conservent le statut juridique d'outils et que des personnes morales conservent le contrôle sur ces machines et la responsabilité envers elles en tout temps.
5. Veiller à une IA sans distinction de genre ni préjugés	Dans la conception et la maintenance de l'IA, il est vital que le système subisse des contrôles afin d'en éliminer les préjugés négatifs ou préjudiciables envers l'être humain, et de veiller à ce que tout préjugé (genre, origine, orientation sexuelle, âge) soit identifié et ne soit pas propagé par le système.
6. Faire partager les avantages des systèmes d'IA	La prospérité économique créée par l'IA devrait être répartie largement et de manière égale, au profit de l'ensemble de l'humanité. Des politiques mondiales autant que nationales visant à surmonter le fossé numérique économique, technologique et social sont donc nécessaires.
7. Assurer une transition juste et garantir le soutien des	Au fur et à mesure du développement des systèmes d'IA et de la création de réalités augmentées, les travailleurs perdront leur emploi et les tâches professionnelles disparaîtront. Il est vital que des politiques soient mises en place pour garantir une transition juste menant à la

droits et libertés fondamentales	réalité numérique, y compris des mesures gouvernementales spécifiques pour aider les travailleurs qui ont perdu leur emploi à en trouver un nouveau.
8. Établir un mécanisme de gouvernance mondial	Il est nécessaire de créer des instances de gouvernance sur le travail décent et l'IA éthique, regroupant toutes les parties prenantes au niveau mondial comme au niveau régional. Ces instances devraient inclure les concepteurs d'IA, les fabricants, les propriétaires, les développeurs, les chercheurs, les employeurs, les juristes, les organisations de la société civile et les syndicats.
9. Interdire l'attribution de responsabilités aux robots	Les robots devraient être conçus et exploités dans la mesure du possible de manière à respecter la législation existante et les droits et libertés fondamentales, y compris le respect de la sphère privée.
10. Interdire la course aux armements d'IA	Les armes létales autonomes, y compris la cyberguerre, devraient être interdites. UNI Global Union appelle à l'adoption d'une convention mondiale sur l'IA éthique pour prévenir et gérer les conséquences négatives indésirables de l'IA, tout en optimisant les avantages de l'IA pour les travailleurs et la société. UNI Global Union souligne que les agents responsables doivent être les êtres humains et les entreprises.

Source : Colclough (2018[56]), « Ethical artificial intelligence – 10 essential ingredients », https://www.oecd-forum.org/channels/722-digitalisation/posts/29527-10-principles-for-ethical-artificial-intelligence.

Références

Allemagne (2018), *Artificial Intelligence Strategy*, Gouvernement fédéral allemand, https://www.ki-strategie-deutschland.de/home.html. [5]

Allemagne (2017), *Automated and Connected Driving*, BMVI Ethics Commission, https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile. [6]

Benhamou, S. et L. Janin (2018), *Intelligence artificielle et travail*, France Stratégie, http://www.strategie.gouv.fr/publications/intelligence-artificielle-travail. [28]

Brésil (2018), *Brazilian digital transformation strategy « E-digital »*, Ministério da Ciência, Tecnologia, Inovações e Comunicações, http://www.mctic.gov.br/mctic/export/sites/institucional/sessaoPublica/arquivos/digitalstrategy.pdf. [7]

Chine (2018), *AI Standardisation White Paper*, Gouvernement de Chine, traduit en anglais par Jeffrey Ding, Chercheur pour le Future of Humanity's Governance of AI Program, https://baijia.baidu.com/s?id=1589996219403096393. [14]

Chine (2017), *Guideline on Next Generation AI Development Plan*, Gouvernement de Chine, Conseil d'État, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm. [12]

Chine (2016), *Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018-2020)*, Ministère chinois de l'Industrie et des Technologies de l'information, http://www.miit.gov.cn/n1146290/n1146392/c4808445/content.html. [11]

CIFAR (2017), *Stratégie pancanadienne en matière d'intelligence artificielle*, CIFAR, https://www.cifar.ca/fr/ia/strategie-pancanadienne-en-matiere-dintelligence-artificielle. [8]

Colclough, C. (2018), « Ethical Artificial Intelligence – 10 Essential Ingredients », *A.Ideas Series*, n° 24, Réseau du Forum, Éditions OCDE, Paris, https://www.oecd-forum.org/channels/722-digitalisation/posts/29527-10-principles-for-ethical-artificial-intelligence. [56]

Corée (2016), *Mid- to Long-term Master Plan in Preparation for the Intelligent Information Society*, Gouvernement de Corée, Exercice interministériel, http://english.msip.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf. [16]

