

Intelligence artificielle et développement humain

Vers un programme de recherche

LIVRE BLANC



IDRC | CRDI

International Development Research Centre
Centre de recherches pour le développement international

Canada



Chercheurs et auteurs

Matthew Smith

msmith@idrc.ca

Spécialiste de programmes principal, Centre de recherches pour le développement international

Sujaya Neupane

Boursier postdoctoral, *McGovern Institute for Brain Research*,
Department of Brain and Cognitive Sciences, MIT

Révision et soutien à la recherche

Greg Leonard

Conception graphique

Claudio Mendonca

ccmdesign.ca

Les auteurs souhaitent remercier Urs Gasser, Amar Ashar, et leurs collègues au *Berkman Klein Centre for Internet and Society* à l'Université Harvard, Laurent Elder, Maroussia Lévesque, Marshall Smith, Pascal Kropf, Sunisha Neupane, Aaron Martin, Nola Haddadian, Raul Zambrano, et Cheryl Chan pour leurs commentaires et leurs points de vue inestimables.

Avril 2018

© Centre de recherches pour le développement international 2018



Diffusé en vertu de la licence d'attribution Creative Commons 4.0.

La recherche présentée dans la présente publication a été financée en partie par le Département du développement international (DFID) du R.-U. Les opinions exprimées ne reflètent pas nécessairement celles du DFID, du CRDI ni du Conseil des gouverneurs du CRDI.



Contenu

RÉSUMÉ	9
INTRODUCTION	21
QU'EST-CE QUE L'IA?	29
AVANTAGES DU DÉVELOPPEMENT	37
Soins de santé.....	39
Prestation de services et de renseignements gouvernementaux.....	41
Agriculture	45
Éducation	49
Économie et affaires.....	55
DÉFIS ET RISQUES	59
Défis contextuels	60
Secteurs de risque	63
Extrapolation dans le futur : risques liés à l'IA dans les pays du Sud.....	93
CONCLUSION ET RECOMMANDATIONS.	97
Recommandations	103
ANNEXE : COMMENT L'IA FONCTIONNE-T-ELLE ?	111
Références	123
Notes	129



ABRÉVIATIONS

IA	Intelligence artificielle
STI	Systèmes de tutorat intelligent
PFR-PRI	Pays à faible revenu et pays à revenu intermédiaire
AA	Apprentissage automatique
CLOT	Cours en ligne ouvert à tous
AR	Apprentissage par renforcement
ODD	Objectifs de développement durable

FIGURES ET TABLEAUX

Tableau 1. Définitions de l'IA	30
Tableau 2. Exemples de comportements de l'intelligence artificielle	32
Figure 1. Structure d'un système de tutorat intelligent	50
Figure 2. Répartition selon le sexe dans les applications d'intelligence artificielle	65
Figure 3. Les risques liés à l'IA et comment ils pourraient entraîner des répercussions sociales négatives	94
Figure 4. Un réseau neuronal à trois couches	112
Figure 5. La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond, au fil du temps.	114
Figure 6. Dévoilement des couches cachées de l'apprentissage profond	116
Figure 7. L'apprentissage par renforcement et sa mise en oeuvre	118
Figure 8. Composants des systèmes experts	120



Intelligence artificielle et développement humain



Résumé

Les applications de l'intelligence artificielle (IA) exerceront une incidence profonde sur les sociétés des pays à faible revenu et pays à revenu intermédiaire (PFR-PRI), tant positivement que négativement. L'intelligence artificielle est une nouvelle classe de technologies à partir de laquelle d'autres technologies et applications sont en train d'être conçues. Propulsées par la disponibilité accrue de la puissance de calcul, une connectivité améliorée, et les mégadonnées, les applications de l'IA offrent de fascinantes possibilités de promouvoir la croissance économique et de s'attaquer à un large éventail de problèmes de longue date dans les pays du Sud. L'incidence perturbatrice qui en résultera sera massive et pourrait bien être révolutionnaire.

La mise en place d'applications de l'IA dans les pays du Sud offre un potentiel énorme de faire le bien comme de nuire. L'intelligence artificielle rend possibles les innovations techniques novatrices, axées sur les données, pour aider à résoudre des problèmes sociaux urgents. L'intelligence artificielle peut favoriser les percées scientifiques, améliorer les diagnostics médicaux, accroître la productivité agricole, optimiser les chaînes d'approvisionnement et assurer un accès équitable à l'éducation grâce à un apprentissage hautement personnalisé. Toutefois, comme la plupart des nouvelles technologies, l'IA possède aussi le potentiel d'exacerber les problèmes existants et d'en créer de nouveaux. L'IA contribue aux problèmes sociaux, par exemple, en renforçant les inégalités structurelles et les préjugés, en perpétuant les inégalités entre les sexes, en menaçant des emplois et en introduisant d'autres risques actuellement inconnus et des conséquences inattendues.



Par conséquent, le déploiement de futures applications de l'IA exige une perspective critique saine et un dialogue public permanent. Des efforts concertés sont nécessaires pour assurer un développement et un déploiement plus équitable et inclusif de l'IA dans les PFR-PRI. Ces efforts devraient se concentrer sur l'habilitation de ces pays à tirer parti du potentiel de l'IA afin qu'ils puissent tout de même bénéficier de l'immense valeur que la technologie peut apporter. En même temps, il faudra faire preuve de prudence et de vigilance pour atténuer les risques connus et pour cerner et juguler les risques imprévus et les conséquences néfastes inattendues. Une telle approche est essentielle pour empêcher l'exacerbation des inégalités existantes et l'instabilité sociale.

Le présent livre blanc propose un programme de recherche proactif en vue de l'application éthique et équitable de l'IA dans les pays du Sud.

QU'EST-CE QUE L'IA?

L'IA est un domaine de l'informatique qui s'attache à la mise au point de systèmes auxquels on peut enseigner ou qui peuvent apprendre à prendre des décisions et à faire des prévisions dans des contextes précis. Les applications d'IA peuvent exécuter un large éventail de comportements intelligents : optimisation (p. ex., les chaînes d'approvisionnement); reconnaissance et détection de formes (p. ex., reconnaissance faciale); prévision et mise à l'essai d'hypothèses (p. ex., prévision d'éclosions de maladies); traitement du langage naturel; et traduction automatique.

Les technologies de l'IA sont sur le point d'avoir une incidence importante sur la société étant donné qu'elles tirent parti de l'infrastructure existante (Internet, grands ensembles de données) pour réduire radicalement les coûts des activités (à la fois les nouvelles et les anciennes, les bonnes et les mauvaises) sur une grande échelle.

AVANTAGES POTENTIELS DE L'IA POUR LES PAYS EN DÉVELOPPEMENT

Soins de santé: L'IA peut jouer un rôle crucial pour augmenter les capacités en soins de santé en comblant les lacunes sur le plan de l'expertise humaine, en améliorant la productivité et en renforçant la surveillance des maladies. Par exemple, une ONG brésilienne a établi un partenariat avec une jeune entreprise d'IA afin de mettre au point un système permettant de prévoir les incidences futures de maladies.

Prestation de services et de renseignements gouvernementaux: Des groupes du monde entier étudient des moyens de se servir de l'IA pour aider les pays à améliorer leurs efforts en matière de gouvernement en ligne en automatisant les évaluations complexes qui tiennent compte d'un éventail de facteurs techniques, organisationnels et sociaux. Par exemple, un système d'apprentissage automatique a été mis au point pour aider à prévoir l'emplacement des fosses communes des victimes des cartels de la drogue mexicains.

Agriculture: L'IA est employée pour contrer les diverses menaces qui peuvent compromettre une récolte réussie. Par exemple, des systèmes d'IA sont utilisés pour soutenir la gestion de l'eau en Palestine et la surveillance de la sécheresse en Ouganda.

Éducation: L'IA peut faire évoluer l'offre au-delà d'un modèle de prestation universel et industriel vers des possibilités d'apprentissage personnalisées de qualité, à grande échelle. Par exemple, en Inde, on utilise l'intelligence artificielle pour concevoir des systèmes de tutorat intelligents.

Économie et affaires: L'IA offre la possibilité d'accroître la productivité et offre un moyen de croissance grâce à la création d'entreprises, à l'innovation et à l'optimisation des composantes de base économiques. Par exemple, plusieurs entreprises s'appliquent à étendre l'accès aux services financiers de base aux centaines de millions d'Africains qui ne les utilisent pas ou n'y ont pas actuellement accès.

RISQUES POTENTIELS

Équité, biais et responsabilisation: Les systèmes d'IA sont susceptibles de refléter et d'exacerber les préjugés sociétaux et de produire des résultats qui peuvent désavantager des individus et des groupes, surtout ceux qui sont marginalisés. Par exemple, un programme informatique utilisé aux États-Unis pour évaluer les risques de récidive de personnes aux prises avec le système de justice criminelle a signalé que les prévenus noirs présentaient un risque élevé presque deux fois plus souvent que les prévenus blancs.

Surveillance et atteinte au droit à la vie privée: Des algorithmes d'IA survolent la capacité de surveillance et menacent la vie privée. Par exemple, un logiciel de reconnaissance faciale alimenté par l'IA confère aux systèmes



de télévision en circuit fermé la capacité de suivre des individus pendant leurs déplacements dans le paysage urbain. Ceci est préoccupant à la fois socialement et politiquement, puisque la vie privée est essentielle à d'autres droits fondamentaux tels que la liberté d'expression et d'association.

Automatisation, perte d'emplois et perte de recettes fiscales: Avec l'utilisation croissante de systèmes d'apprentissage automatique et d'intelligence artificielle dans pratiquement tous les secteurs de l'économie, l'automatisation à grande échelle dépassera le secteur de la fabrication et s'appliquera aux emplois hautement spécialisés fondés sur le savoir. Bon nombre de ces emplois peuvent être automatisés, en partie ou entièrement, ce qui réduit le besoin de faire appel à des travailleurs humains. Cependant, un argument contraire a également été avancé : l'IA pourrait entraîner un revirement dans la nature et la portée du travail et des emplois, par exemple en faisant en sorte que des robots complètent le travail humain, et en mettant davantage l'accent sur les tâches très spécialisées et mieux rémunérées.

Mise à mal de la démocratie et de l'autodétermination politique: Dans un monde de plus en plus connecté et tributaire de la libre circulation de l'information, la désinformation est une menace réelle et croissante pour la stabilité et la démocratie. Par exemple, en exploitant les données très personnelles recueillies dans le cadre de campagnes sur les médias sociaux, les applications d'IA favorisent des campagnes de propagande et de manipulation du comportement plus efficaces. L'élection présidentielle américaine de 2016 est devenue un exemple notoire du rôle de la désinformation ciblée sur Facebook.

L'AVENIR DE L'IA DANS LES PAYS DU SUD

Il ne fait aucun doute que les technologies d'IA seront d'ordre transformationnel. Des avancées spectaculaires seront réalisées, une richesse extraordinaire sera créée et bon nombre de nos structures sociales et institutionnelles seront transformées. Cependant, nous devons poser la question : des existences seront améliorées, en effet, mais pour qui au juste? Les libertés politiques et économiques connaîtront des avancées, mais pour qui ? Pour qui la privation systématique de ces libertés sera-t-elle adoucie? La conclusion de la présente communication est que, si nous continuons à aller aveuglément de l'avant, nous devons nous attendre à voir les inégalités s'accroître, de concert avec la perturbation économique, l'agitation sociale et, dans certains cas, l'instabilité politique, et ce sont les gens défavorisés et sous-représentés sur le plan technologique qui subiront les pires effets.

Cette sombre prévision tire son origine de l'entrelacement de deux éléments: la nature des applications d'IA, et les prévisions des incidences des applications d'IA dans le contexte mondial actuel. Ce qui est préoccupant, c'est la dynamique avec laquelle nos ensembles actuels d'institutions et de cultures façonnent l'évolution des technologies, et comment, à leur tour, ces technologies façonnent ces institutions et ces cultures. Un élément important de ce contexte est caractérisé par ce que l'on peut appeler le « fossé de l'IA » — c'est-à-dire, un écart entre ceux qui ont la capacité de concevoir et de déployer des applications d'IA, et ceux qui ne l'ont pas. En outre, les assises tant numériques (p. ex., l'infrastructure) qu'analogiques (p. ex., la réglementation) requises en vue d'une application éthique et équitable des technologies d'IA dans de nombreux pays du Sud sont largement absentes, et une asymétrie marquée en matière de pouvoir persiste.

Cette situation rend d'autant plus nécessaire de se prémunir contre les défis posés par l'intelligence artificielle, afin que nous puissions éviter ou atténuer ces résultats négatifs en habilitant les pays en développement à tirer pleinement parti du potentiel positif de l'IA. L'IA offre des promesses si alléchantes et un potentiel si prometteur qu'on ne pourra éviter d'en tenir compte à l'avenir. La question est de savoir si nous serons prêts à l'accueillir.



RECOMMANDATIONS

En nous fondant sur les conclusions du présent document et de la documentation consultée, nous avons cerné trois principaux domaines dans lesquels des mesures peuvent être prises : politiques et réglementations, applications inclusives et éthiques de l'IA et infrastructure et compétences. Dans chaque domaine, nous formulons une série de recommandations sur des recherches qui sont nécessaires pour obtenir des progrès concrets. Veuillez prendre note que cette liste n'est pas exhaustive, et qu'elle vise plutôt à mettre en lumière les interventions les plus pressantes.

STRUCTURES POLITIQUES ET RÉGLEMENTAIRES

FAVORISER LA CONCEPTION DE POLITIQUES ET DE RÈGLEMENTS QUI HABILITENT UNE IA INCLUSIVE ET AXÉE SUR LES DROITS

Mener des recherches de base sur la prévalence des applications d'IA et de politiques dans les pays du Sud. Malgré des poches d'activités liées à l'IA dans les pays du Sud, il n'existe aucune synthèse systématique du niveau de cette activité. La collecte de données de base doit inclure les ensembles de politiques, de règlements, d'applications, d'ensembles de données ouvertes existants et de niveaux de compétences en matière d'intelligence artificielle. Cette recherche doit être effectuée chaque année, ou à tout le moins, tous les deux ans, afin d'appuyer les activités continues, l'élaboration de politiques et le programme de recherche.

Se renseigner au sujet de modèles réglementaires efficaces. Documenter et évaluer les modèles réglementaires de l'intelligence artificielle élaborés pour composer avec l'émergence de nouvelles activités qui utilisent l'IA comme les services de police prédictifs, les véhicules autonomes et les robots conversationnels. Déterminer si les risques potentiels des applications d'IA sont pris en compte adéquatement dans la réglementation existante, ou si la réglementation existante doit être adaptée ou encore si une nouvelle réglementation doit être élaborée. Déterminer les réponses réglementaires à des cas d'utilisation de l'intelligence artificielle donnés, ainsi que les niveaux de risque qui sont appropriés pour les contextes associés à une capacité institutionnelle réduite. Même si les leçons apprises dans les pays du Nord sont utiles, il est essentiel de ne pas importer directement les approches institutionnelles et réglementaires dans les contextes du Sud où les formes institutionnelles et culturelles sont différentes.

Surveiller l'incidence de l'IA sur l'emploi et le travail. Mener des recherches sur la politique sociale et économique afin de comprendre les effets de l'IA sur l'emploi, la nature du travail et les marchés du travail. Dans quelle mesure l'automatisation grâce à l'IA modifie-t-elle les modèles d'emplois et transforme-t-elle le milieu du travail ? Quels sont les autres modèles de revenus et de distribution des ressources, d'éducation et de perfectionnement professionnel professionnel dans les différents contextes ?

Examiner les approches pour composer avec la responsabilité, et les mécanismes de réparation eu égard à la prise de décisions fondées sur l'IA. Concevoir des systèmes et des cadres réglementaires qui déterminent la responsabilité eu égard à la prise de décisions fondées sur l'IA qui sont erronées, biaisées ou discriminatoires, et établir des mécanismes de réparation. Ces mesures peuvent comprendre notamment des politiques qui stipulent la transparence eu égard au processus décisionnel automatisé, des procédures d'évaluation en vue de déterminer la compétence des systèmes d'IA, et la certification des systèmes d'IA qui exécutent des tâches nécessitant un certain degré de compétence ou de formation. La nécessité d'agir est particulièrement urgente dans le cas de systèmes de décision qui influent sur le bien-être ou la liberté des personnes, comme ceux qui font appel à l'usage de la force ou à l'incarcération. La recherche est essentielle dans ce cas pour découvrir et documenter quels sont les systèmes efficaces, et dans quels contextes, en matière de responsabilité et de réparation.

Étudier l'incidence de l'IA sur les droits de la personne. De manière générale, les Nations Unies reconnaissent que les droits qui s'appliquent hors ligne doivent s'appliquer en ligne, témoignant de la pertinence des droits analogiques dans des environnements régis numériquement. Des organismes professionnels demandent spécifiquement que les droits de la personne soient pris en considération dans le contexte de la conception et de l'exploitation de l'IA. Calquer les évaluations d'impact sur les risques liés à l'IA contribuerait à encourager les programmes de développement à intégrer la technologie d'IA de manière à respecter et à promouvoir les droits de la personne, y compris le droit à la vie privée, à l'égalité et à la liberté d'expression.



APPLICATIONS

CATALYSER LE DÉVELOPPEMENT D'APPLICATIONS D'IA INCLUSIVES ET ÉTHIQUES

Appuyer le développement et le déploiement d'applications d'IA

innovatrices au service du bien social. Investir dans le développement, le déploiement et l'utilisation d'applications pour l'éducation, la santé, l'environnement, la sécurité alimentaire, etc., et faire en sorte que ces applications soient éthiques et inclusives. Comme pour les innovations en matière de réglementation, même s'il est important de s'inspirer d'exemples de partout dans le monde, les applications d'IA nécessiteront souvent des solutions locales pour être efficaces.

Faire de la recherche sur les répercussions sociales des innovations utilisant l'IA.

Des recherches sont nécessaires afin de mieux comprendre quelles sont les applications d'IA qui fonctionnent (ou celles qui ne fonctionnent pas), pour qui, et dans quels contextes. Nous devons savoir qui profite des applications d'IA, et comment, ainsi que qui est laissé pour compte ou victime de ces applications. On insistera plus particulièrement sur l'examen des répercussions différentielles sur divers groupes, surtout lorsque les différences sont imputables au genre, au statut social et économique, à la race, etc. Cette recherche doit aller au-delà des effets de premier ordre, comme l'amélioration des gains d'efficacité ou de l'exactitude des diagnostics, pour inclure des effets sociaux plus larges. Il est possible que l'on doive faire appel à de nouvelles méthodologies d'évaluation et d'évaluation d'impact.

Mettre à l'essai et surveiller les biais dans les applications d'IA.

Les systèmes d'intelligence artificielle qui prennent des décisions ou qui informent la prise de décisions qui ont une incidence sur le bien-être des humains (par exemple diagnostics médicaux, fournir à un juge une évaluation du potentiel de récidive) devraient être mis à l'essai et surveillés en vue d'y repérer des biais et des erreurs dans les différents contextes et collectivités, avant la diffusion et continuellement par la suite.