Corée (2016), « MSIP announces development strategy for the intelligence information industry », *Science, Technology & ICT Newsletter, Ministry of Science and ICT, Government of Korea,* 16, https://english.msit.go.kr/english/msipContents/contentsView.do?cateId=msse44&artId=1296203. [15]

Danemark (2018), *Strategy for Denmark's Digital Growth*, Gouvernement du Danemark, https://em.dk/english/publications/2018/strategy-for-denmarks-digital-growth. [18]

EOP (2018), *Artificial Intelligence for the American People*, Executive Office of the President, Gouvernement des États-Unis, https://www.whitehouse.gov/briefings-statements/artificial-intelligence-american-people/. [21]

Finlande (2017), *Finland's Age of Artificial Intelligence - Turning Finland into a Leader in the Application of AI*, page web, Ministère finlandais de l'Emploi et de l'Économie, https://tem.fi/en/artificial-intelligence-programme. [24]

FLI (2017), *Asilomar AI Principles*, Future of Life Institute (FLI), https://futureoflife.org/ai-principles/. [53]

Fonds de recherche du Québec (2018), « Québec lays the groundwork for a world observatory on the social impacts of artificial intelligence and digital technologies », *communiqué de presse*, 29 mars, https://www.newswire.ca/news-releases/quebec-lays-the-groundwork-for-a-world-observatory-on-the-social-impacts-of-artificial-intelligence-and-digital-technologies-678316673.html. [9]

G20 (2018), *Ministerial Declaration – G20 Digital Economy*, Réunion ministérielle du G20 sur l'économie numérique, 24 août 2018, Salta, Argentine, https://g20.argentina.gob.ar/sites/default/files/digital_economy_-_ministerial_declaration.pdf. [1]

G7 (2018), *Résumé des présidents : Réunion ministérielle sur le thème « Se préparer aux emplois de l'avenir »*, https://international.gc.ca/world-monde/international_relations-relations_internationales/g7/documents/2018-03-27-chairs_summary-resume_presidents.aspx?lang=fra. [47]

G7 (2017), *Artificial Intelligence (Annex 2)*, http://www.g7italy.it/sites/default/files/documents/ANNEX2-Artificial_Intelligence_0.pdf. [46]

G7 (2016), *Proposal of Discussion toward Formulation of AI R&D Guideline*, Ministère japonais des Affaires intérieures et des Communications, http://www.soumu.go.jp/joho_kokusai/g7ict/english/index.html. [45]

Hall, W. et J. Pesenti (2017), *Growing the Artificial Intelligence Industry in the UK*, Wendy Hall and Jérôme Pesenti, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf. [39]

IEEE (2017), *Ethically Aligned Design (Version 2)*, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf. [51]

Inde (2018), « National Strategy for Artificial Intelligence #AI for All, Discussion Paper », *Discussion Paper*, NITI Aayog, http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf. [29]

Italie (2018), *Artificial Intelligence at the Service of the Citizen*, Agency for Digital Italy, https://libro-bianco-ia.readthedocs.io/en/latest/. [30]

ITI (2017), *AI Policy Principles*, Information Technology Industry Council, https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf. [54]

Japon (2019), *Social Principles for Human-Centric AI*, Japan Cabinet Office, April, https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf. [33]

Japon (2018), *Draft AI Utilization Principles*, Ministry of Internal Affairs and Communications, Japon, http://www.soumu.go.jp/main_content/000581310.pdf. [35]

Japon (2018), *Integrated Innovation Strategy*, Japan Cabinet Office, June, https://www8.cao.go.jp/cstp/english/doc/integrated_main.pdf. [32]

Japon (2017), *Artificial Intelligence Technology Strategy*, Strategic Council for AI Technology, http://www.nedo.go.jp/content/100865202.pdf. [31]

Japon (2017), *Draft AI R&D Guidelines for International Discussions*, Ministry of Internal Affairs and Communications, Japon, http://www.soumu.go.jp/main_content/000507517.pdf. [34]

Jing, M. et S. Dai (2018), *Here's what China is doing to boost its artificial intelligence capabilities*, 10 May, https://www.scmp.com/tech/science-research/article/2145568/can-trumps-ai-summit-match-chinas-ambitious-strategic-plan. [10]

Kaevats, M. (2017), *Estonia's Ideas on Legalising AI*, exposé présenté à la conférence AI: Intelligent Machines, Smart Policies, Paris, les 26 et 27 octobre 2017, https://prezi.com/yabrlekhmcj4/oecd-6-7min-paris/. [19]

Kaevats, M. (25 septembre 2017), *Estonia considers a 'kratt law' to legalise artifical intelligence (AI)*, E-residency blog, https://medium.com/e-residency-blog/estonia-starts-public-discussion-legalising-ai-166cb8e34596. [20]