Examiner les modèles de conception participative pour l'IA.

Mener des recherches sur les pratiques qui appuient le développement d'applications d'IA inclusives. Quelles sont les techniques efficaces pour obtenir des processus véritablement participatifs qui mobilisent des populations diverses dans la conception et le déploiement d'applications d'IA ? Comment et dans quels contextes ces pratiques parviennent-elles à contrer les biais de conception et les biais appris et à faire en sorte que l'IA soit pertinente pour les collectivités

marginalisées ? Les intervenants de l'intelligence artificielle sur le terrain devraient publier des données sur la diversité de la participation à la conception et au développement.

Une recherche-action pour mieux comprendre comment mettre à l'échelle efficacement et équitablement des applications d'IA éprouvées.

Démontrer la validité du concept est différent de diffuser l'application dans toute une population donnée tout en maintenant la qualité et l'équité des avantages. La recherche sur le processus de mise à l'échelle des applications d'IA, tant à la verticale pour englober les fonctionnalités additionnelles, qu'à l'horizontale pour s'étendre dans de nouveaux endroits, est essentielle pour propager les avantages de ces applications. La généralisation dans divers contextes nécessite la compréhension des raisons pour lesquelles une application a bien fonctionné dans un certain contexte et une appréciation du fait que l'application pourrait avoir besoin de modifications pour réussir dans un nouvel environnement. Nous devons échafauder des théories sous-jacentes à la mise en oeuvre des applications d'IA afin de reproduire de façon fiable la réussite au sein de populations différentes. Dans ce contexte, il faudrait mettre l'accent sur les variations entre les contextes plutôt que s'en tenir strictement aux modalités de l'une ou l'autre des mises en oeuvre. Les difficultés particulières liées aux données comprendront notamment la mise à l'échelle au-delà de la portée des ensembles de données existants et l'élaboration de moyens pour produire rapidement de nouveaux ensembles de données.

INFRASTRUCTURE ET COMPÉTENCES

CONSTRUIRE L'INFRASTRUCTURE ET LES COMPÉTENCES EN VUE D'UNE IA INCLUSIVE ET ÉTHIQUE

Appuyer des programmes destinés à renforcer l'expertise en matière d'IA dans les gouvernements. Promouvoir l'expertise en matière d'IA dans toutes les branches et à tous les niveaux des gouvernements, y compris les organismes réglementaires et, éventuellement, les nouveaux organismes consultatifs.

Favoriser les capacités locales à la tête de la conception, du développement et du déploiement des applications d'IA. Ces activités pourraient comprendre notamment : soutenir la croissance de centres d'excellence multidisciplinaires en intelligence artificielle dans les pays du Sud en vue de s'engager dans le développement et la recherche à l'échelle locale, et de fournir des données probantes pour contribuer à l'élaboration de la politique nationale et aux décisions réglementaires; établir des liens entre les experts techniques et les

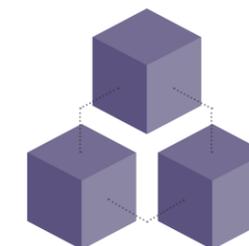


communautés marginalisées et à faible revenu dans les pays du Sud; et soutenir les collaborations Sud–Sud.

Élaborer et mettre à l'épreuve des approches rentables pour renforcer des compétences pertinentes en matière d'IA, plus particulièrement chez les femmes et dans les populations marginalisées. Élaborer et soutenir des programmes qui se concentrent sur le développement des capacités des femmes et d'autres populations marginalisées en vue de les engager à différentes étapes du développement et de l'application des technologies d'IA. La recherche devrait stimuler cette activité au moyen de l'examen de modèles à faibles coûts pour perfectionner les compétences techniques en matière d'IA ainsi que ainsi que mettre à l'essai mettre à l'essai des programmes et des pédagogies efficaces.

Élargir l'accès aux données et aux ressources informatiques. Dans la mesure du possible, la recherche, les outils et les ensembles de données de formation liés à l'IA doivent être fournis gratuitement. Appuyer le développement et le partage d'ensembles de données diversifiées et inclusives qui sont nécessaires pour les applications d'IA dans différents contextes.

Étudier les avantages et les risques liés à l'intelligence artificielle ouverte. Mener des recherches sur les risques et les avantages à court et à moyen terme de l'ouverture dans le domaine de l'IA (c'est-à-dire le partage des ressources d'IA, des ensembles de données, etc). Si possible, cette recherche devrait établir un lien entre les questions du côté de l'approvisionnement (quel serait le meilleur moyen d'offrir un accès ouvert aux algorithmes d'IA, aux outils et aux ensembles de données) en approfondissant la compréhension de l'engagement nécessaire pour s'assurer que les ressources d'IA ouvertes sont disponibles en vue d'une utilisation ou d'une réutilisation et d'une adaptation par diverses populations (et pas seulement par ceux qui sont déjà compétents et dotés de ressources). Il conviendrait d'accorder une attention particulière à la question de l'équilibre entre le partage d'ensembles de données et la protection de la vie privée.



Intelligence artificielle et développement humain

Vers un programme de recherche



L'intelligence artificielle exercera une incidence profonde dans les pays dits en développement, tant positivement que négativement, au cours des prochaines décennies

Introduction

Il ne fait aucun doute que les applications de l'intelligence artificielle (IA) exerceront une incidence profonde sur les sociétés des pays à faible revenu et pays à revenu intermédiaire (PFR-PRI), tant positivement que négativement (Stone et coll., 2016). L'intelligence artificielle est une nouvelle classe de technologies à partir de laquelle d'autres technologies et applications seront construites.¹ Propulsées par la disponibilité accrue de la puissance de calcul, une connectivité améliorée, et les mégadonnées, les applications de l'IA offrent de fascinantes possibilités de promouvoir la croissance économique et de s'attaquer à un large éventail de problèmes de longue date dans les pays du Sud. L'incidence perturbatrice qui en résultera sera massive et pourrait bien être révolutionnaire.

La promesse de l'IA et son potentiel pour nous aider à atteindre les objectifs de développement durable (ODD)² suscite énormément d'intérêt dans la communauté du développement. Par exemple, le Sommet sur L'intelligence artificielle au service du bien social, organisé en 2017 par l'Union internationale des télécommunications, a examiné les nombreux moyens que pourrait prendre l'IA pour contribuer au développement durable grâce à un large éventail d'applications, tels que³:

Photo par Igor Ovsyannykov



Applications de l'intelligence artificielle

Cartographier la pauvreté à partir de l'espace pour permettre une allocation des ressources en temps réel.

Accroître la productivité agricole au moyen de l'automatisation et de l'analyse prédictive.

Analyser les données en soins de santé afin de faciliter les percées scientifiques.

Révolutionner les salles de classe en offrant des parcours d'apprentissage individualisés.

Encourager les méthodes d'embauche équilibrées et braquer l'attention sur les inégalités entre les sexes.

Utiliser des capteurs pour prévoir les schémas de consommation en vue d'assurer un approvisionnement en eau efficace et sûr.

Améliorer la capture de l'énergie photovoltaïque, et ce faisant, abaisser le coût de l'énergie solaire.

Accroître la productivité en vue de la croissance économique grâce à l'automatisation, et permettre une utilisation plus efficace des ressources au moyen de chaînes d'approvisionnement, de parcours logistiques et d'un planification.

Bâtir une société plus égalitaire et plus inclusive au moyen de la robotique au service du handicap.

Soutenir l'urbanisme afin de rendre les villes plus intelligentes et plus durables.

Réduire les déchets en prévoyant et en déterminant les niveaux de production optimaux.

Prévoir les catastrophes climatiques au moyen de modélisations climatiques améliorées.

Combattre la pêche illicite en surveillant les mouvements des bateaux de pêche.

L'engouement à leur égard est comparable à celui que l'on avait ressenti pour d'autres technologies prometteuses comme Internet et les chaînes de blocs. On espère que ces applications de l'IA permettront d'adopter des approches radicalement plus efficaces et efficientes dans la promotion du développement, lesquelles pourront, à leur tour, nous permettre d'atteindre les cibles ambitieuses des ODD. En résumé, l'IA est vue comme renfermant un énorme potentiel de créations de solutions techniques novatrices, généralement axées sur les données, à des problèmes sociaux complexes et pressants.

Toutefois, comme la plupart des nouvelles technologies, l'IA possède aussi le potentiel d'exacerber les problèmes existants et d'en créer de nouveaux. L'IA contribue aux problèmes sociaux, par exemple, en renforçant les inégalités structurelles et les préjugés, en perpétuant les inégalités entre les sexes, en menaçant des emplois et en introduisant d'autres risques actuellement inconnus et des conséquences inattendues. Des exemples récents très médiatisés d'algorithmes de l'IA ayant mal tourné illustrent certains de ces risques: un algorithme de Google ayant étiqueté deux personnes noires sur une photo comme des gorilles⁴, un permis de conduire révoqué lorsqu'un algorithme de reconnaissance faciale a signalé une photo comme étant probablement frauduleuse⁵, des algorithmes prédictifs de la récidive aiguillant des décisions déterminantes par des juges fondés sur des données fortement empreintes de préjugés⁶, et une informatisation fondée sur l'IA exerçant une incidence négative sur les emplois dans certains secteurs et en menaçant d'autres⁷. Toutefois, les risques posés par l'IA ne proviennent pas seulement d'algorithmes biaisés. La technologie peut aussi être mise à contribution pour promouvoir des résultats sociaux « négatifs » comme saper la gouvernance démocratique et permettre des activités contraires à l'éthique et criminelles. Comme l'a écrit James Hendler, « L'IA peut se mettre au service du bien social. Mais elle peut aussi entraîner d'autres types de répercussions sociales qui illustrent bien l'adage, ce qui est une nourriture pour l'un, est un poison pour l'autre. Il faut en tenir compte. »⁸

Effectivement, les potentiels parallèles de l'IA de promouvoir d'extraordinaires réalisations pour le bien social, et d'entraîner des dommages significatifs sont remarquables. Du point de vue du développement international, il est clair qu'il y a et qu'il continuera d'y avoir de nombreuses applications déterminantes de l'IA qui peuvent être déployées avec beaucoup de succès dans les pays du Sud. Cependant, il est aussi clair que les risques et les dommages potentiels à court et à moyen terme sont significatifs et justifient qu'on y porte une attention sérieuse. De plus, les risques pour les pays du Sud sont amplifiés dans



Photo par Clem Onojeghuo sur Unsplash

les contextes de revenu faible et de revenu intermédiaire. Cette perspective repose sur la compréhension que les effets sociaux des technologies sont façonnés par le contexte institutionnel, politique et économique dans lequel elles sont déployées. Malheureusement, il existe des facteurs fondamentaux généralement présents dans les PFR-PRI qui contribueront à rendre la prise en compte des défis et des risques liés à l'IA encore plus difficile. Voici ces facteurs:

- ▶ Une distribution très inégale des ressources et des connaissances requises pour la mise en oeuvre des techniques d'IA (y compris de très larges écarts sur le plan ethnique et de la sexospécificité), et de ce fait, une capacité inégale de prendre des décisions au sujet des applications qui seront mises au point et pour qui;
- ▶ Un large fossé numérique entre les pays du Nord et les pays du Sud, de même qu'un fossé entre les milieux urbains et ruraux dans les pays du Sud;
- ▶ Des niveaux élevés d'instabilité politique et sociale (dans certains pays);
- ▶ Des taux élevés de chômage et une faible assiette fiscale qui limitent les mesures prises par les gouvernements pour réagir à la perte d'emplois entraînée par l'automatisation;
- ▶ Une faible capacité institutionnelle de protéger les droits individuels, comme la vie privée.

Dans ce contexte, les risques qui découlent d'applications de l'IA pourraient s'avérer particulièrement nocifs dans les pays affichant des taux élevés de chômage et d'instabilité politique. Compte tenu des facteurs en jeu, nous pouvons anticiper les résultats généraux suivants:

- ▶ Une répartition inégale des avantages, qui penche lourdement en faveur de ceux qui sont déjà bien nantis;
- ▶ La persistance d'un fossé lié aux compétences et aux ressources en matière d'IA qui limite la « démocratisation » des avantages dérivés de l'IA et l'inclusion de voix diversifiées et habituellement exclues des décisions concernant la conception, le développement et le déploiement des applications de l'IA;
- ▶ La perpétuation et même l'exacerbation de la marginalisation sociale par le truchement de l'automatisation de préjugés non reconnus;
- ▶ Une expansion de la surveillance de la vie privée et des menaces à cet égard, avec des intervenants mauvais ou voyous exerçant leurs activités avec un savoir-faire grandissant visant à favoriser la discorde sociale et l'agitation politique.

Ces effets nocifs qui s'autorenforcent, et qui touchent de façon disproportionnée les populations marginalisées et économiquement défavorisées. (Nous fournissons davantage de détails sur chacun de ces risques à la section 4.2 ci-après.)



Photo par Daniel Von Appen sur Unsplash

Par conséquent, le déploiement de futures applications de l'IA exige une perspective critique saine et un dialogue public permanent. Des efforts concertés sont nécessaires pour assurer un développement et un déploiement plus équitable et inclusif de l'IA dans les PFR-PRI. Ces efforts devraient se concentrer sur l'habilitation de ces pays à tirer parti du potentiel de l'IA afin qu'ils puissent tout de même bénéficier de l'immense valeur que la technologie peut apporter. En même temps, il faudra faire preuve de prudence et de vigilance pour atténuer les risques connus et pour cerner et juguler les risques imprévus et les conséquences néfastes inattendues. Une telle approche est essentielle pour empêcher l'exacerbation des inégalités existantes et l'instabilité sociale.

En dépit de ces défis, nous pouvons entrevoir un monde où les innovations liées à l'IA sont mises en oeuvre de manière éthique et produisent des avantages équitables qui font progresser la condition sociale. L'IA peut même être un moyen d'encourager une compréhension sociale plus profonde et de surmonter des préjugés sociaux et culturels de longue date. La question qui se pose est la suivante : comment nous y prendrons-nous?

Un ancrage solide dans la recherche sera incontournable pour répondre à cette question. Actuellement, cependant, il existe de précieuses petites recherches empiriques pour guider la conception, le développement et le déploiement de l'IA dans les PFR-PRI, ou pour éclairer les mesures réglementaires et stratégiques nécessaires. Et les questions ouvertes abondent : Comment des sociétés pauvres peuvent-elles tirer parti d'applications de l'IA pour relever les principaux défis du développement ? Comment devraient-elles s'y prendre pour s'attaquer aux enjeux compliqués sur le plan moral, légal et de la reddition de comptes que présente



l'IA ? Dans quelle mesure l'automatisation formée par l'IA perturbera-t-elle les économies et entraînera-t-elle l'élimination potentielle de millions d'emplois ? Quelles sont les mesures décisionnelles judicieuses et réalisables?

En conséquence, actuellement, la réflexion sur les répercussions sociales de l'IA sont purement spéculatives; il reste beaucoup d'inconnues, et les opinions peuvent diverger grandement. Étant donné que « les décisions [relatives à l'IA] en matière de conception et de politiques sont susceptibles d'entraîner des répercussions durables sur la nature et les orientations de ces développements » [Traduction] (Stone et coll., 2016: 5), il est impératif de prendre des mesures pour combler ces lacunes dès que possible.

Le présent document vise à présenter un programme proactif en vue de l'application éthique et équitable de l'IA pour le développement, en mettant l'accent sur le rôle de la recherche. Pour servir d'assise à ce programme, nous présentons un large aperçu des technologies associées à l'IA et des possibilités et des défis qu'elles présentent.

Dans la section suivante, nous offrons une définition de travail de l'IA et abordons les éléments de base des algorithmes de l'IA qui ont connu des progrès rapides au cours de la dernière décennie. Ensuite, nous décrivons les avantages et les risques potentiels de l'IA dans le contexte des pays du Sud. Nous concluons par un survol de recommandations visant à faire avancer le développement, le déploiement et la mise en oeuvre éthiques et inclusifs de l'IA dans ces contextes.

Photo par Ferran Fusalba Rosello sur Unsplash





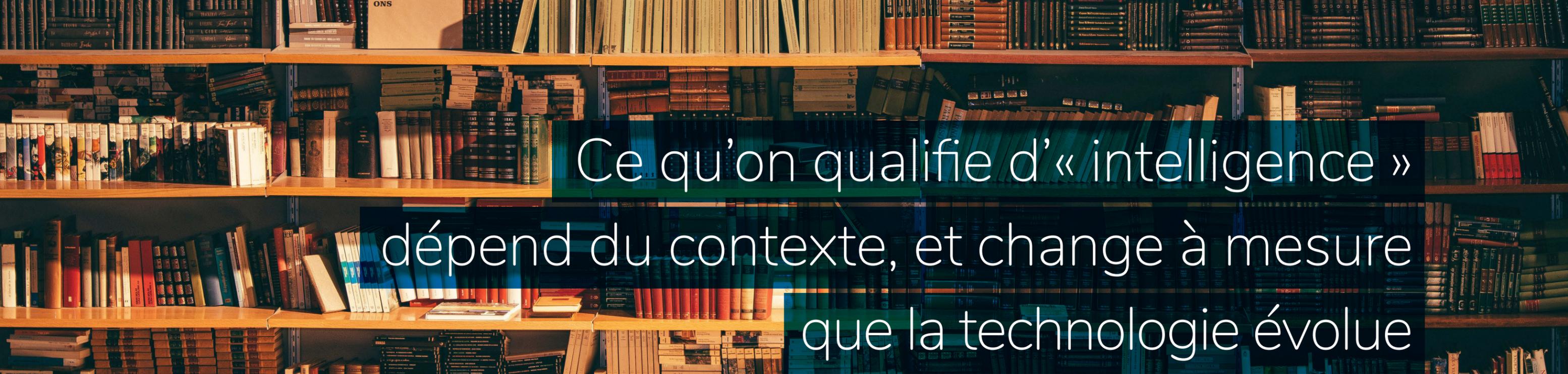
L'IA n'appartient
plus à un avenir
fictif quelconque

Qu'est-ce que l'IA?

En dehors du domaine de l'informatique et du développement de logiciels, le concept d'IA est connu de la majorité des gens comme un thème de la science-fiction — souvent associé à des répercussions menaçantes ou apocalyptiques. Mais l'IA n'appartient plus à un avenir fictif quelconque: elle fait déjà partie de notre monde et elle est appelée à exercer rapidement une incidence croissante sur la vie de tous les jours.

Pour pouvoir examiner d'un oeil critique les avantages et les risques associés à l'IA dans le contexte du développement et fournir des solutions pour en atténuer les risques, nous devons avoir une compréhension claire de ce qu'elle est.

De manière générale, l'IA est un domaine de l'informatique qui s'attache à la mise au point de systèmes auxquels on peut enseigner (p. ex. par l'entremise des connaissances de l'expert en codage) ou qui peuvent apprendre (à partir de données) à prendre des décisions ou à faire des prévisions dans des contextes précis. Même si le domaine existe depuis les années 1950, l'émergence des « mégadonnées » (grands ensembles de données rendus possibles grâce à l'adoption répandue d'Internet, des téléphones cellulaires, et des médias sociaux) et de récents progrès dans les techniques d'apprentissage automatique (AA), jumelés aux progrès réalisés dans les principales technologies habilitantes comme la robotique et la détection, ont propulsé une explosion d'intérêt pour l'IA et ses applications dans le monde réel.