Kania, E. (2018), « China's AI agenda advances », *The Diplomat* 14 février, https://thediplomat.com/2018/02/chinas-ai-agenda-advances/. [13]

Mexico (2018), *Towards an AI strategy in Mexico: Harnessing the AI Revolution*, British Embassy Mexico City, Oxford Insights, C minds, http://go.wizeline.com/rs/571-SRN-279/images/Towards-an-AI-strategy-in-Mexico.pdf. [36]

Muller, C. (2017), *Avis du CESE: L'intelligence artificielle*, Comité économique et social européen, Bruxelles, https://www.eesc.europa.eu/fr/our-work/opinions-information-reports/opinions/lintelligence-artificielle. [48]

Nordic (2018), *AI in the Nordic-Baltic Region*, Nordic Council of Ministers, https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf. [49]

OCDE (2019), *Recommandation du Conseil sur l'intelligence artificielle*, OCDE, Paris, https://legalinstruments.oecd.org/api/print?ids=648&lang=fr. [3]

OCDE (2019), *Scoping Principles to Foster Trust in and Adoption of AI – Proposal by the Expert Group on Artificial Intelligence at the OECD (AIGO)*, Éditions OCDE, Paris, http://oe.cd/ai. [2]

OGCR (2018), *Analysis of the Development Potential of Artificial Intelligence in the Czcech Republic*, Office of the Government of the Czcech Republic, https://www.vlada.cz/assets/evropske-zalezitosti/aktualne/AI-Summary-Report.pdf. [37]

Peng, T. (2018), « South Korea aims high on AI, pumps $2 billion into R&D », *Medium,* 16 May, https://medium.com/syncedreview/south-korea-aims-high-on-ai-pumps-2-billion-into-r-d-de8e5c0c8ac5. [17]

Porter, M. (1990), « The competitive advantage of nations », *Harvard Business Review,* March-April, https://hbr.org/1990/03/the-competitive-advantage-of-nations. [4]

Pretz, K. (2018), « IEEE Standards Association and MIT Media Lab form council on extended intelligence », *IEEE Spectrum*, http://theinstitute.ieee.org/resources/ieee-news/ieee-standards-association-and-mit-media-lab-form-council-on-extended-intelligence. [52]

Price, A. (2018), « First international standards committee for entire AI ecosystem », *IE e-tech,* 03, https://iecetech.org/Technical-Committees/2018-03/First-International-Standards-committee-for-entire-AI-ecosystem. [50]

RU (2018), *AI Sector Deal*, Department for Business, Energy & Industrial Strategy and Department for Digital, Culture, Media & Sport, Gouvernement du Royaume-Uni, https://www.gov.uk/government/publications/artificial-intelligence-sector-deal. [41]

RU (2018), *Centre for Data Ethics and Innovation Consultation*, Department for Digital, Culture, Media & Sport, Gouvernement du Royaume-Uni, https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation. [42]

RU (2017), *Industrial Strategy: Building a Britain Fit for the Future*, Gouvernement du Royaume-Uni, https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future. [40]

RU (2017), *UK Digital Strategy*, Gouvernement du Royaume-Uni, https://www.gov.uk/government/publications/uk-digital-strategy. [38]

Russie (2017), *Digital Economy of the Russian Federation*, Gouvernement de la Fédération de Russie, http://pravo.gov.ru. [23]

Singapour (2018), *Digital Economy Framework for Action*, Infocomm Media Development Authority, https://www.imda.gov.sg/-/media/imda/files/sg-digital/sgd-framework-for-action.pdf?la=en. [43]

Sivonen, P. (2017), *Ambitious Development Program Enabling Rapid Growth of AI and Platform Economy in Finland*, exposé présenté à la conférence AI Intelligent Machines, Smart Policies, Paris, les 26 et 27 octobre 2017, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-sivonen.pdf. [25]

The Public Voice (2018), *Universal Guidelines for Artificial Intelligence*, The Public Voice Coalition, October, https://thepublicvoice.org/ai-universal-guidelines/memo/. [55]

Thompson, N. (2018), « Emmanuel Macron talks to WIRED about France's AI strategy », *WIRED,* 31 March, https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy. [27]

US (2017), « Delaney launches bipartisan artificial intelligence (AI) caucus for 115th Congress », Congressional Artificial Intelligence Caucus, communiqué de presse, 24 mai, https://artificialintelligencecaucus-olson.house.gov/media-center/press-releases/delaney-launches-ai-caucus. [22]

Villani, C. (2018), *Donner un sens à l'intelligence artificielle - Pour une stratégie nationale et européenne*, AI for Humanity, consulté en décembre 2018, https://www.aiforhumanity.fr/. [26]

Vinnova (2018), *Artificial Intelligence in Swedish Business and Society*, Vinnova, 28 octobre, https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf. [44]

Notes

[1] Ce rapport a été commandité par l'ambassade britannique au Mexique, financé par le Prosperity Fund du Royaume-Uni et élaboré par Oxford Insights et C Minds, avec la collaboration du gouvernement mexicain et la contribution d'experts de l'ensemble du Mexique.