Ce qu'on qualifie d'« intelligence » dépend du contexte, et change à mesure que la technologie évolue

Photo par Alfons Morales sur Unsplash

Définitions de l'intelligence artificielle

L'expression « intelligence artificielle » (IA) a été utilisée pour la première fois en 1955 dans une proposition d'étude sur l'utilisation d'ordinateurs pour « résoudre le genre de problèmes qui sont actuellement réservés aux humains⁹ ».

Cette activité consacrée à rendre les machines intelligentes, et l'intelligence est cette qualité qui permet à une entité de fonctionner de manière appropriée et avec clairvoyance dans son environnement¹⁰.

Le domaine qui étudie la synthèse et l'analyse d'agents virtuels qui agissent de façon intelligente¹¹.

Une constellation de technologies, y compris l'apprentissage automatique, la perception, le raisonnement et le traitement du langage naturel¹².

Un terme collectif pour décrire des systèmes informatiques capables de percevoir leur environnement, de penser, d'apprendre et d'accomplir des comportements en fonction de ce qu'ils ont perçu et de leurs objectifs¹³.

Construire des machines qui sont intelligentes, qui peuvent faire des choses que les humains font, et pour y arriver, elles doivent posséder des connaissances sur le monde, et être capables d'utiliser ces connaissances pour faire des choses utiles¹⁴.

L'une des façons de penser l'IA consiste à la voir comme une forme avancée d'analyse de données selon laquelle un algorithme logiciel peut tirer des conclusions en se fondant sur les résultats de l'analyse, puis agir en fonction de ces conclusions, ce qui revient à se comporter de façon « intelligente ». Les algorithmes de l'IA sont souvent utilisés lorsque nous « sommes incapables de comprendre un problème suffisamment bien pour pouvoir le traduire dans un algorithme [standard] » (Privacy International, 2016).

La collecte de données sur leurs grands parcs d'utilisateurs en ligne a transformé de nombreuses sociétés comme Google, Facebook et Amazon en chefs de file des mégadonnées. La disponibilité de ces ensembles massifs de données — et l'argent qui peut être réalisé en les explorant — représente un point de basculement qui a propulsé la croissance de l'IA.

Dans les années qui viennent, il faut s'attendre à un autre tsunami de données issues de l'Internet des objets. À mesure que des objets comme les automobiles jusqu'aux réfrigérateurs, en passant par les stimulateurs cardiaques deviendront connectés à des systèmes en ligne, des industries complètes se retrouveront assises sur d'énormes ensembles de données prêtes à être analysées – et exploitées. Les sociétés d'investissement en capital de risque dans l'IA connaissent une croissance très rapide depuis 2013, et la tendance semble vouloir se maintenir¹⁵. La « révolution des mégadonnées » est en passe de devenir la révolution de l'IA.



Comportement intelligent	Optimisation
Application	Chaînes d'approvisionnement, parcours logistiques, établissement de prix
Comportement intelligent	Reconnaissance/détection de formes
Application	Reconnaissance faciale, diagnostics médicaux, détection de la fraude
Comportement intelligent	Prévision/mise à l'essai d'hypothèses
Application	Prévision d'inondation, prévision de catastrophe, éclosion de maladie, récidivisme
Comportement intelligent	Traitement du langage naturel ¹⁶
Application	Reconnaissance vocale
Comportement intelligent	Traduction automatique ¹⁷
Application	Traduction de textes ou de paroles, à partir d'une langue dans une autre

Tableau 2. Exemples de comportements de l'intelligence artificielle.

Il est utile de penser à l'IA en deux dimensions : ce qu'elle peut faire (comportements) et comment elle le fait (techniques). Les applications de l'IA peuvent exécuter un large éventail de comportements, qu'il s'agisse de jouer aux échecs ou de reconnaître la voix pour conduire des véhicules autonomes. Le tableau 2 fournit des exemples de comportements que l'on associe généralement au domaine de l'IA. Du point de vue de l'IA pour le développement, comprendre ce que fait l'IA (les comportements) et comment ces comportements peuvent être appliqués est essentiel pour comprendre toute la gamme de ce qui est possible.¹⁸

Ces comportements peuvent aussi être combinés en vue d'obtenir des capacités intelligentes encore plus complexes. À mesure que des comportements particuliers sont de plus en plus regroupés sous l'étiquette de « compétences » ou offerts à titre de services dans le nuage, le développement et le déploiement de ces comportements en fonction de différents contextes devient de plus en plus disponible. À mesure que divers comportements sont combinés, la diversité et la sophistication des applications de l'IA s'accroissent.

Par exemple, un assistant au service à la clientèle virtuel doit posséder des capacités de reconnaissance vocale et de traitement du langage naturel, en plus d'une représentation des connaissances spécialisées de l'entreprise. Des assistants virtuels, comme Alexa, d'Amazon, Siri, d'Apple et Google Home sont de plus en plus répandus, et à mesure que leurs compétences s'améliorent, nous nous attendons à ce qu'ils soient déployés de plus en plus dans différents secteurs.

Les véhicules autonomes eux aussi réunissent de nombreuses techniques de collecte de données et d'IA. Le véhicule autonome respecte les règles de la circulation et évite les obstacles grâce à une combinaison de règles en codage dur, d'algorithmes relatifs à l'évitement des obstacles, de modélisations prédictives et de discrimination d'objet « intelligente » (p. ex., connaître la différence entre une bicyclette et une motocyclette). Le système réagit à son environnement immédiat en se servant de données fournies par un réseau de lasers et d'autres capteurs, et atteint l'objectif principal consistant à atteindre sa destination à l'aide de données fournies par des systèmes de navigation comme Google Maps¹⁹.

Cependant, pour comprendre comment ces comportements sont obtenus, il faut examiner les diverses techniques d'IA sur lesquelles ils reposent. Les plus importantes de ces techniques sont celles des algorithmes d'AA et des systèmes experts. Nous incluons un bref aperçu de ces techniques à titre de base pour reconnaître leurs limitations et risques potentiels, ainsi que pour



savoir comment les atténuer. (Pour obtenir plus de précisions, voir l'encadré: «Comment l'apprentissage automatique fonctionne-t-il?», et l'annexe: Comment l'IA fonctionne-t-elle?)

Les techniques d'apprentissage automatique, comme l'apprentissage profond et l'apprentissage par renforcement, habilitent les applications de l'IA à apprendre en absorbant les données se trouvant dans les ensembles de données existants ou au moyen de la rétroaction provenant de l'interaction avec leur environnement. S'il est vrai que la disponibilité des mégadonnées a permis à ces algorithmes de s'attaquer à des problèmes d'une grande complexité, leur dépendance à l'égard des données a ses limites. Premièrement, ces algorithmes privilégient certains types de problèmes et de connaissances par rapport aux autres. Toutes les choses ne peuvent être réduites à des chiffres et encodées numériquement. De ce fait, les problèmes qui se prêtent facilement à la représentation sous la forme d'ensembles de données ou qui comportent des résultats souhaités clairs sont plus faciles à maîtriser par l'apprentissage automatique (AA). C'est l'une des raisons pour lesquelles l'IA a progressé si rapidement dans certains domaines (par exemple jouer à des jeux et reconnaître des images) mais pas dans d'autres (par exemple comprendre des concepts de haut niveau)²⁰.

Deuxièmement, les techniques d'AA qui apprennent en repérant des modèles dans les ensembles de données, comme l'apprentissage profond, comportent des risques inhérents plus élevés de biais et de menaces pour la vie privée. Si un

ensemble de données d'apprentissage contient des données biaisées, l'algorithme apprendra ce biais. Si l'algorithme s'alimente à partir de processus décisionnels entraînant des conséquences, il peut renforcer ou même exacerber les inégalités sociales existantes (voir Secteurs de risque, p. 63). De plus, si un algorithme est conçu pour extraire des renseignements personnels de certaines données, il peut compromettre la vie privée des personnes dont les données figurent dans cet ensemble de données. Selon le contexte, cette atteinte au droit à la vie privée risque de miner sérieusement les droits fondamentaux de la personne (voir Surveillance et atteinte au droit à la vie privée, p. 75). Pour en apprendre davantage sur les limites de l'apprentissage profond, voir Marcus (2018).

Contrairement aux algorithmes d'AA, les systèmes experts tentent d'émuler les aptitudes à la résolution de problèmes d'un expert humain en utilisant des connaissances encodées explicitement et des procédures de déduction pour résoudre des problèmes. De ce fait, plutôt que de travailler avec des ensembles de données existants, les systèmes experts opèrent dans des situations où il existe déjà un corpus d'expertise explicite sur la manière d'effectuer une tâche ou de résoudre un problème particulier. Les systèmes experts comptent parmi les techniques d'IA initiales mises au point avant les mégadonnées, à une époque où la puissance de calcul était limitée.

COMMENT L'APPRENTISSAGE AUTOMATIQUE FONCTIONNE-T-IL?

Il existe trois types généraux d'algorithmes d'apprentissage automatique : apprentissage supervisé, apprentissage non supervisé et apprentissage par renforcement.

Les algorithmes d'apprentissage supervisé sont alimentés successivement par des données sur des exemples; pour chaque

exemple, l'algorithme fournit une réponse (un résultat ou une prévision). L'algorithme apprend en rajustant ses paramètres internes de manière à ce que sa réponse à cet exemple particulier soit plus précise la fois suivante.

Les algorithmes d'apprentissage non supervisé sont alimentés par des ensembles de données

sans résultats correspondants; l'algorithme extrait les modèles (classes) de données. L'apprentissage non supervisé est particulièrement utile pour faire l'extraction de perspectives à partir de données.

L'apprentissage par renforcement (AR) entraîne les algorithmes au moyen d'une rétroaction positive ou

négative fondée sur une action prise dans un environnement particulier. L'AR est approprié pour les tâches axées sur un but, comme apprendre un jeu (par exemple les échecs ou le jeu de go) ou pour enseigner à un robot une compétence (par exemple comment manipuler un objet).

Voir aussi l'annexe : Comment l'IA fonctionne-t-elle?



Avantages pour le développement

EXEMPLES CHOISIS

Même si les voitures autonomes et les logiciels de reconnaissance vocale comptent parmi les applications les plus conviviales de l'IA, la technologie a été appliquée dans de nombreux domaines différents. Toutefois, la majorité des applications de l'IA ont été mises en oeuvre dans les pays du Nord, où le contexte est plus favorable à ce genre d'applications. C'est pourquoi, même si l'IA pour le développement suscite beaucoup d'intérêt, il existe peu d'applications à l'heure actuelle dont il est possible de tirer des enseignements.

Cette section présente des exemples dans quelques domaines de la manière dont l'IA peut contribuer au développement social, politique et économique. Elle ne vise pas à entrer dans les détails des applications possibles de l'IA dans les pays du Sud; elle a plutôt pour but de fournir des descriptions de certaines applications réelles de l'IA pour donner un aperçu de l'état actuel du développement et du déploiement de l'IA dans ces contextes.

Photo par Alain Pham



L'intelligence artificielle peut combler les lacunes sur le plan de l'expertise humaine en améliorant la productivité des travailleurs de la santé disponibles

SOINS DE SANTÉ

Une plus faible espérance de vie et des ressources insuffisantes en soins de santé constituent des sujets de préoccupation constants dans les pays du Sud, où bon nombre de conditions évitables ne sont pas diagnostiquées et où les épidémies de maladies infectieuses débordent fréquemment l'infrastructure disponible. L'IA peut jouer un rôle crucial en augmentant les capacités en comblant les lacunes sur le plan de l'expertise humaine, en améliorant la productivité des travailleurs de la santé disponibles et en améliorant la surveillance des maladies (Quinn et coll., 2014).

DIAGNOSTICS AU POINT DE SERVICE

Le paludisme touche plus de 200 millions de personnes chaque année, et elle entraîne le décès de centaines de milliers d'entre elles. Le meilleur moyen de diagnostiquer la maladie consiste à analyser des échantillons sanguins au microscope, mais cette analyse doit être effectuée par des personnes possédant l'expertise technique appropriée.

Le groupe de recherche sur l'IA de l'Université de Makerere, en Ouganda²¹, a mis au point un logiciel d'IA à l'aide de techniques de vision par ordinateur, et l'a formé au moyen d'échantillons de paludisme. Son système a obtenu de meilleurs résultats que les tests de dépistage des anticorps qui ont tendance à produire des taux élevés de faux positifs. Les systèmes automatisés comme celui-ci contribuent à renforcer les capacités en facilitant le triage des échantillons afin que les fournisseurs de soins de santé dans les centres très achalandés puissent travailler plus efficacement ou en étendant les capacités de diagnostic dans les régions rurales ou éloignées où l'expertise n'est pas disponible (Quinn et coll., 2014).

SURVEILLANCE DES MALADIES

En réponse à l'épidémie de dengue survenue en 2011 dans la province du Punjab, au Pakistan, un système de surveillance des maladies a été élaboré en vue de fournir des alertes rapides en cas d'épidémies futures. Le Système de surveillance des maladies intelligent du Punjab qui en est résulté utilise des algorithmes d'apprentissage statistique pour analyser des données en provenance d'une ligne d'urgence sur la dengue et de fils de nouvelles sur Internet, offrant aux hôpitaux et aux organismes gouvernementaux un suivi en temps réel des épidémies avec un niveau élevé de précisions géographiques (Ahmad et coll., 2013).

Photo par Ousa Chea sur Unsplash



L'intelligence artificielle fournit des outils pour aider à optimiser la prestation des services gouvernementaux

PRESTATION DE SERVICES ET DE RENSEIGNEMENTS GOUVERNEMENTAUX

Des efforts importants sont consentis mondialement afin de rendre les services et les renseignements gouvernementaux disponibles par voie électronique. On espère que les gouvernements atteindront une plus grande efficacité et une transparence améliorée, tandis que les citoyens bénéficieront d'un meilleur accès à ces services et à ces renseignements grâce à Internet et aux plateformes mobiles, et qu'ils trouveront ainsi qu'il est plus facile de participer à tous les aspects de la vie publique. Bon nombre de pays font face à de nombreux défis sur ce plan, notamment une infrastructure de technologie de l'information limitée, un approvisionnement irrégulier en électricité, le fait que les populations peuvent s'exprimer dans plusieurs langues et dialectes différents, et les grandes disparités dans l'accès à Internet et aux services mobiles. Les applications de l'IA possèdent le potentiel d'améliorer tant la compréhension que la mise en place d'applications et de services du gouvernement en ligne. En fin de compte, l'IA fournit des outils pour aider à optimiser la prestation des services gouvernementaux, et idéalement, à maximiser les retombées sociales tout en réduisant les coûts financiers.

ÉVALUATION DE L'EFFICACITÉ DU GOUVERNEMENT EN LIGNE

Évaluer l'efficacité des initiatives du gouvernement en ligne est une tâche complexe qui doit prendre en compte toute une gamme de facteurs techniques, organisationnels et sociaux comme la convivialité, l'accessibilité, l'inclusivité, les systèmes d'arrêt, les exigences liées à l'entretien, la sécurité, la protection de la vie privée, la satisfaction de la clientèle et les répercussions sur la productivité. En ayant recours à des systèmes experts et à des approches analytiques capables de prendre en compte des données dont la quantifiabilité et la spécificité sont variables, des groupes de partout autour du monde examinent des moyens de se servir de l'IA pour automatiser les évaluations des gouvernements en ligne afin d'aider les pays qui manquent d'argent à améliorer leurs efforts en matière de gouvernement en ligne. Les exemples comprennent notamment des recherches effectuées au Bangladesh (Hossain et coll., 2015), à Taïwan (Yang et coll., 2012), et en Grèce (Magoutas et Mentzas, 2010).

AMÉLIORATION DES SERVICES GOUVERNEMENTAUX

L'enjeu particulier de la diversité linguistique peut faire en sorte que des systèmes électroniques deviennent inaccessibles à des groupes entiers. En Afrique du Sud, par exemple, il existe 11 langues officielles. Le Centre de recherche sur l'intelligence de ce pays travaille à des méthodes de traduction automatique en vue d'élargir l'accès aux services gouvernementaux (Fondation World Wide Web, 2017). Dans le même ordre d'idées, le groupe de recherche sur l'IA de l'Université de Makerere cité ci-dessus, travaille à l'élaboration d'un ensemble de données sources pour quelques-unes des dizaines de langues parlées en Ouganda. Sans ces ensembles de données, il est impossible de mettre au point le traitement des langues naturelles nécessaire à la traduction automatique.

Photo par Mavis CW sur Unsplash



SOUTIEN EN CAS DE CATASTROPHE NATURELLE

L'une des toutes premières applications de l'IA dans le contexte du développement a été le soutien à la planification et à l'atténuation dans l'éventualité d'une catastrophe naturelle. Les heures qui suivent un événement catastrophique comme un tremblement de terre ou un ouragan sont chaotiques, et le flux de l'information est énorme et difficile à filtrer. Deux projets représentent les efforts réalisés en vue de tirer parti du soutien participatif et du déluge en temps réel des médias sociaux. Artificial Intelligence for Disaster Response (AIDR) est un projet de logiciel ouvert qui explore, classifie et attribue des balises aux fils Twitter durant les crises humanitaires²². Afin de donner un sens au flux d'information sur les médias sociaux qui dépasse la capacité d'analyse manuelle, AIDR utilise l'apprentissage automatique supervisé pour informatiser le processus, transformant les gazouillis bruts en une source organisée de renseignements susceptibles d'améliorer le processus décisionnel et les délais d'intervention.

En 2016, il s'est produit un séisme majeur en Équateur. Peu de temps après, une plateforme Web pour la recherche participative appelée Zooniverse a lancé un site Web au moyen duquel des bénévoles purent rapidement analyser des centaines d'images satellite afin d'aiguiller les secours dans les zones où l'on avait le plus besoin d'une assistance immédiate. La fiabilité de l'analyse de chaque contributeur était ensuite vérifiée par un système d'AA, et pondérée. Deux heures après son lancement, 1 300 images satellite avaient été examinées au moins 20 fois chacune, et une carte de densité des dommages était produite deux heures plus tard²³.

Du point de vue de l'évaluation des risques, des chercheurs de la Suède ont utilisé des techniques de systèmes experts semblables à celles de l'exemple en matière de gestion de l'eau ci-après (voir Agriculture, p. 45) pour modéliser les risques d'inondation au Bangladesh. Le système a été formé à l'aide d'un ensemble de données complexes regroupant une diversité de facteurs liés au

Photo par Ray Hennessy sur Unsplash



risque et aux répercussions d'une inondation. L'utilisation de données réelles sur une localité précise dans le pays a appuyé la production de scénarios potentiels pour faciliter la planification des secours et le processus décisionnel²⁴.

CONSERVATION DE LA FAUNE

En Afrique, de nombreuses espèces sont à la merci de braconniers. Tigres, éléphants, rhinocéros, et autres grands mammifères qui sont essentiels à la santé des écosystèmes et qui représentent des attractions majeures pour le tourisme sont exposés au dépeuplement régional et à l'extinction pure et simple. Afin d'améliorer l'efficacité de la lutte contre les braconniers, des chercheurs de l'University of Southern California ont mis au point un outil d'IA destiné aux patrouilles de rangers. Mise au point à l'origine en partenariat avec l'Uganda Wildlife Authority, le logiciel utilise l'AA formé à partir de données chronologiques sur les activités de braconnage locales pour produire des itinéraires de patrouille menant vers des endroits où les braconniers sont le plus susceptibles de se trouver. Les ensembles de données comprennent notamment des renseignements topographiques assortis de coordonnées GPS, des endroits où les animaux ont été vus et des preuves de braconnage comme des lieux où ont été retrouvés des carcasses et des pièges. Les versions d'essai du système ont été mises à l'épreuve en Malaisie et dans le parc national Queen Elizabeth, en Ouganda²⁵.

JUSTICE PUBLIQUE

Depuis de nombreuses années, le Mexique est aux prises avec une violence endémique liée à la drogue. L'un des résultats de cette violence est la disparition de plus de 30 000 personnes depuis 2006. Il n'est pas rare de voir les corps de ces desaparecidos être éventuellement retrouvés dans des fosses communes, mais tomber sur ces charniers tient bien souvent du hasard. Une collaboration entre trois groupes du Mexique et des États-Unis utilise l'AA afin de rendre le processus plus systématique. Les 2 457 comtés du Mexique sont utilisés à titre d'unités géographiques. Un profil sociodémographique détaillé est élaboré pour chaque comté et entré dans le système, de même que les données tirées des reportages publiés dans la presse sur chacune des découvertes de charniers. Le modèle utilise ces données pour déterminer quels sont les comtés les plus susceptibles d'avoir été utilisés pour se débarrasser des cadavres des victimes. Même si la collaboration fait état d'un degré élevé d'exactitude (100 % en 2014), cette exactitude dépend des données, et la collecte de données issues des médias est laborieuse et coûteuse en temps²⁶.



AGRICULTURE

Dans de nombreux pays du Sud, l'agriculture est une composante importante de l'économie, et une bonne partie de la population dépend de l'exploitation agricole comme source de nourriture. Cependant, les cultures saines et les récoltes abondantes peuvent subir les effets de la maladie, des insectes et de la sécheresse. Les applications de l'IA peuvent fournir des perspectives et des solutions critiques qui peuvent améliorer l'efficacité et la qualité des activités agricoles.

Des applications de l'IA sont capables de détecter les maladies et les infestations dans les cultures en analysant des photos prises avec des téléphones cellulaires

SURVEILLANCE DES MALADIES DES CULTURES

La surveillance des maladies des cultures est une entreprise coûteuse en temps qui requiert souvent des connaissances approfondies qui ne sont pas toujours disponibles localement en temps utile. Voici deux exemples d'applications de l'IA qui sont capables de détecter les maladies et les infestations dans les plantes cultivées en analysant des photos prises avec des téléphones cellulaires.

En Afrique, le manioc est un aliment de première nécessité dont les rendements sont appelés à diminuer en raison de maladies virales. Une enquête approfondie pour produire un diagnostic sur les cultures atteintes et pour cartographier l'ampleur de la propagation de la maladie peut prendre des mois et exige des déplacements importants de la part des enquêteurs. Les chercheurs spécialisés en IA de l'Université de Makerere ont également mis au point une méthode pour optimiser ce processus habituellement réalisé sur papier en recueillant des images d'échantillons — prises avec des téléphones cellulaires — en vue d'en faire l'analyse et la classification à l'aide d'un système d'IA. Les images des symptômes de la maladie, comme les dommages causés aux racines et l'accumulation de mouches blanches sur les feuilles, sont introduits dans un algorithme d'AA pour produire un diagnostic rapide et une rétroaction, et une fonction de cartographie peut être utilisée pour évaluer et prévoir la propagation au fil du temps. Ce même groupe a mis au point des systèmes comparables pour évaluer les maladies des plants de bananes en utilisant des images des feuilles (Quinn, 2013).

Des chercheurs de la Pennsylvania State University et de l'École polytechnique fédérale de Zurich ont adopté une approche plus généralisée, en mettant au point une application d'IA qui peut analyser des photographies pour identifier des cultures et des maladies avec un taux d'exactitude de presque 100 %. Au moyen d'une approche d'apprentissage profond, ils ont formé le système avec une base de données de plus de 53 000 photos de plantes saines et de plantes malades à partir d'archives d'images en libre accès. Le système peut reconnaître 14 cultures et 26 maladies parmi 38 combinaisons. Un système comme celui-là peut être utilisé par les agriculteurs pour repérer rapidement les maladies des cultures dans le champ, à l'aide de photos prises avec des téléphones intelligents²⁷.



GESTION DE L'EAU ET SURVEILLANCE DES SÉCHERESSES

En réponse aux problèmes de rareté de l'eau en Palestine, un groupe en Autriche a collaboré avec des collègues palestiniens en vue de mettre au point un système expert pour atténuer les fuites dans les systèmes d'aqueduc. En ayant pour but d'optimiser une stratégie de gestion des fuites d'eau, le système a évalué un éventail de mesures précises – comme un contrôle actif et passif des fuites, la gestion de la pression d'eau, les compteurs d'eau et les campagnes de sensibilisation du public – en fonction d'un ensemble de critères d'évaluation économiques, environnementaux, techniques et socioéconomiques. Pour s'attaquer au problème complexe consistant à intégrer de nombreuses variables et une combinaison de données qualitatives et quantitatives, les développeurs ont eu recours à un type de mathématiques analytiques pour classer les mesures disponibles appropriées aux conditions locales²⁸.

L'accès à des informations opportunes sur les conditions climatiques et agricoles d'endroits spécifiques peut aider à détecter, prévenir ou répondre à des catastrophes telles que la sécheresse. Le groupe de l'Université de Makerere, en Ouganda, a élaboré un système visant à produire de tels rapports. À partir de sujets ciblés déterminés manuellement (par exemple l'eau, les précipitations, l'état du sol) et de critères de recherche, le système balaie le Web, télécharge les résultats de recherche qui figurent en tête de liste, extrait le contenu et résume l'information pour déterminer les tendances pertinentes pour les sujets indiqués (Quinn et coll., 2014).



L'intelligence artificielle peut contribuer à faire évoluer l'offre au-delà d'un modèle de prestation universel de l'éducation qui n'a pas changé considérablement depuis un siècle

ÉDUCATION

L'éducation sous-tend le développement humain et social. En dépit de gains massifs dans les démarches visant à obtenir l'éducation pour tous, il arrive trop souvent que cette éducation soit de mauvaise qualité. Les systèmes d'enseignement dans les pays du Sud font face à de nombreux défis, y compris des ressources insuffisantes ou de mauvaise qualité et une pénurie d'enseignements qualifiés.

Les promesses de l'IA en matière d'éducation sont importantes, étant donné qu'elle peut contribuer à faire évoluer l'offre au-delà d'un modèle de prestation universel et industriel qui n'a pas changé considérablement depuis un siècle. En effet, les techniques d'IA peuvent être utilisées pour appuyer (et peut-être même parfois pour remplacer) les rôles des enseignants, des tuteurs et des administrateurs en vue d'améliorer l'enseignement et les méthodes d'apprentissage et de les axer sur les étudiants et de les personnaliser — un progrès fondamental requis pour transformer l'éducation (Winthrop et McGivney, 2017). Plus particulièrement, les techniques d'IA peuvent offrir des possibilités d'apprentissage personnalisées de qualité, à grande échelle, et peuvent aussi faciliter la création d'un contenu de qualité.

APPRENTISSAGE PERSONNALISÉ

Peut-être que le principal changement par rapport à un processus d'apprentissage axé sur l'enseignant est l'élaboration de l'apprentissage personnalisé. L'un des moyens de l'offrir est par l'entremise de systèmes tutoriels intelligents (STI), aussi appelés tuteurs cognitifs. Un STI est habituellement un système expert qui tente de recréer l'instruction individuelle en adaptant et en personnalisant l'expérience d'apprentissage en fonction de l'apprenant:



Un STI évalue chacune des actions de l'apprenant dans le cadre de ces environnements interactifs et élabore un modèle de ses connaissances, de ses compétences et de son expertise. En se fondant sur le modèle de l'apprenant, il peut personnaliser des stratégies d'enseignement, sur le plan du contenu et du style, et fournir des explications pertinentes, des conseils, des exemples, des démonstrations et des exercices aux apprenants à titre individuel (Phobun et Vicheanpanya, 2010 : 4065).

À titre de système expert, un STI intègre trois types de modèles de connaissances : le modèle du domaine (représentant l'expertise de l'enseignant dans le domaine), le modèle de l'apprenant (les modèles de la manière dont l'individu apprend), et le modèle pédagogique (pour prendre des décisions sur les pratiques pédagogiques). Voir Figure 1 ci-après.

Les STI commencent aussi à intégrer d'autres techniques d'IA en vue d'améliorer l'enseignement. Certains STI – appelés systèmes tutoriels intelligents avec

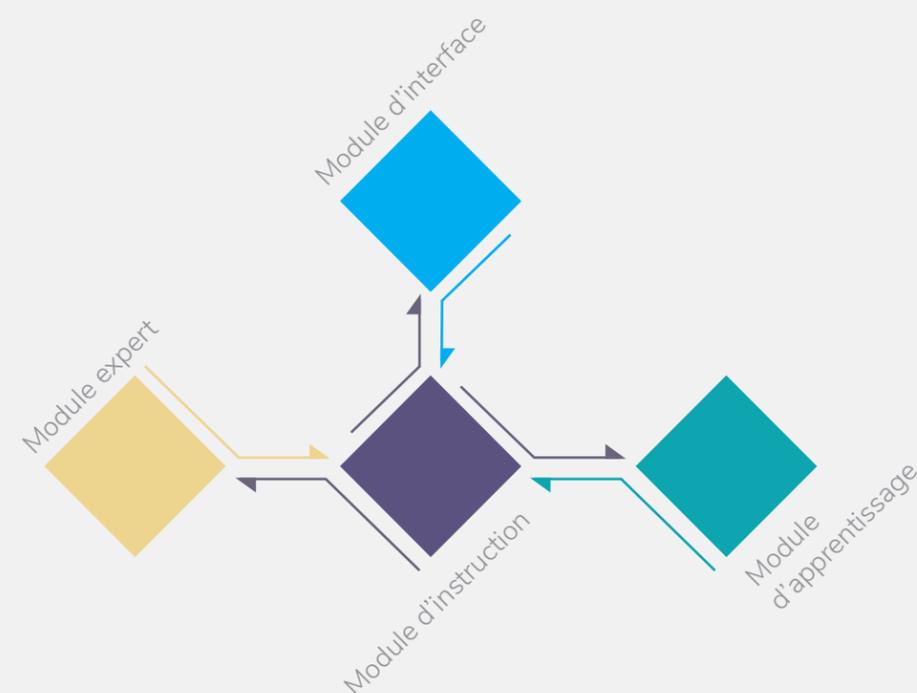
reconnaissance émotionnelle – intègrent aussi la reconnaissance émotionnelle à titre de moyen d'améliorer l'adaptation du tutoriel à l'étudiant (Petrovica et coll., 2017). Les STI peuvent aussi appliquer le traitement du langage naturel et la reconnaissance vocale pour repérer plus facilement les erreurs de langage ou pour interagir avec les étudiants de façon novatrice²⁹.

De nouvelles données probantes indiquent que l'enseignement personnalisé au moyen d'un STI peut se révéler très efficace. Un compte rendu d'études menées en 2016 sur l'efficacité des STI a permis de constater des améliorations considérables du rendement dans 78 % des 50 études examinées (Kulik et Fletcher, 2016). Il existe de solides données probantes à l'appui de l'utilisation des STI dans les contextes des pays du Sud également, et ce, en dépit d'obstacles connus (Nye, 2015). Lorsqu'ils sont bien utilisés, les STI peuvent se révéler un moyen rentable et très inclusif d'améliorer les pratiques d'enseignement.



Photo par Annie Spratt

FIGURE 1 Structure d'un système de tutorat intelligent



Structure type d'un système de tutorat intelligent intégrant trois modèles de connaissances : l'expert du domaine (l'enseignant), l'apprenant, et l'instruction (les tactiques pédagogiques).
Tiré de : Phobun et Vicheanpanya 2010 : 4066

Mindspark, une plateforme tutorielle intelligente mise au point en Inde, en est un bon exemple. Mindspark détermine les modèles d'erreurs commises par les utilisateurs étudiants, et personnalise les leçons et les tutoriels suivants en conséquence. Une évaluation rigoureuse de Mindspark a révélé que le programme était tout à la fois rentable et « efficace pour cibler l'instruction avec précision en fonction du niveau de réussite de l'étudiant et pour composer avec une grande variété de préparation des cours pour la même année d'enseignement » [Traduction] (Muralidharan et coll., 2017 : 23).

L'analytique de l'apprentissage est la discipline consacrée à « la mesure, la collecte, l'analyse et la présentation de rapports fondés sur des données sur les apprenants en contexte d'apprentissage dans le but de comprendre et d'optimiser l'apprentissage et le contexte » (Long et Siemens, 2011 : 34). Pour y arriver, l'analytique de l'apprentissage tire parti de vastes quantités de données sur l'enseignement et de techniques d'AA. Par exemple, appliquer l'analytique de l'apprentissage à de grandes quantités de données sur les apprenants peut contribuer à améliorer les modèles de connaissances dans les STI, par exemple en repérant les tâches d'apprentissage qui comportent les gains les plus efficaces

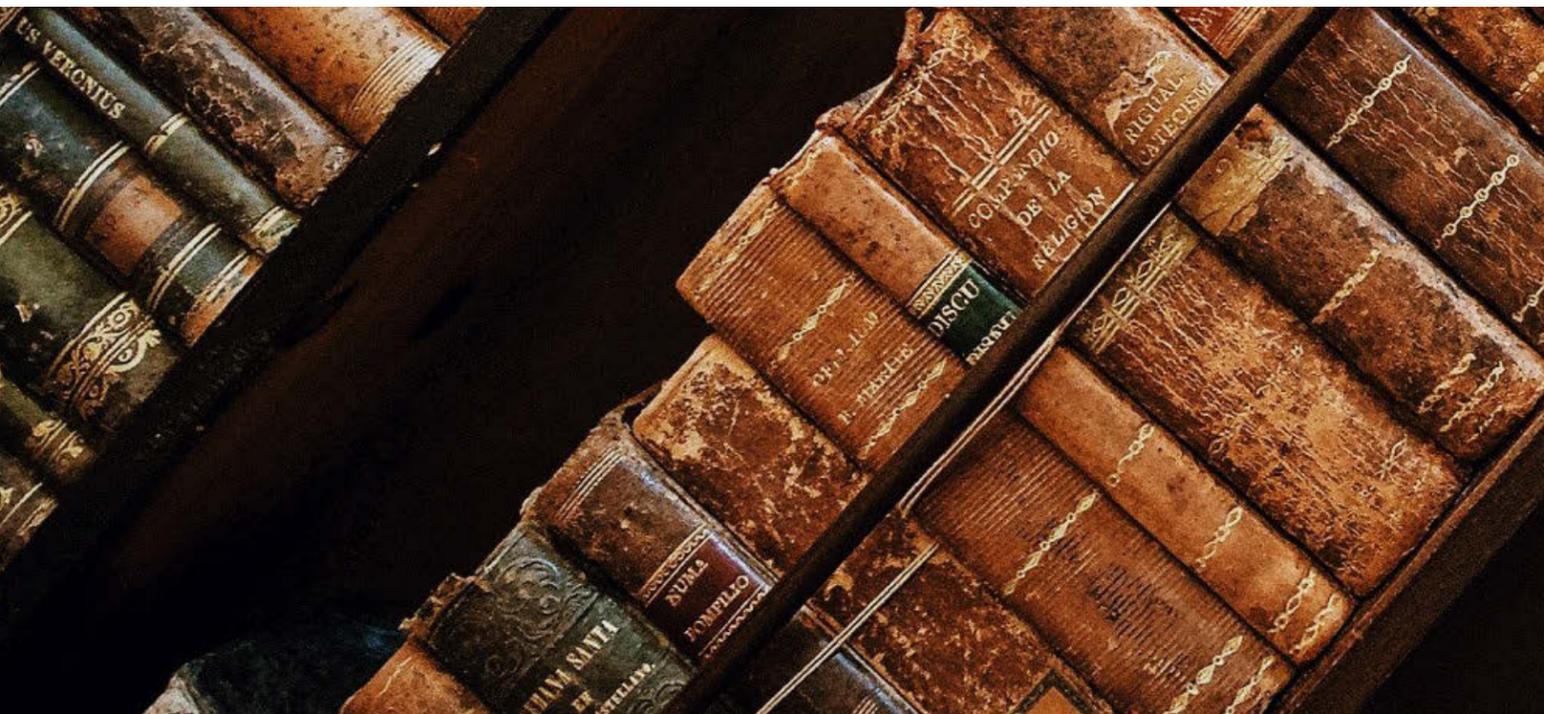


(Lim et Tinio, 2018). L'analytique de l'apprentissage peut aussi recourir à l'analytique prédictive pour repérer les étudiants qui risquent d'échouer un cours.

Les premiers exemples recueillis dans les pays du Sud sont encourageants. À titre d'exemple, Mwalumbwe et Mtebe (2017) ont appliqué l'analytique de l'apprentissage aux données sur l'engagement des étudiants recueillies à l'aide d'un système de gestion de l'apprentissage pour prévoir le rendement scolaire des étudiants. Reyes (2018) s'est servi de données comparables pour fournir des intrants à un système d'aide aux étudiants à l'Université ouverte des Philippines.

APPRENTISSAGE À GRANDE ÉCHELLE

L'une des principales difficultés inhérentes à la fourniture de l'apprentissage de qualité à grande échelle consiste à le faire à faible coût. L'accès accru à une connectivité qui ne cesse de s'améliorer, joint à des méthodes d'apprentissage numériques en ligne comme les cours en ligne ouverts à tous, a accru la capacité d'offrir à un grand nombre de personnes un contenu éducatif et des expériences de qualité. Même si les premières données probantes sur les cours en ligne ouverts à tous montrent que les personnes affichant un niveau d'instruction et un statut socioéconomique plus élevés ont tendance à en profiter de façon disproportionnée (Christensen et coll., 2013; Hansen et Reich, 2015), des recherches révèlent que certains types de cours en ligne ouverts à tous – comme ceux qui dispensent de la formation liée à l'emploi – peuvent bénéficier aux utilisateurs des populations de faible revenu et de revenu intermédiaire dans les pays du Sud (voir, par exemple, Garrido et coll., 2016).