[2] Voir www.unicri.it/news/article/2017-09-07_Establishment_of_the_UNICRI.

[3] Voir https://www.itu.int/en/ITU-T/AI/.

[4] Voir https://openai.com/about/#mission.

[5] Voir http://ai-initiative.org/ai-consultation/.

[6] Pour plus de détails, voir https://thepublicvoice.org/ai-universal-guidelines/memo/.

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

L'OCDE est un forum unique en son genre où les gouvernements oeuvrent ensemble pour relever les défis économiques, sociaux et environnementaux que pose la mondialisation. L'OCDE est aussi à l'avant-garde des efforts entrepris pour comprendre les évolutions du monde actuel et les préoccupations qu'elles font naître. Elle aide les gouvernements à faire face à des situations nouvelles en examinant des thèmes tels que le gouvernement d'entreprise, l'économie de l'information et les défis posés par le vieillissement de la population. L'Organisation offre aux gouvernements un cadre leur permettant de comparer leurs expériences en matière de politiques, de chercher des réponses à des problèmes communs, d'identifier les bonnes pratiques et de travailler à la coordination des politiques nationales et internationales.

Les pays membres de l'OCDE sont : l'Allemagne, l'Australie, l'Autriche, la Belgique, le Canada, le Chili, la Corée, le Danemark, l'Espagne, l'Estonie, les États-Unis, la Finlande, la France, la Grèce, la Hongrie, l'Irlande, l'Islande, Israël, l'Italie, le Japon, la Lettonie, la Lituanie, le Luxembourg, le Mexique, la Norvège, la Nouvelle-Zélande, les Pays-Bas, la Pologne, le Portugal, la République slovaque, la République tchèque, le Royaume-Uni, la Slovénie, la Suède, la Suisse et la Turquie. La Commission européenne participe aux travaux de l'OCDE.

Les Éditions OCDE assurent une large diffusion aux travaux de l'Organisation. Ces derniers comprennent les résultats de l'activité de collecte de statistiques, les travaux de recherche menés sur des questions économiques, sociales et environnementales, ainsi que les conventions, les principes directeurs et les modèles développés par les pays membres.

ÉDITIONS OCDE, 2, rue André-Pascal, 75775 PARIS CEDEX 16
ISBN 978-92-64-60811-5 – 2019

L'intelligence artificielle dans la société

Le paysage technique de l'intelligence artificielle (IA) s'est métamorphosé depuis 1950, lorsqu'Alan Turing s'interrogeait pour la première fois sur la capacité des machines à penser. Aujourd'hui, l'IA transforme les économies et les sociétés. Elle promet de générer des gains de productivité, d'améliorer le bien-être et de contribuer à apporter des solutions aux défis mondiaux que sont, par exemple, le changement climatique, l'épuisement des ressources et les crises sanitaires. Cependant, à l'heure où ces applications sont adoptées à travers le monde, leur utilisation soulève un certain nombre d'interrogations et de difficultés ayant trait, entre autres, aux valeurs humaines, à l'équité, à la détermination humaine, à la protection de la vie privée, à la sécurité et à la responsabilité. Le présent rapport contribue à faire émerger une compréhension commune de l'IA, sous sa forme actuelle et dans son évolution à court terme, à travers des relevés du paysage technique, économique, pratique et réglementaire de l'IA et la mise en évidence de grandes considérations de politique publique. Il contribue également à un débat coordonné et cohérent entre les diverses enceintes nationales et internationales.

Cette publication s'inscrit dans le cadre du projet « Going Digital » de l'OCDE. Dans un monde résolument tourné vers le numérique et les données, ce projet vise à fournir aux décideurs les outils dont ils ont besoin pour aider leurs économies et leurs sociétés à prospérer.

Pour plus d'informations, consultez ***www.oecd.org/going-digital***

#GoingDigital

Veuillez consulter cet ouvrage en ligne : ***https://doi.org/10.1787/b7f8cd16-fr***.

Cet ouvrage est publié sur OECD iLibrary, la bibliothèque en ligne de l'OCDE, qui regroupe tous les livres, périodiques et bases de données statistiques de l'Organisation.
Rendez-vous sur le site ***www.oecd-ilibrary.org*** pour plus d'informations.

ISBN 978-92-64-60811-5

éditionsOCDE
www.oecd.org/editions