Les techniques d'IA possèdent le potentiel de tabler sur l'apprentissage en ligne pour atteindre l'objectif ambitieux d'offrir un apprentissage de haute qualité à grande échelle, et particulièrement aux populations marginalisées. Le secret pour y parvenir consiste à surmonter les goulots d'étranglement qui surviennent lorsqu'on s'engage auprès d'un grand nombre d'étudiants – principalement le manque de ressources humaines pour formuler une rétroaction personnalisée, des conseils et l'évaluation du rendement de l'étudiant. Une combinaison de techniques d'IA peut aider à prendre en charge des volumes élevés d'étudiants. Les tutoriels intelligents et l'analytique de l'apprentissage, dont il a été question ci-dessus, peuvent aider à fournir des expériences d'apprentissage personnalisées de grande qualité à très faible coût par étudiant additionnel (Laurillard et coll., en cours d'impression). De plus, les progrès réalisés dans l'évaluation par score automatisée, au moyen du traitement du langage naturel et d'autres techniques permettent la correction massive d'interrogations, d'examens et de devoirs.

PERSONNALISATION/LOCALISATION DU CONTENU

Les techniques d'IA peuvent aussi contribuer à améliorer le contenu éducatif à faible coût, par exemple en fournissant la traduction automatique des travaux existants dans de nouvelles langues ou en misant sur l'IA pour créer de nouveaux contenus.

À titre d'exemple, la plateforme Pratham Books de StoryWeaver collabore avec Google.org en vue d'utiliser l'outil de traduction propulsé par l'IA de Google pour traduire des livres pour enfants dans autant que 60 langues différentes³⁰. Bien entendu, la possibilité de recourir à la traduction automatique devrait grandement faciliter la localisation et l'utilisation de ressources pédagogiques ouvertes et de qualité (matériel pédagogique, habituellement sous forme numérique, offert gratuitement) dans les pays du Sud (Smith, 2013).

Une autre entreprise, Content Technologies, Inc.³¹ utilise les techniques d'AA pour créer des livres sur mesure. Les enseignants entrent leur plan de cours et leur matériel dans le moteur d'IA, et le système crée des manuels et du matériel pédagogique établis en fonction des concepts de base qu'il a extraits³². On peut également envisager la création automatisée de guides d'étude, d'interrogations et de tests, lesquels seraient particulièrement utiles dans les environnements d'apprentissage en ligne ouverts à tous.

Photo par Roman Kraft sur Unsplash



L'intelligence artificielle pourrait contribuer à hauteur de
15,7 billions de dollars à l'économie mondiale en 2030

PwC 2017: 5

ÉCONOMIE ET AFFAIRES

L'IA peut se révéler un important levier pour le développement économique grâce à divers moyens. Selon un rapport de PwC de 2017, l'« IA pourrait contribuer à hauteur de 15,7 billions de dollars à l'économie mondiale en 2030, soit plus que le PIB combiné actuel de la Chine et de l'Inde » (p. 5), grâce aux gains de productivité entraînés par l'automatisation des processus des entreprises et l'augmentation de leur effectif actuel avec les technologies d'IA, ainsi qu'à la hausse de la demande des consommateurs pour des produits et services de qualité supérieure améliorés par l'IA. L'IA offre la possibilité d'accroître la productivité en créant des gains d'efficacité, en automatisant des méthodes de travail et en optimisant des activités commerciales clés comme les chaînes d'approvisionnement et l'établissement des prix³³. Cependant, le rapport de PwC qualifie les avantages économiques potentiels dans les pays à faible revenu et à revenu intermédiaire, en affirmant que les augmentations y seront « plus modestes... compte tenu du taux d'adoption des technologies d'IA beaucoup plus faible qui est prévu » (PwC, 2017: 9).

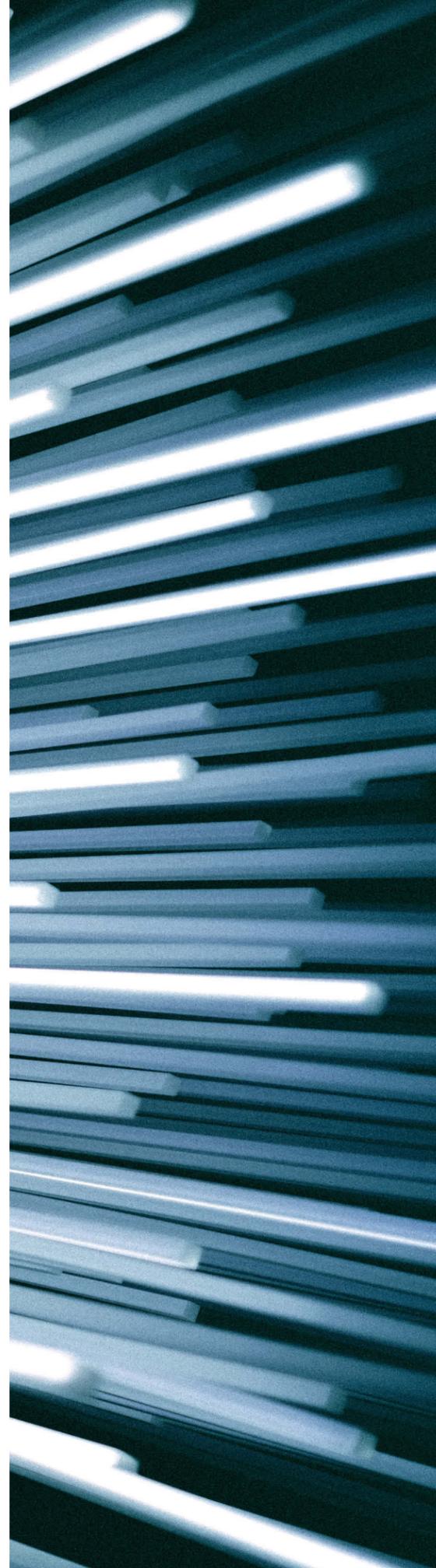
Voici des exemples de deux domaines où l'IA commence à contribuer à la croissance dans les pays du Sud : 1) le soutien des nouvelles entreprises et l'innovation, 2) l'optimisation des composantes de base de l'économie, telles que les services financiers.

Photo par Joshua Ness

ENTREPRENEURIAT ET INNOVATION

La disponibilité d'une puissance de calcul considérablement améliorée pour une fraction de ce qu'il en aurait coûté il y a dix ans, et le développement d'un logiciel ouvert ont déclenché une explosion de jeunes entreprises techniques partout dans le monde. Compte tenu des nombreuses applications potentielles de l'IA, il existe des possibilités considérables sur le plan entrepreneurial dans les pays du Sud, et elles devraient probablement s'accroître avec les progrès réalisés dans l'infrastructure et les nouvelles méthodes de collecte de données. On peut déjà constater la montée des jeunes entreprises en IA dans certaines économies émergentes. En voici seulement quelques exemples:

- ▶ L'Inde compte de nombreuses jeunes entreprises qui misent sur l'IA. Parmi les applications on retrouve des outils pour développeurs en vue de créer des systèmes d'IA et des réseaux neuronaux, des plateformes d'apprentissage, la planification des ressources humaines, un assistant personnel pour les voyageurs, l'optimisation de la prestation et de la logistique ainsi qu'un diagnostic à distance des maladies cardiaques (Manyika et coll., 2017).
- ▶ LangBot, une jeune entreprise d'Addis-Abeba, en Éthiopie, travaille à la mise au point de « robots conversationnels propulsés par l'IA et ludifiés pour l'enseignement des langues ». L'entreprise en démarrage a récemment remporté le volet éthiopien du concours pour les jeunes entreprises de SeedStars qui se concentre sur les marchés émergents³⁴.
- ▶ À Pretoria, en Afrique du Sud, la jeune entreprise spécialisée dans les soins de santé en ligne hearX Group met au point des outils numériques pour évaluer la perte d'audition et détecter les maladies de l'oreille. L'un des produits en développement est un otoscope bon marché connecté à un téléphone cellulaire pour reproduire une image du canal auditif. L'appareil se servira de l'analyse d'images et de l'IA pour produire des diagnostics automatisés des maladies courantes de l'oreille. La campagne de sociofinancement du groupe a été submergée d'inscriptions³⁵.
- ▶ La question de la santé sexuelle est taboue dans bien des milieux africains. Cette situation contribue aux taux élevés d'infection au VIH et de MTS ainsi que de grossesses non désirées. Un incubateur de technologies du Kenya a récemment financé les développeurs de Sophie Bot, un robot conversationnel utilisant l'IA qui permet aux utilisateurs d'obtenir en privé des renseignements exacts sur la santé sexuelle et reproductive³⁶.



SERVICES FINANCIERS

Dans les pays du Sud, de grands segments de la population n'utilisent pas les services financiers ou le secteur bancaire structurés. En Afrique, ces segments dépassent les 325 millions de personnes; cependant, la pénétration importante de la téléphonie mobile dans bien des régions offre une plateforme pour avoir accès à ces services. Une entreprise appelée MyBucks se sert de l'IA pour offrir la prestation de services virtuels dans au moins neuf pays africains. L'entreprise a aussi fait l'acquisition de plusieurs institutions financières ayant pignon sur rue en Afrique subsaharienne dans le but d'étendre sa portée pour rejoindre les communautés pauvres et mal desservies. MyBucks utilise l'IA pour automatiser des tâches comme l'établissement de la fiabilité financière, la détection de fraudes et l'optimisation afin de maintenir les coûts bas pour les prêts d'un petit montant et le microcrédit, les comptes d'épargne, l'assurance et le portefeuille de transaction qu'elle offre³⁷.

Si MyBucks compte parmi les premières entreprises à avoir fait appel à l'IA dans les services financiers en Afrique, d'autres se préparent à lui emboîter le pas. En 2016, Barclays Africa a lancé un robot conversationnel pour son service à la clientèle qui se sert de l'IA pour simuler une « conversation intelligente » destinée à répondre à des requêtes simples³⁸.

Au Nigeria, un autre robot conversationnel, Kudi AI, est un assistant bancaire personnel qui utilise le traitement du langage naturel pour permettre aux utilisateurs d'effectuer des transactions financières simples (comme payer des factures ou virer des fonds) au moyen de l'application Messenger par Facebook. Étant donné que le service est offert par l'entremise de la plateforme Free Basics par Facebook, il n'y a aucuns frais d'utilisation. Les virements bancaires, d'une institution à une autre, sont gratuits, et les autres transactions (comme le paiement de factures) coûtent l'équivalent d'environ 30 cents³⁹.



Les assises nécessaires en
vue d'une application éthique
et équitable de l'IA sont
largement absentes

Défis et risques

Comme nous l'avons illustré dans la section précédente, les applications d'IA possèdent le potentiel de favoriser des changements sociaux positifs – en effet, dans certains domaines, leur incidence pourrait être révolutionnaire. Cependant, comme avec toute nouvelle technologie, l'atteinte réelle de ces résultats positifs comporte des défis, et n'est pas sans risque. Dans cette section, nous commençons par exposer certains facteurs contextuels qui peuvent représenter des obstacles potentiels à l'utilisation efficace de l'IA pour le développement. Nous abordons ensuite une série de risques d'obtenir des retombées sociales négatives issues du déploiement d'applications d'IA dans des contextes difficiles des pays du Sud.

DÉFIS CONTEXTUELS

Selon le Rapport sur le développement dans le monde 2016 : Les dividendes du numérique (Banque mondiale, 2016), la capacité des pays d'obtenir une répartition positive et équitable des avantages tirés des technologies numériques repose sur l'existence d'un ensemble d'assises « analogiques » préalables : réglementation, compétences et institutions. Nous savons aussi, d'après des recherches et des expériences réalisées dans le passé, qu'un contre-pouvoir est requis pour remettre en question les inégalités qui persistent (Benkler, 2011). Dans le cas de l'IA, les assises tant numériques qu'analogiques requises en vue d'une application éthique et équitable de la technologie dans de nombreux pays du Sud sont largement absentes, et une asymétrie marquée en matière de pouvoir demeure.

Voici quelques-uns des principaux défis contextuels:

Compétences

Des lacunes en matière de financement et de politiques pour les sciences et l'innovation dans les pays du Sud freinent le développement de l'expertise et de la recherche locales en IA. De ce fait, les possibilités de développement d'applications et d'innovations organiques en matière d'IA sont limitées, et ils doivent se contenter d'être des importateurs plutôt que des créateurs de la technologie. Cette situation freine aussi la conception d'applications d'IA pertinentes à l'échelle locale.

On constate une pénurie générale de compétences concernant le développement et le déploiement d'applications d'IA, et un manque de diversité bien connu chez ceux qui possèdent les compétences requises. Ce manque de diversité contribue à l'introduction de préjugés dans les applications d'IA. Par exemple, comme dans le secteur technique de manière générale, il y a peu de femmes dans le domaine de l'intelligence artificielle⁴¹, et la majorité des hommes qui y travaillent sont blancs⁴². Cette disparité tire son origine d'influences systémiques et culturelles. Aux États-Unis, les garçons sont beaucoup plus nombreux que les filles à se présenter aux examens de l'Advanced Placement Computer Science (AP), et en 2013, seulement 26 % des informaticiens professionnels du pays étaient des femmes. On peut s'attendre à trouver un schéma comparable lié à la disparité entre les sexes dans les pays du Sud.

Il existe un écart de compréhension important entre les chercheurs en sciences sociales, les décideurs, les ONG et ceux qui ont une compréhension technique de l'IA. En effet, ceux qui mettent en œuvre l'IA ne sont généralement pas conscients de ses répercussions sur le plan social, politique et éthique, et ceux qui sont responsables de la réglementation ou de l'application de la technologie pour le bien social comprennent assez mal comment elle fonctionne.

Capacité/réglementation institutionnelle

La gouvernance et la réglementation de la technologie sont rudimentaires dans bien des pays du Sud, ce qui mène à des occasions d'innovations technologiques qui ne sont pas toujours possibles, ou qui mettent beaucoup de temps à gagner du terrain dans les pays du Nord. L'application bancaire mobile MPesa, au Kenya, en est un exemple. Toutefois, la réglementation limitée s'accompagne du potentiel d'entraîner de plus grands dommages sur le plan éthique, social et politique. Certains experts, par exemple, ont servi la mise en garde que l'innovation technologique sans entraves dans des endroits comme l'Afrique subsaharienne constitue, selon eux, une bombe à retardement.⁴⁰

Les gouvernements, universités et société civile ne disposent que d'une capacité et des ressources limitées pour concevoir des cadres réglementaires pour l'innovation technologique, et plus précisément, pour des cadres qui établissent un équilibre entre favoriser l'innovation tout en protégeant la vie privée, la sécurité ou l'environnement.

La forte poussée exercée par le secteur privé dans les pays du Sud pour promouvoir les solutions d'IA particulières de ce secteur peuvent évincer le développement d'une technologie locale.

Infrastructure

Malgré des progrès rapides depuis 2010 dans la prolifération des téléphones mobiles et des connexions Internet à large bande, l'infrastructure de base en matière de télécommunications et de services numériques demeure insuffisante dans bien des contextes, et plus particulièrement ceux qui se trouvent en milieu rural et de faible revenu. Même si nous anticipons que ce fossé en matière d'accès diminuera considérablement au cours de la prochaine décennie, il subsistera très probablement un fossé persistant entre la « voie rapide » et la « voie lente » numérique (Hilbert, 2016), qui suscite des questions à savoir qui peut innover de façon réaliste dans l'IA dans un avenir rapproché.

On manque d'ensembles de données pour former les algorithmes d'IA propres aux pays du Sud, ce qui contribue à créer ce « fossé des données ». Une mauvaise connectivité et des technologies de connecte de données limitées dans bien des régions (surtout dans les zones rurales) continueront de rendre la collecte de données extrêmement difficile, pour ne pas dire impossible. Au cours des cinq prochaines années, les données des téléphones mobiles et des services financiers constitueront les sources de données les plus probables dans ces pays.



Photo par Marcus Cramer

Un élément important de ce contexte est caractérisé par ce que l'on peut appeler le « fossé de l'IA » – c'est-à-dire, un écart entre ceux qui ont la capacité de concevoir et de déployer des applications d'IA, et ceux qui ne l'ont pas. Il peut être vu comme l'écart entre ceux qui participent et ont une voix et une organisation pour façonner les applications qui sont mises au point et déployées, et la manière dont on s'y prend pour le faire, et ceux qui sont laissés pour compte dans le processus.

Ce fossé de l'IA transcende les limites géographiques, socioéconomiques, sexospécifiques et raciales. L'infrastructure requise pour le développement d'applications d'IA restreint cette activité, majoritairement, aux endroits où l'on dispose d'une puissance de calcul suffisante, d'un accès à des données pertinentes (ou de ressources pour les recueillir) ainsi que des compétences en IA. La géographie de l'écart entre le taux de participation est peut-être le mieux illustrée par la dominance relative de quelques pays (et de quelques grandes sociétés technologiques) dans le développement de l'IA. Une analyse réalisée par PwC a estimé que 70 % des retombées économiques mondiales de l'IA vont revenir à la Chine et à l'Amérique du Nord (PwC, 2017). Ainsi, par exemple, la majorité des véhicules autonomes sont conçus et mis à l'essai dans les contextes des pays du Nord où l'on trouve des routes de grande qualité, une bonne signalisation et un bon marquage ainsi qu'une circulation automobile ordonnée. Il est probable que la circulation de ces véhicules dans certains contextes des pays du Sud serait problématique. Ces technologies doivent être au moins formées et mises à l'épreuve dans ces contextes, si on entend les y utiliser.

Au-delà de la géographie, la disparité dans les compétences en matière d'IA nécessaires pour concevoir et déployer les applications d'IA se reflète dans le genre, l'origine ethnique et la classe socioéconomique des personnes qui font le travail. Les personnes qui possèdent des compétences en matière d'IA sont habituellement de sexe masculin, blanches et vivent dans les pays du Nord, là où la majorité des activités liées à l'IA se déroulent. Un rapport du Forum économique mondial (2016) signalait que dans le monde, seulement 32 % des diplômés en sciences, technologie, ingénierie et mathématiques (STIM) sont des femmes. Par conséquent, ceux qui décident de ce qui sera conçu et déployé devront être conscients des limites de leurs connaissances contextuelles et de leurs propres préjugés sociaux éventuels.



SECTEURS DE RISQUE

Dans cette section, nous abordons cinq types de risques, à court et à moyen terme, associés à l'application de l'IA⁴³. Deux de ces risques – les préjugés et le manque de transparence du processus décisionnel exécuté par les applications d'IA – sont inhérents à la technologie en tant que telle. Les trois autres – la surveillance accrue et l'atteinte au droit à la vie privée, la désinformation ciblée, et l'automatisation entraînant des pertes d'emplois – sont le résultat d'applications particulières de l'IA dans différents domaines. Même si ces risques sont présents dans n'importe quel contexte social ou économique (la majorité des exemples sont tirés de pays du Nord), il reste que les contextes politiques et institutionnels des PFR-PRI exacerbent potentiellement les risques en question.

ÉQUITÉ, BIAIS ET RESPONSABILISATION

Les humains sont, de nature, biaisés et sujets à commettre des erreurs. Même lorsque nous faisons des efforts conscients pour agir de façon impartiale, nos préjugés et nos biais sont tellement ancrés en nous que nous ne nous rendons même pas compte de leur influence. La justice, par exemple, est « aveugle » en théorie, mais la partialité est clairement présente dans les systèmes de justice partout dans le monde.

Ordinateurs et algorithmes, en revanche, sont les symboles de l'efficacité et de l'exactitude. À titre de machines et de codes numériques, on s'attend à ce qu'ils fonctionnent de manière équitable et impartiale. Il semblerait donc logique de concevoir des logiciels et des algorithmes d'IA pour surmonter ces limites humaines, aussi des applications d'IA ont été mises au point pour aider les juges dans l'espoir qu'ils imposeront des peines plus justes et moins biaisées⁴⁴.

Cependant, notre confiance dans l'impartialité des systèmes informatiques est mal placée. Parce que ces systèmes sont conçus et mis au point par des humains, formés à partir de données produites par des humains, et qu'ils intègrent les préjugés de leurs concepteurs et programmeurs, ainsi que de la culture en général au sein de laquelle ils sont créés (Friedman et Nissenbaum, 1996; Friedman et coll., 1996).

Friedman et Nissenbaum (1996) définissent les systèmes informatiques biaisés comme ceux qui « exercent systématiquement et injustement une discrimination à l'égard de certains individus ou groupes d'individus au profit d'autres. » Cette définition demeure appropriée pour les applications d'IA. En effet, on dispose



déjà de données probantes montrant que ces applications peuvent systématiquement et injustement faire preuve de discrimination dans des aspects importants de la vie des gens. Cela n'a rien de surprenant étant donné la capacité potentielle de l'IA d'exercer une vaste influence sur les décisions concernant l'accès individuel à des services, à l'emploi et au soutien financier. Compte tenu d'une telle influence, les applications d'IA biaisées peuvent être la source d'iniquités et d'injustices considérables dans la société.

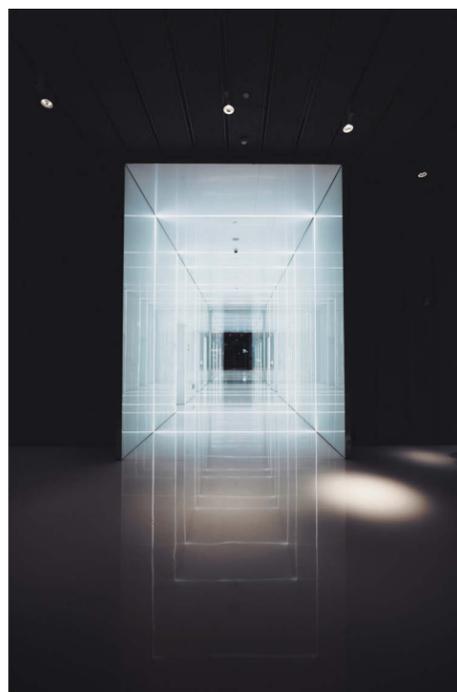


Photo par Alessio Lin

Biais de conception

Le processus de conception constitue une source potentielle de biais dans les applications d'IA. Il est bien compris que les technologies reflèteront bon nombre des valeurs souvent implicites de ceux qui les ont conçues et construites (Nissenbaum, 2001; Van de Poel et Kroes, 2014). Nombre d'articles publiés récemment portaient sur le penchant masculin qui prévaut dans la Silicon Valley et dans l'industrie du logiciel en général. Ce type de biais façonne la sélection des applications en vue du développement et les caractéristiques de ces applications. On peut voir un exemple de ces préjugés dans les choix effectués par les concepteurs d'attribuer un genre aux applications d'IA (voir la Figure 2 ci-après). Jusqu'ici, les assistants numériques sont principalement féminins, tandis que les personnalités de l'IA dans les films sont majoritairement de sexe masculin. Il faut reconnaître qu'il s'agit d'un territoire relativement nouveau sur le plan commercial, et l'industrie semble hésiter à choisir quelle devrait être la meilleure option pour les robots conversationnels et les assistants numériques – femme, homme, la possibilité de choisir l'un ou l'autre, ou un genre neutre.

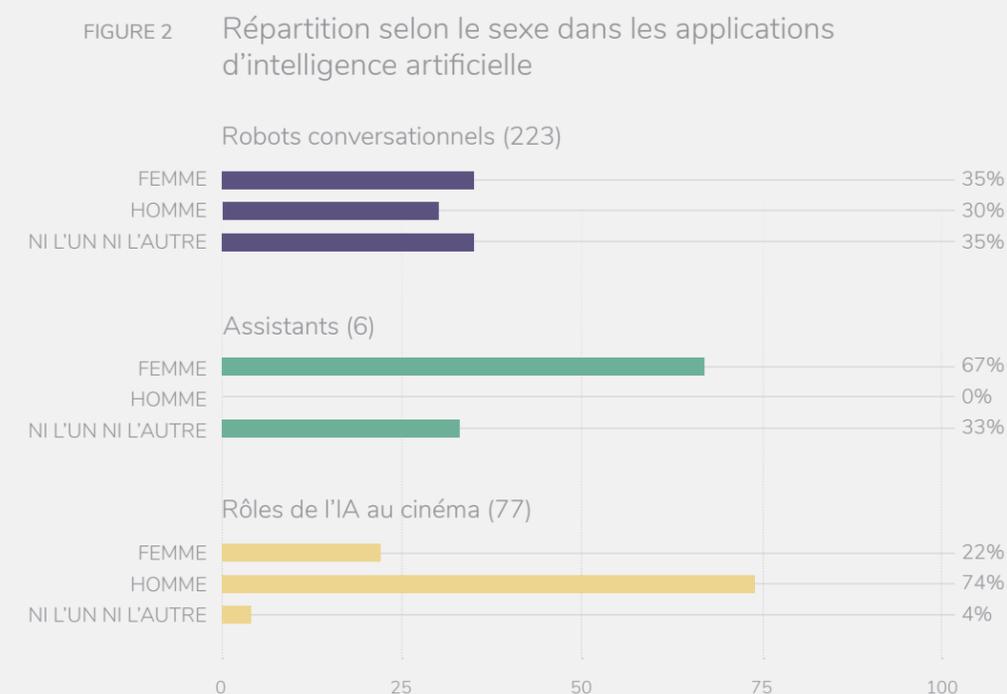
Les autres préjugés pourraient inclure le langage dans lequel un algorithme fonctionne, le système d'exploitation pour lequel il a été construit, et les hypothèses au sujet de ses utilisateurs. Chacun de ces aspects de la conception comporte des points de décision qui peuvent, ou non, être

reconnus comme étant chargés de valeur, mais qui peuvent néanmoins introduire des biais susceptibles d'influencer qui peut utiliser une application d'IA et en tirer parti.

On peut imaginer qu'il existe aussi un potentiel de biais de conception dans les systèmes experts, lesquels sont construits à partir de bases de connaissances qui reflètent l'expertise de contributeurs humains. Pour cette raison, on a tenté de mettre en place des processus d'accréditation destinés à évaluer les résultats produits par ces systèmes, un peu sur le modèle de l'accréditation des médecins et des enseignants. Cependant, le reste de cette section portera sur les algorithmes d'AA, étant qu'il existe beaucoup moins de documentation sur le sujet des biais dans les systèmes experts.

Biais appris

Les algorithmes d'AA ne sont pas biaisés en tant que tels; ils apprennent à être biaisés. Ce biais algorithmique (Danks et London, 2017) a suscité beaucoup d'attention. Il se produit lorsque l'algorithme d'apprentissage est formé à partir d'ensembles de données biaisées, et qu'il apprend



SOURCE: Crowdfunder | bit.ly/2IKzPcr



Lorsque l'intelligence artificielle apprend des humains, les choses se gâtent.

James Crowder*

ultérieurement « avec exactitude » les schémas de préjugés inhérents dans les données (voir par exemple Caliskan et coll., 2017). Dans certains cas, les représentations apprises à l'intérieur des algorithmes d'AA peuvent même exagérer ces biais (Zhao et coll., 2017). Les biais algorithmiques ont deux sources : des ensembles de données incomplets et des ensembles de données qui représentent des phénomènes de biais sociaux.

Les ensembles de données incomplets sont ceux qui ne sont pas représentatifs de la gamme complète des exemples potentiels. Par conséquent, un algorithme formé à partir d'un ensemble de données incomplet donnera un rendement insatisfaisant lorsqu'on lui présentera un exemple qui se situe en dehors de la portée des données disponibles. Par exemple, un algorithme de reconnaissance faciale qui n'est pas formé à partir d'un large échantillon de couleurs de peau pourrait ne pas fonctionner avec précision pour toutes les carnations. En fait, une chercheuse en IA ayant elle-même la peau foncée a découvert qu'un algorithme de reconnaissance faciale qui fonctionnait plutôt bien par ailleurs ne la reconnaissait que si elle mettait un masque blanc⁴⁵. De nombreux exemples comparables ont été rapportés⁴⁶.

Lorsque des décisions critiques sont prises en se fondant sur les données d'un algorithme formé à partir d'une base de données qui n'est pas représentative de la population entière d'utilisateurs, les résultats peuvent avoir une incidence sur la santé et le bien-être. La majorité des essais cliniques utilisent des critères de sélection très pointus qui excluent les femmes (et particulièrement les femmes enceintes), les personnes âgées, et les personnes atteintes de conditions autres que celles qui sont à l'étude. Par conséquent, les participants ont tendance à être des hommes blancs⁴⁷. Dans certains cas, il peut être difficile d'effectuer la généralisation des conclusions de ces études dans une population plus large, et les résultats sont susceptibles d'être faussés pour les personnes non représentées dans la recherche.

* bit.ly/2HFetvY Photo par Marius Masalar

De ce fait, les algorithmes d'IA formés à partir de ces données sont biaisés, et sont potentiellement dangereux pour les populations qui ont été exclues. Ce n'est que depuis quelques années que les bailleurs de fonds de la recherche médicale ont commencé à exiger la parité entre les sexes et une plus grande inclusion.

Les ensembles de données sur les phénomènes sociaux peuvent potentiellement refléter des biais sociaux, et les algorithmes formés à partir de ces derniers peuvent apprendre ces schémas de préjugés et les reproduire. À titre d'exemple:

Ciblage publicitaire

Des chercheurs ont constaté que les femmes ont moins de chances de se voir présenter des publicités ciblées sur des emplois très rémunérateurs⁴⁸, et les recherches en ligne de noms associés d'après les stéréotypes à des personnes à la peau noire ont 25 % plus de chances de déclencher des annonces au sujet de dossiers criminels et d'arrestations⁴⁹.

Services de police prédictifs

Un programme informatique utilisé aux États-Unis pour évaluer les risques de récidive de personnes aux prises avec le système de justice criminelle a signalé de façon erronée que les prévenus noirs présentaient un risque élevé presque deux fois plus souvent que les prévenus blancs⁵⁰.

On peut trouver d'autres exemples de préjugés dans des domaines tels que la présélection en vue d'une entrevue d'embauche, l'évaluation du rendement d'enseignants, ainsi que l'approbation d'emprunts et de prêts hypothécaires⁵¹.

Photo par Michal Parzuchowski sur Unsplash



Photo par Ryoji Iwata sur Unsplash

Dans certains cas, les biais algorithmiques peuvent aussi provenir de l'interaction entre les choix de conception et les données biaisées. Un bon exemple de cette situation, est le tristement célèbre robot conversationnel Tay, de Microsoft, un robot autonome qui a dû être retiré de Twitter après seulement 16 heures parce qu'il a commencé à proférer des gazouillis racistes et misogynes⁵². Dans un sens, il est raisonnable d'imputer une partie du blâme aux données, parce que les utilisateurs de Twitter tentaient délibérément de manipuler le robot conversationnel. Cependant, le fait que les algorithmes sous-jacents étaient neutres par rapport au contenu et à la signification des gazouillis absorbés à titre de données signifie que Tay a effectivement traité des expressions racistes ou misogynes comme s'il s'agissait simplement d'autres points de données à traiter.

Un autre exemple illustre la précarité des algorithmes d'apprentissage à contenu neutre. À partir de données dérivées du contenu partagé par les utilisateurs de Facebook et de leurs activités en ligne, les algorithmes de Facebook créent

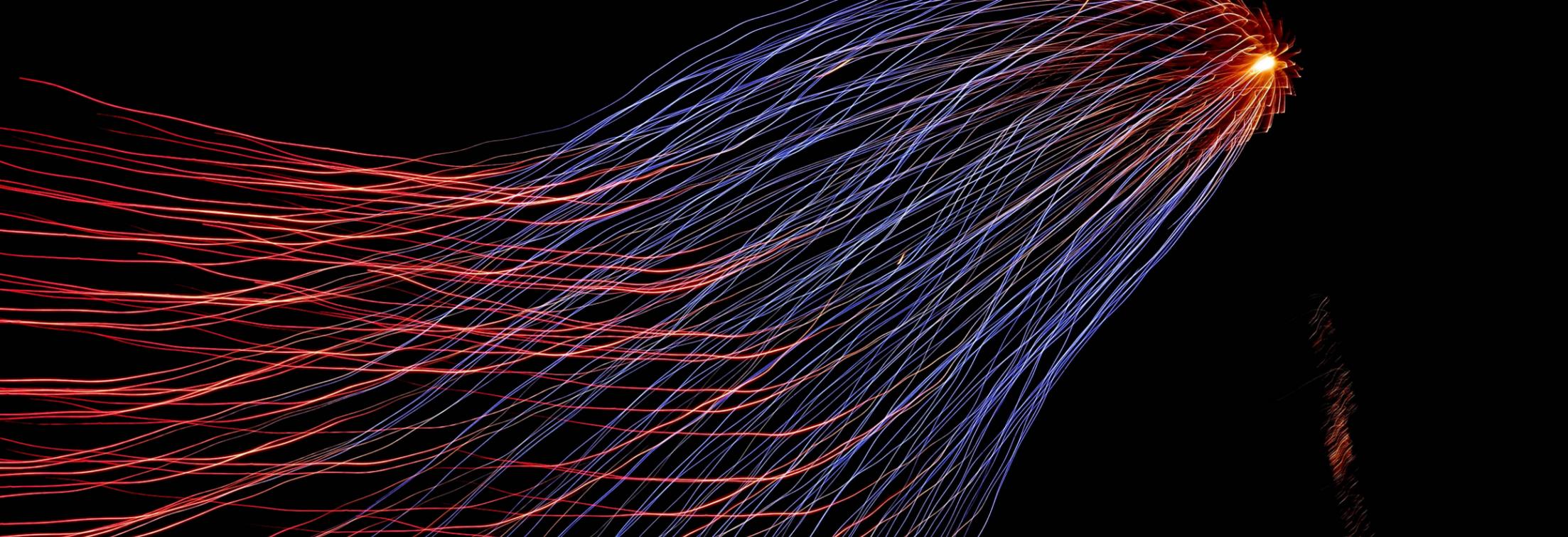


Photo par Hunter Harritt

automatiquement des catégories d'utilisateurs pour aider les publicitaires à cibler leurs annonces. Des catégories fondées sur des expressions durement antisémites comme « déteste les juifs » ont permis à des publicitaires d'atteindre quelque 2 300 utilisateurs identifiés comme ayant un intérêt pour ces sujets⁵³.

Comme l'illustrent ces exemples, le choix d'un algorithme d'apprentissage « neutre » peut entraîner des conséquences chargées de valeur. Dans le cas du robot conversationnel Tay, l'algorithme s'est révélé incapable de déterminer si une énonciation était positive ou négative du point de vue de valeurs sociales. Les concepteurs de tels systèmes font régulièrement des choix involontairement chargés de valeur qui éclairent sur la manière dont les algorithmes d'apprentissage vont se comporter par la suite.

Il ressort clairement que des biais peuvent facilement être introduits dans les systèmes d'IA et absorbés par ces derniers. Ce qui plaide en faveur de l'introduction de mesures de protection aux étapes de la conception et de l'opération de ces systèmes afin de veiller à ce que les résultats soient équitables et impartiaux. De plus, la transposition d'un algorithme d'apprentissage d'un contexte à un autre pourrait entraîner des résultats non désirés : les processus appris d'un système formé à partir de données nord-américaines, par exemple, pourraient produire des résultats inappropriés s'ils étaient mis en oeuvre en Amérique centrale, en Afrique ou en Asie. À ce titre, l'absence d'ensembles de données propres à un pays en particulier représente un obstacle évident à la production de systèmes d'IA équitables dans les pays du Sud.

Boîtes noires d'IA : transparence et responsabilisation

Un autre obstacle à la mise au point d'applications d'IA équitables (algorithmes d'AA, en particulier) tient au fait que, contrairement à d'autres logiciels, les choix et les actions de ces programmes peuvent être difficiles ou même impossibles à expliquer⁵⁴. Examiner les mécanismes d'un algorithme d'AA et identifier les biais codés n'est pas vraiment simple. La complexité des représentations abstraites apprises au moyen de l'encodage de schémas selon des millions de paramètres potentiels rend très difficile pour les humains d'analyser pourquoi un algorithme d'AA se comporte comme il le fait. L'apprentissage automatique présente un problème de transparence⁵⁵.

Étant donné que les systèmes d'IA peuvent prendre des décisions que l'on peut considérer comme injustes, inexactes ou contraires à l'éthique, la difficulté de déterminer comment ces décisions sont prises dans certaines circonstances particulières soulève plusieurs questions : Selon quels critères les décisions seront-elles considérées comme inappropriées ? Comment procédera-t-on pour déterminer la responsabilité en regard des conséquences de ces décisions ? Comment les personnes lésées pourront-elles demander réparation ?

Ces questions n'ont pas encore été abordées de façon concrète, et il faut se pencher sur elles tant dans les politiques que dans la législation. Le rapport intitulé *The Hundred Year Study* (Stone et coll., 2016) expose la question de la responsabilité civile et pénale comme suit :



Photo par Jose Lopez Franco sur Unsplash

” Étant donné que l’IA est organisée de manière à influencer directement sur le monde, et ce, même physiquement, la responsabilité en regard des dommages causés par l’IA prendra davantage de relief. La perspective que l’IA se comporte d’une manière que les concepteurs n’ont pas prévue remet en question l’hypothèse qui prévaut dans le droit de la responsabilité civile délictuelle selon laquelle les tribunaux ne peuvent réparer un préjudice que pour des dommages prévisibles. Les tribunaux pourraient attribuer de façon arbitraire la responsabilité à un acteur humain, et ce, même si la responsabilité se situe ailleurs pour des raisons d’équité ou d’efficacité. En revanche, les tribunaux pourraient refuser de chercher à qui appartient la responsabilité parce que le défendeur qui comparaît n’a pas pu, et ne pouvait pas, prévoir les préjudices occasionnés par l’IA. La responsabilité retomberait donc par défaut sur la victime innocente. Le rôle de la responsabilité associée aux produits – et de la responsabilité qui échoit aux entreprises qui fabriquent ces produits – est appelé à croître lorsque les acteurs humains deviennent moins responsables pour les actions d’une machine. (...).

Si des applications d’IA adoptent un comportement qui, s’il avait été commis par un humain aurait constitué un crime, les tribunaux et autres intervenants juridiques devront naviguer à travers les méandres de qui doit être tenu responsable et en fonction de quelle théorie.

Calo (2017:12) résume le vaste enjeu comme suit:

” L’objectif ultime de la conception de systèmes qui reflètent la justice et l’équité nécessitera des efforts interdisciplinaires considérables et pourrait se révéler l’une des questions de politique déterminantes de notre temps.

Pour concevoir des systèmes d’IA justes et équitables, l’une des approches à privilégier consiste à rendre les mécanismes internes des algorithmes plus transparents. Voici quelques-unes des nouvelles approches pour atteindre une plus grande transparence:



Vérification des algorithmes

Il s’agit d’un domaine naissant dans lequel les chercheurs inspectent et évaluent des algorithmes⁵⁶. Pour les algorithmes d’AA, cela pourrait vouloir dire visualiser ce que le logiciel a appris⁵⁷.

Explicabilité

Les concepteurs peuvent faire en sorte qu’au moins une partie du raisonnement qui sous-tend les décisions algorithmiques soit plus transparente⁵⁸. Dans la mesure du possible, les systèmes d’IA peuvent être développés avec la capacité de fournir le raisonnement qui sous-tend une décision⁵⁹.

Recours judiciaire

Rendre obligatoire l’ouverture du logiciel derrière des algorithmes d’AA critiques permettrait d’avoir accès à ce qui pourrait ne pas être disponible en raison des brevets et des contrats de licence de protection. Cela s’est produit récemment lorsqu’un juge de la cour fédérale des États-Unis a ordonné à un laboratoire médico-légal de rendre public son logiciel d’analyse des preuves génétiques afin que le code puisse être analysé⁶⁰.

Certaines administrations ont déjà pris des mesures concrètes pour imposer juridiquement la transparence dans le processus décisionnel de systèmes automatisés. À titre d’exemple, l’Union européenne a adopté une loi qui exige désormais que les décisions prises par des machines soient explicables. Les sanctions imposées aux entreprises incapables de se conformer pourraient se chiffrer en milliards de dollars⁶¹. Cependant, le libellé de la loi est sans doute suffisamment imprécis pour la rendre sans effet (Wachter et coll., 2017). Voilà qui fournit une intéressante cause-type pour les autres pays.

Bien entendu, il n’est pas clair si l’approche retenue par l’UE est faisable, compte tenu de l’extrême complexité que peuvent présenter parfois les algorithmes d’AA. Une autre approche pourrait consister à régir les processus décisionnels au moyen de l’IA sur le plan de l’optimisation de l’algorithme plutôt que sur celui des résultats en tant que tels. Par exemple, David Weinberger expose l’approche suivante : « Les véhicules autonomes pourraient être optimisés d’abord en vue de réduire les décès, ensuite de réduire les blessures, puis de réduire leurs répercussions environnementales, et ensuite de réduire la durée du trajet, et ainsi de suite. Les hiérarchies exactes des priorités sont des choses avec lesquelles les organismes de réglementation devront se colleter⁶² ».



La surveillance et la vie privée sont les deux côtés d'une même médaille

SURVEILLANCE ET ATTEINTE AU DROIT À LA VIE PRIVÉE

La surveillance et la vie privée sont les deux côtés d'une même médaille. La surveillance est définie comme « le traitement de renseignements personnels pour les fins de fournir des soins ou d'assurer le contrôle, d'influencer ou de gérer des personnes et des populations » [Traduction] (Lyon, 2010 : 108), alors que la vie privée consiste à conserver l'emprise sur ses propres renseignements personnels. La surveillance peut être menée pour des motifs relativement bénins, comme découvrir les préférences personnelles en vue de cibler des annonces et des promotions en ligne ou d'améliorer les services gouvernementaux. Mais elle peut aussi avoir des buts plus sinistres, comme suivre les allées et venues et les associés de rivaux politiques – des buts qui sont particulièrement préoccupants aujourd'hui, en raison de la montée de la surveillance étatique (Deibert, 2013) et des effets potentiellement paralysants qui y sont associés sur les libertés individuelles (Penney, 2017). Quel qu'en soit le but, cette pratique porte atteinte à la vie privée, et les individus disposent de moins en moins de moyens de l'éviter ou de l'empêcher.

Du point de vue de la vie privée, les applications d'IA sont préoccupantes parce que les puissants algorithmes de reconnaissance des formes peuvent extraire des renseignements personnels des données

Intelligence artificielle et vie privée

En 2017, dans une présentation au Comité spécial de l'intelligence artificielle de la Chambre des Lords du parlement britannique, *Privacy International* a désigné quatre buts des applications d'IA qui sont particulièrement préoccupants sur le plan de la vie privée:

Identifier et suivre des individus;

Évaluer des individus ou des groupes et leurs comportements ou prévoir ces derniers;

Prendre automatiquement des décisions ou contribuer à la prise de décisions déterminantes au sujet de personnes ou de leur environnement;

Générer, recueillir et partager des données.

Pour en apprendre davantage, voir: [privacyinternational.org/node/1525](https://www.privacyinternational.org/node/1525)



Les techniques d'IA survolent les capacités de surveillance déjà impressionnantes

Photo par Serge Kutuzov sur Unsplash

disponibles (Calo, 2017; *Privacy International*, 2017). L'utilisation croissante de la surveillance a entraîné la création de vastes dépôts de données brutes sur les activités des individus, tant en ligne que dans le monde réel⁶³, et les techniques d'IA simplifient le processus de triage à travers l'infinité de points de données. Les ensembles de données peuvent être explorés et mis en correspondance, des schémas subtils peuvent être repérés, des inférences peuvent être établies et des conclusions tirées, et tout cela sans intervention humaine active. La surveillance est survoltée par les algorithmes d'IA qui traitent les données plus rapidement qu'il n'a jamais été possible de le faire, et qui repèrent des schémas trop subtils pour que les analystes humains puissent les découvrir.

Les résultats sont l'érosion de la vie privée, un concept que certains cercles technologiques considèrent déjà comme désuet. Cette érosion est particulièrement préoccupante en regard d'une perspective de développement social et politique, étant donné que la vie privée est « le pilier de valeurs individuelles indispensables comme la dignité humaine, l'autonomie individuelle, la liberté d'expression, la liberté d'association et la liberté de choix » [Traduction] (*Privacy International*, 2017).

Les exemples de la manière dont les applications d'IA peuvent envahir la vie privée des individus sont légion. Dans un cas, une analyse des amis Facebook d'un individu a permis de déterminer avec exactitude l'orientation sexuelle de cette personne⁶⁴. Plus récemment, l'analyse des mentions « j'aime » sur Facebook à l'aide de l'apprentissage automatique a permis de déterminer la race, les préférences sexuelles et l'allégeance politique d'un individu avec un degré élevé de précision⁶⁵.

Les gouvernements se servent des techniques d'IA pour survolter leurs capacités de surveillance déjà impressionnantes. Par exemple, les logiciels de reconnaissance faciale confèrent aux systèmes de télévision en circuit fermé la capacité de suivre des individus pendant leurs déplacements dans le paysage urbain⁶⁶. Les algorithmes de reconnaissance faciale peuvent reconnaître les visages, et ce, même s'ils sont flous ou pixelisés, ce qui permet de déjouer certaines technologies actuelles en matière de protection de la vie privée⁶⁷. Dans le même ordre d'idées, les algorithmes de reconnaissance faciale ont réussi à identifier des personnes qui avaient tenté de masquer leur identité en portant une casquette, un foulard

ou des lunettes (Singh et coll., 2017). La Chine a récemment utilisé cette technologie pour arrêter 25 suspects lors d'un festival de la bière, dont un homme qui échappait aux autorités depuis 10 ans⁶⁸. L'usage de ces systèmes pour suivre des criminels ou des terroristes fournit un bien public, mais là où les mesures de protection contre de telles pratiques sont faibles ou inexistantes, les mêmes outils peuvent facilement se retourner contre les rivaux politiques, les concurrents en affaires ou d'autres personnes. Les régimes autoritaires dans les PFR-PRI ont déjà investi dans de puissantes technologies de surveillance et les ont tournées vers les dissidents et les militants pour la protection des droits de la personne⁶⁹, entre autres. Nous nous attendons à ce que cette tendance se maintienne, et à ce qu'elle ait recours à des outils encore plus puissants.

L'absence actuelle de transparence quant aux renseignements personnels qui sont recueillis auprès des utilisateurs en ligne et à la manière dont ces données sont utilisées – habituellement sans aucune forme de consentement – rend difficile pour les individus de protéger leur vie privée en ligne. Maintenant que l'Internet des objets devient une réalité, et que des objets de tous les jours (véhicules, réfrigérateurs, etc.) se connectent numériquement, de tous nouveaux horizons de données deviendront disponibles en vue d'être utilisés. Les enjeux relatifs au contrôle de ses propres données et la possibilité d'obtenir réparation dans les cas de mauvaise utilisation ou de résultats inappropriés deviendront encore plus urgents alors que des applications d'IA prennent des décisions déterminantes sur la foi de ces données⁷⁰.



MISE À MAL DE LA DÉMOCRATIE ET DE L'AUTODÉTERMINATION POLITIQUE

Le processus démocratique repose sur le droit des citoyens d'accéder librement à l'information qui nourrit les choix politiques. Les risques inhérents à une surveillance accrue et à une réduction de la vie privée décrits ci-dessus mettent à mal ces principes démocratiques fondamentaux. Il existe au moins deux autres mécanismes au moyen desquels l'IA peut potentiellement contribuer à mettre à mal la démocratie : la propagande ciblée et la manipulation, ainsi que la fragmentation sociale.

S'il ne s'agit pas de phénomènes nouveaux, il reste que la confluence des algorithmes d'IA avec d'autres changements relativement récents survenus dans l'écosystème en ligne contribue à rendre la désinformation et la manipulation beaucoup plus efficaces sur une plus grande échelle que jamais auparavant. Premièrement, la centralisation de l'attention et des activités canalisées par les plateformes en ligne de seulement quelques entreprises comme Google, Twitter et Facebook⁷¹. La situation est comparable en Chine, où des sociétés comme Tencent dominent l'expérience en ligne des citoyens. Selon une recherche menée en 2017 par le Pew Research Center, près de la moitié des Américains se renseignent sur l'actualité « parfois » ou « souvent » dans les médias sociaux⁷². Ces plateformes recueillent de vastes quantités de données sur les utilisateurs à partir desquelles les algorithmes d'IA extraient des renseignements personnels; les plateformes vendent ensuite un accès très ciblé à ces utilisateurs.

De plus, de récents progrès dans l'étude de la manipulation du comportement ont accentué l'impact de campagnes destinées à influencer les utilisateurs de médias sociaux. Le résultat est une prolifération de campagnes de désinformation et de manipulation très personnalisées et efficaces (propagande de précision)⁷³ qui représentent une réelle menace pour l'autodétermination politique⁷⁴.

Nous pouvons nous attendre, à mesure que les techniques d'IA vont progresser, à ce que leur utilisation en vue d'automatiser la manipulation et la tromperie sociales ne puisse que gagner en sophistication, élargissant la portée des menaces potentielles (Brundage et coll., 2018).

À mesure que les techniques
d'IA vont progresser,
leur utilisation en vue
d'automatiser la manipulation
et la tromperie sociales ne
fera qu'augmenter

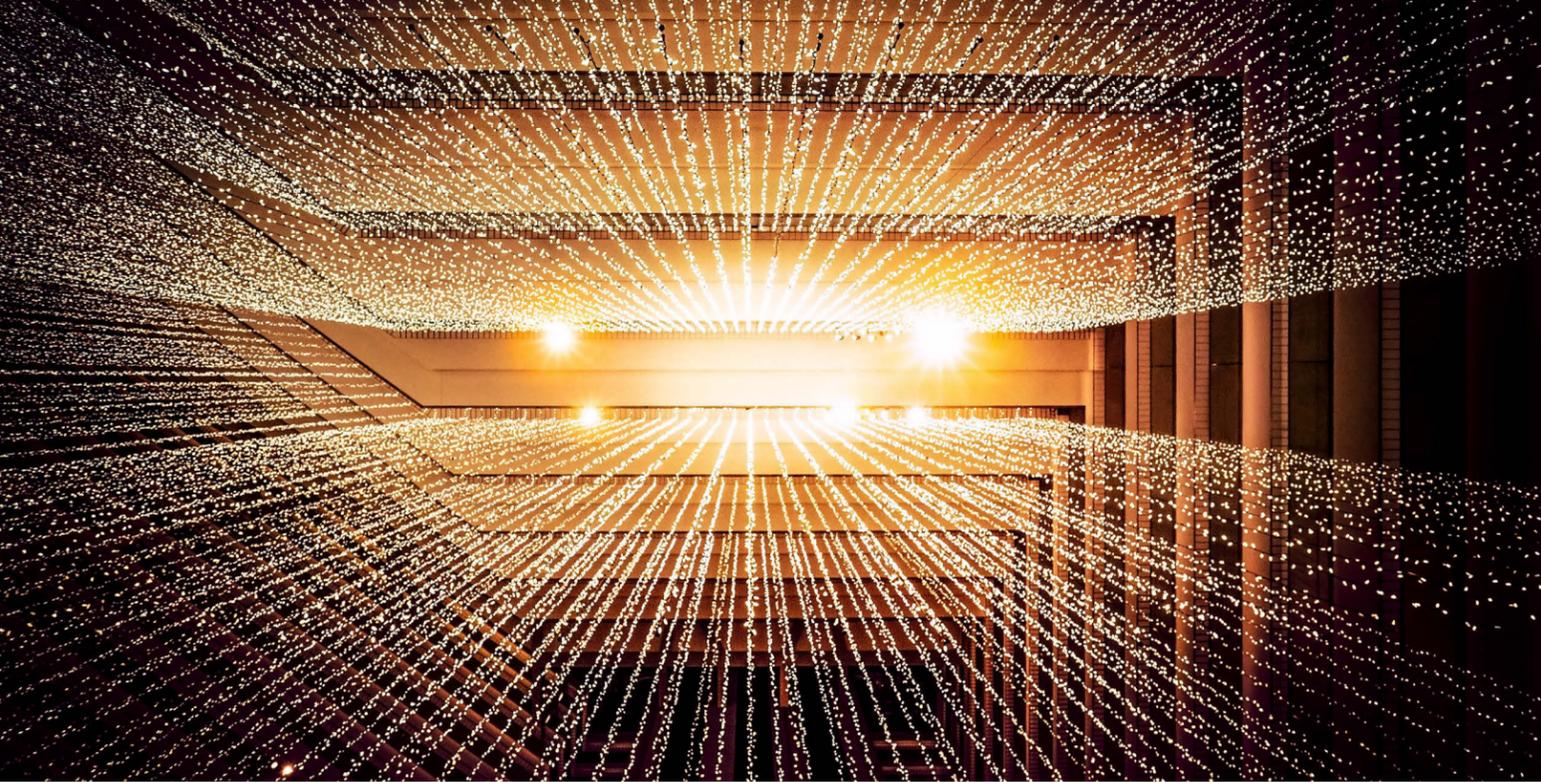


Photo par Joshua Sortino sur Unsplash

Propagande et manipulation

Les techniques d'IA sous-tendent l'émergence de la « propagande virtuelle » (Bolsover et Howard, 2017), qui peut prendre de nombreuses formes, dont celle de robots conversationnels, de fausses nouvelles et de désinformation et de manipulation très ciblées. Comme le fait remarquer Howard, « Algorithmes et fausses nouvelles vont de pair⁷⁵ ».

Les élections présidentielles américaines de 2016 sont devenues un exemple tristement célèbre de l'utilisation de la désinformation, au cours desquelles des robots conversationnels activés par l'IA sur Twitter et des fausses nouvelles étaient en pleine action. On estime que 33 % des gazouillis pro-Trump et 20 % des gazouillis pro-Clinton publiés durant le premier débat présidentiel ont été créés par des robots dotés d'une IA⁷⁶ qui ont gonflé artificiellement le volume de trafic et le soutien apparent aux candidats. Une profusion de désinformation très ciblée a également été distribuée, sous la forme de faux articles de nouvelles qui ont contribué à influencer encore plus les opinions des gens. Facebook estime qu'entre janvier 2015 et août 2017, la désinformation russe pourrait avoir atteint autant que 146 millions d'utilisateurs⁷⁷. Des initiatives comparables de désinformation et de propagande ont été utilisées dans une tentative d'influencer d'autres campagnes politiques.

Ces tactiques peuvent aussi faire progresser d'autres genres de programmes. Au Myanmar, par exemple, on peut avancer que Facebook a été un vecteur essentiel en vue de la transmission du sentiment anti-Rohingya, ayant contribué



à l'épuration technique dans ce pays⁷⁸. Dans le même ordre d'idées, il ne fait aucun doute que les fausses nouvelles diffusées sur les médias sociaux ont contribué à une hausse du nombre de personnes qui, aux États-Unis, s'interrogent à savoir si les changements climatiques sont dus à l'intervention humaine, et s'il existe un lien entre le tabac et le cancer⁷⁹.

Les progrès dans le domaine de l'intelligence artificielle contribueront seulement à rendre la désinformation encore plus convaincante. Tout comme les logiciels de retouche d'images permettent de façon relativement simple de créer de fausses images convaincantes, les nouveaux logiciels dotés d'IA apportent cette facilité de modifier des bandes vidéo et audio, ce qui pourrait suralimenter les campagnes de fausses nouvelles et de désinformation⁸⁰. À titre d'exemple, la jeune entreprise de Montréal Lyrebird a récemment diffusé sur YouTube une capsule audio dans laquelle on entend, apparemment, les voix de Barack Obama et de Donald Trump qui discutent de l'entreprise⁸¹. La capsule audio, entièrement fabriquée, utilise des voix artificielles pour reproduire le style de discours des deux hommes, mais un auditeur naïf pourrait prendre la capsule audio pour un enregistrement d'une conversation réelle traitée pour sembler avoir été informatisée. Avec ce logiciel, l'entreprise ouvre la porte à la possibilité de produire de faux vidéos à l'aide de voix synthétisées, de la gestuelle, de la tonalité affective, et ainsi de suite. Dans la même optique, Adobe a mis au point un outil qui permet aux utilisateurs de modifier un discours enregistré de la même manière que l'on corrige un texte. Il permet même d'insérer des mots qui n'avaient pas été prononcés⁸². (Voir aussi Suwajanakorn et coll., 2017.)⁸³

Photo par Mike Kononov sur Unsplash



Et fait encore plus préoccupant, on peut aujourd'hui trouver en accès libre des outils informatiques pour fabriquer des « deepfakes » ou faux montages vidéo. En se servant de la puissance de l'apprentissage profond, cet outil permet à quiconque de remplacer un visage par un autre dans une vidéo. La seule exigence est un nombre et une variété suffisants de vidéos de la personne que vous souhaitez remplacer dans la vidéo. Étant donné la capacité des fausses histoires de se répandre rapidement parmi les gens, et la facilité avec laquelle les humains forgent de faux souvenirs, et plus particulièrement ceux qui viennent confirmer nos préjugés, la possibilité d'abus est énorme⁸⁴.

Les manipulations de citoyens peuvent s'effectuer non seulement au moyen de la propagande et de la désinformation, mais aussi par l'entremise de la consolidation et de l'analyse automatisées de données massives recueillies quotidiennement sur diverses personnes. Le Citizen Score, proposé par la Chine, en est un exemple extrême quoique tiré du monde réel. Établi à partir de facteurs comme l'historique financier d'une personne, ses activités en ligne, son appartenance politique et même le comportement de ses amis et connaissances, ce score unique servira à déterminer l'accès de cette personne à un emprunt, un logement, un emploi et des visas pour voyager. S'il est entièrement mis en oeuvre comme prévu en 2020, il serait un mécanisme pour manipuler et contrôler le comportement d'une population entière de milliards de personnes⁸⁵.

Photo par Mike Enerio

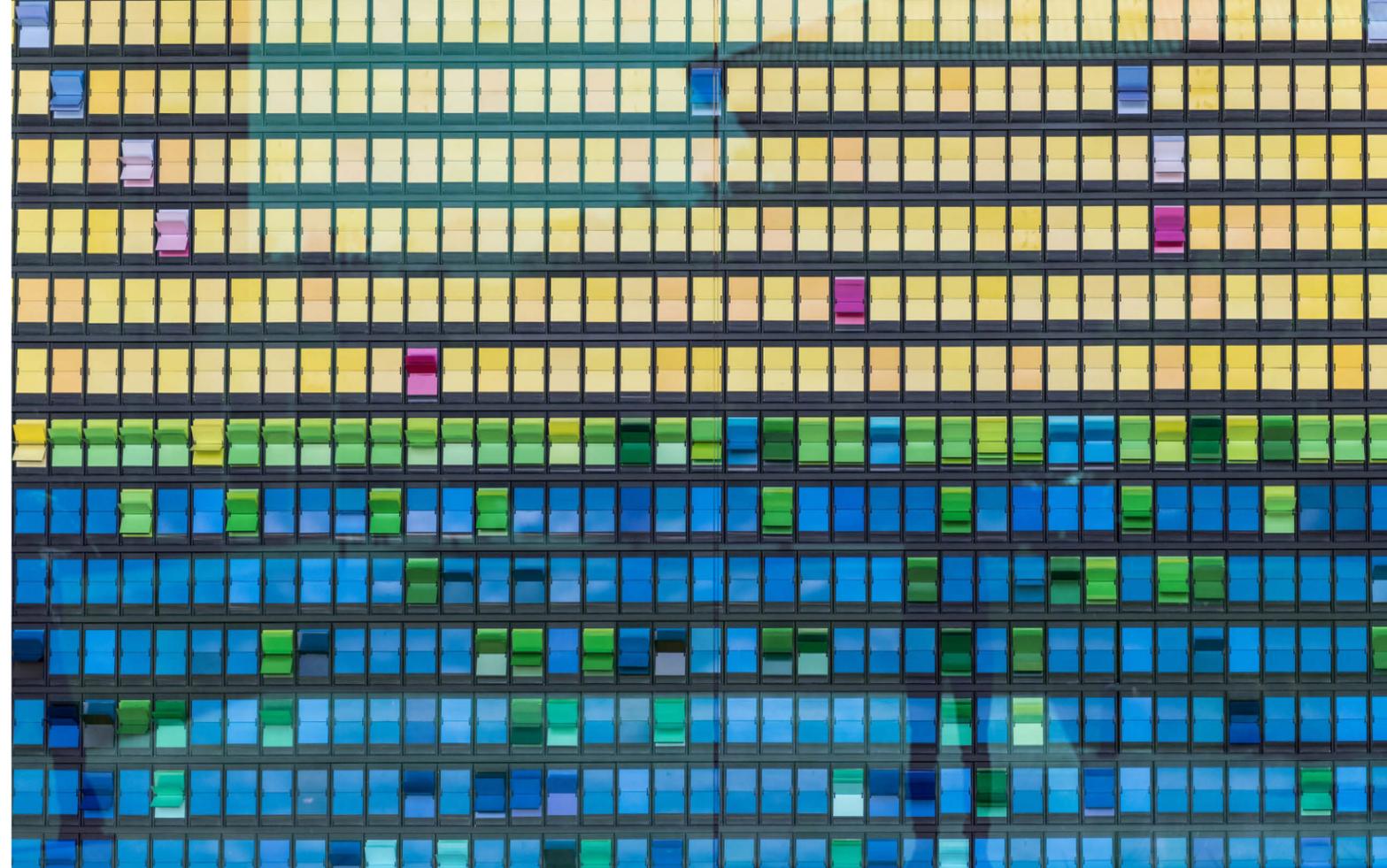


Photo par AI x sur Unsplash

Fragmentation

Pris ensemble, les efforts pour fausser et manipuler l'opinion publique, ainsi que la prédominance dans beaucoup de régions d'un seul ou de seulement quelques moteurs de recherche et plateformes de médias sociaux, peuvent limiter le discours public. L'adaptation à l'utilisateur et la personnalisation utilisées pour améliorer les résultats du marketing peuvent aussi rétrécir le champ des nouvelles et des opinions que les utilisateurs rencontrent, ce qui entraîne un effet de « caisse de résonance ». Ainsi, les opinions politiques et sociales deviennent autorenforçants, et les différents groupes finissent par avoir moins d'interactions et de terrains d'entente. Les débats deviennent polarisés, et la collaboration et les compromis deviennent difficiles. Cette tendance est présente dans l'état de la politique aux États-Unis. Au fil du temps, ce processus peut mener à la « fragmentation, et même éventuellement à la désintégration, de la société⁸⁶. »



L'automatisation ne se limitera plus au seul secteur de la fabrication

AUTOMATISATION, PERTE D'EMPLOIS ET PERTE DE RECETTES FISCALES

L'automatisation n'est pas un phénomène nouveau: elle contribue à la perte générale des emplois dans le secteur manufacturier depuis les années 1990. Cependant, avec l'utilisation croissante de systèmes d'apprentissage automatique et d'intelligence artificielle dans pratiquement tous les secteurs de l'économie, l'automatisation ne se limitera plus au seul secteur de la fabrication. Beaucoup d'emplois et de professions vus traditionnellement comme nécessitant la participation et l'expertise humaines seront touchés. Par exemple, un algorithme d'apprentissage profond compétent peut aujourd'hui déceler le cancer aussi bien qu'un dermatologue formé à cet effet⁸⁷. Le côté positif, ce sont les avantages que cela comporte pour les collectivités où les ressources sont limitées et qui manquent de médecins ou de techniciens compétents. Le côté négatif, tient à la menace réelle qu'une automatisation de ce genre présente pour des emplois qui sont considérés comme hautement qualifiés.

Bon nombre de ces emplois peuvent être automatisés, en partie ou entièrement, ce qui réduira le besoin de faire appel à des travailleurs humains. Une combinaison d'algorithmes d'apprentissage profond et de traitement de langage naturel peut potentiellement accomplir une bonne partie ou la totalité du travail que font les tuteurs, les agents de voyage, les préparateurs de déclarations du revenu, les conseillers en santé personnels, les chercheurs juridiques, les conseillers financiers, les aides de bureau et les chauffeurs. Le ratio déjà faible d'emplois dans bien des pays du Sud peut être affecté par d'autres pertes d'emplois attribuables à l'intelligence artificielle.

Les estimations de la mesure dans laquelle l'automatisation dotée de l'IA influera sur l'emploi

Photo par Giovanni Randisi

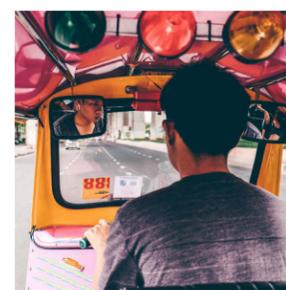
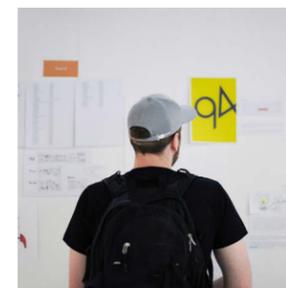
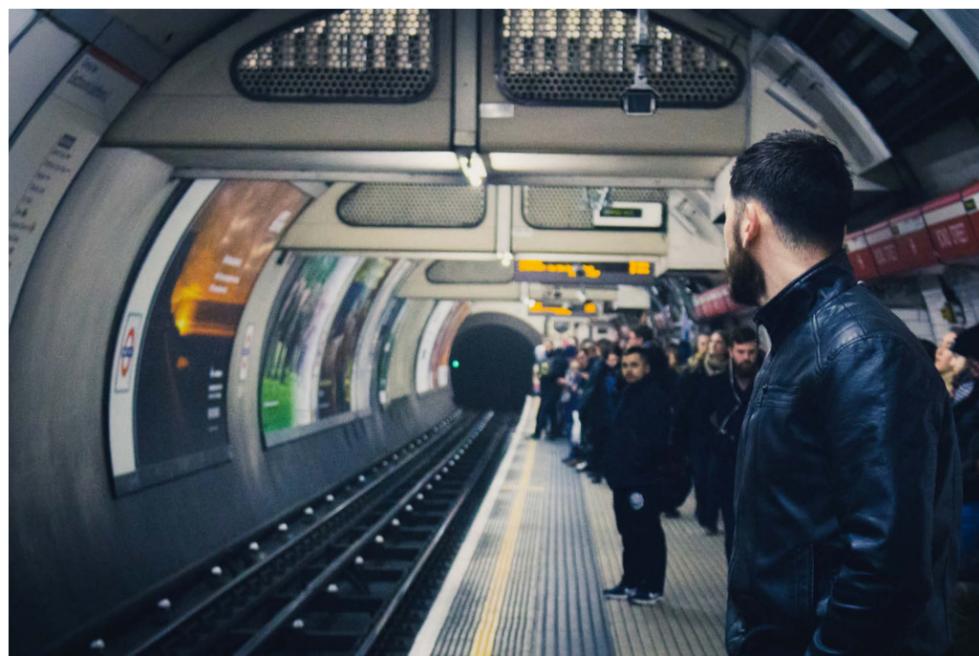


varient considérablement. Une étude réalisée à l'Université d'Oxford a indiqué que 47 % des emplois pourraient être automatisés d'ici 10 à 20 ans (Frey et Osborne, 2013), tandis que l'Organisation de coopération et de développement économiques (OCDE) arrivait à 9 %, en se fondant sur le fait qu'un nombre important d'emplois ne peuvent être entièrement automatisés (Arntz et coll., 2016). Selon le McKinsey Global Institute, « l'adaptation des technologies actuellement en démonstration pourrait toucher 50 % des heures de travail dans l'économie mondiale » (Manyika et coll., 2017).

Un tel changement cataclysmique pourrait désavantager considérablement les pays moins sophistiqués sur le plan technologique. Un autre rapport estimait que l'automatisation menacerait 55 % des emplois en Ouzbékistan et jusqu'à 85 % des emplois en Éthiopie, et que les économies émergentes comme la Chine et l'Inde pourraient voir au moins 70 % de leurs emplois à risque élevé de disparaître⁸⁸. Dans certaines industries, le risque potentiel est vu comme relativement imminent. En juillet 2017, craignant des pertes d'emplois massives, le ministère indien des transports a banni les véhicules autonomes sur les routes du pays⁸⁹.

L'incidence sur l'emploi de l'intelligence artificielle variera selon le secteur économique, le pays et la région du monde visés, mais sans égard aux chiffres précis, beaucoup pensent que l'expansion de l'automatisation grâce à l'IA entraînera une vaste perturbation des marchés du travail, et un changement radical dans la nature même du travail. Par exemple, un rapport du McKinsey

Photo par Hello Lightbulb sur Unsplash



Photos par:
rawpixel.com,
Bethany Legg,
Hanny Naibaho
sur Unsplash

Global Institute a montré qu'avec les technologies actuelles, « 60 % de toutes les professions comportent au moins 30 % d'activités automatisables » (Manyika et coll., 2017 : 32). De plus, on a calculé que « l'adaptation de technologies d'automatisation actuellement en démonstration pourrait toucher 50 % de l'économie mondiale, ou 1,2 milliard d'employés et 14,6 billions de dollars en salaires » (ibid).

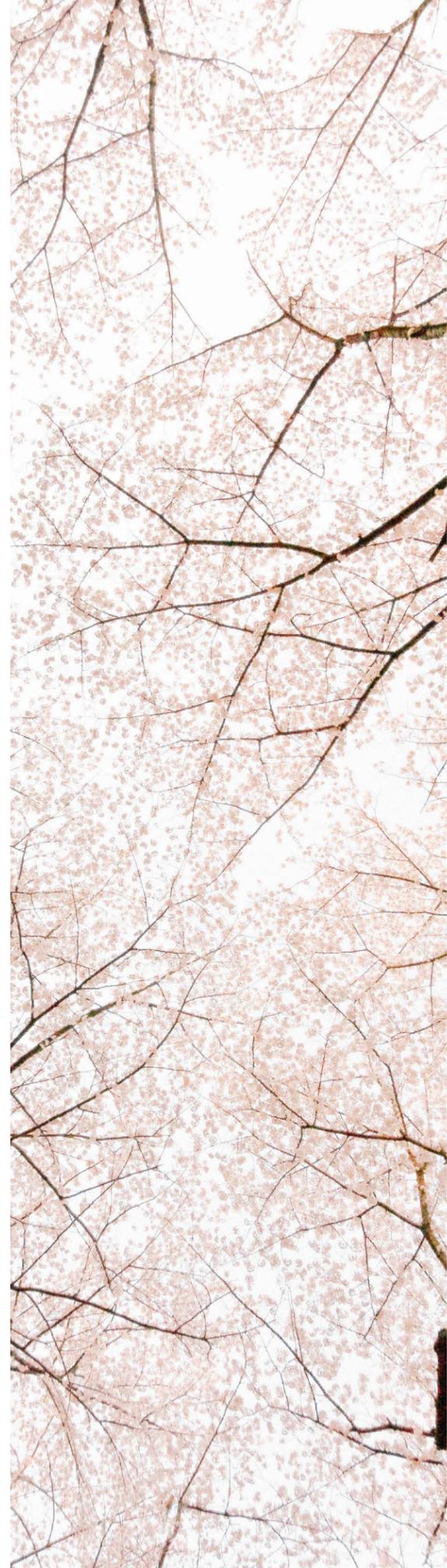
Cela pose des défis pour les modèles de développement économique : « Si la productivité dans le domaine de la fabrication a traditionnellement permis aux pays en développement de combler l'écart avec les pays riches, l'automatisation est susceptible d'influer négativement sur leur capacité de le faire, et de nouveaux modèles de croissance seront nécessaires⁹⁰ ». Nous constatons déjà des exemples d'entreprises qui « relocalisent » la fabrication en Europe et en Amérique du Nord parce que la productivité accrue et la diminution du besoin en main-d'œuvre résultant de l'automatisation compensent pour les frais d'expédition et les délais liés au transport des biens à partir des usines en Asie.

De plus, l'échéancier pour les pertes d'emplois imminentes pourrait être beaucoup plus court que ce que l'on a déjà vu avec les changements technologiques précédents. Dans un rapport de l'industrie publié en 2016, on prévoyait que 6 % de tous les emplois aux États-Unis pourraient être automatisés d'ici 2021, les représentants du service à la clientèle et les employés de centres d'appel figurant parmi les premiers touchés⁹¹. Jack Ma, le fondateur d'Alibaba et un innovateur remarqué dans le commerce de détail en ligne, a déclaré que l'automatisation grâce à l'IA entraînera « plus de peine que de joie » dans les décennies à venir⁹². À mesure que les pertes d'emplois remontent l'échelle des compétences, les experts entrevoient un déclin accéléré de la classe moyenne⁹³. De plus, on s'attend à ce que le fardeau des pertes d'emplois prévues affecte plus lourdement les femmes (Forum économique mondial, 2016).

Il pourrait bien en résulter de l'instabilité sociale et des bouleversements, et les pays du Sud pourraient être les plus durement touchés. Cela s'explique en partie parce que bon nombre de ces pays voient l'industrialisation future comme un moyen d'atteindre la croissance économique, et d'autres pays émergents ont misé largement sur la fabrication pour les progrès qu'ils ont réalisés. Comme nous l'avons déjà mentionné, l'automatisation accrue est en voie d'accélérer les pertes d'emplois dans ce secteur. Même si l'avenir ne risque pas d'être marqué par la « désindustrialisation », il reste qu'il est possible que l'industrialisation ne soit pas la voie du succès qu'elle a été dans le passé. De plus, les populations de jeunes adultes sans emploi ou sous-employés peuvent être instables, comme l'a démontré le phénomène du Printemps arabe⁹⁴.

Parce qu'ils sont dépourvus des filets de sécurité sociale présents dans la majorité des pays à revenus élevés, beaucoup de pays du Sud sont également moins en mesure d'amortir le choc de pertes d'emplois considérables⁹⁵. Le recours aux employés contractuels contribuera à ajouter aux pressions en raison de la demande accrue de soutien de la part des gouvernements, et de la diminution des recettes fiscales. Ces résultats se conjuguent pour entraîner des risques encore plus grands, et pourraient contribuer à créer de l'instabilité dans les pays du Sud. Les bouleversements actuels que connaît la politique américaine pourraient être un signe avant-coureur de ce qui s'en vient⁹⁶.

Il est important de mentionner que la discussion qui précède parle d'un risque potentiel, et que ces perspectives pessimistes sont fondées sur des estimations et des prévisions



sans beaucoup de données probantes à l'appui⁹⁷. En effet, ce n'est pas la première fois que l'on prédit des pertes d'emplois massives en raison de l'automatisation. Par exemple, à l'arrivée des guichets automatiques, dans les années 1970, on avait fait valoir que les banques n'auraient plus besoin de caissiers. Des arguments contraires purement spéculatifs ont aussi été évoqués en regard de l'IA. Ainsi, plutôt que de causer des pertes d'emplois nettes, l'IA pourrait entraîner un revirement dans la nature et la portée du travail et des emplois, par exemple en faisant en sorte que des robots viennent compléter et augmenter la main-d'oeuvre, et en mettant davantage l'accent sur les tâches très spécialisées et mieux rémunérées (International Federation of Robotics, 2017). Dans l'ensemble, même si on peut s'attendre à ce que l'intelligence artificielle contribue à un changement social, la nature ultime de ce changement est inconnue.

Photo par Mike Wilson





La cybersécurité est une nouvelle venue dans la course aux armements

CYBERSÉCURITÉ ET CYBERCRIME

La cybersécurité est une nouvelle venue dans la course aux armements. Des intervenants malveillants (criminels ou pirates appuyés par des États) mettent au point et lancent continuellement de nouvelles formes de logiciels malveillants et de méthodes d'attaque pour accéder aux systèmes informatiques pour voler des données, des identités ou des fonds électroniques ou pour prendre ces systèmes et leurs données en otage contre une rançon. Les spécialistes de la cybersécurité, quant à eux, mettent au point de nouvelles méthodes de défense, de détection et de chiffage afin d'assurer la sécurité de leurs ordinateurs et de leurs données. L'intelligence artificielle ne changera pas ce scénario d'ensemble; elle sera plutôt utilisée des deux côtés, à titre de moyen d'améliorer les cyberattaques et de moyen de les combattre⁹⁸.

L'intelligence artificielle accroîtra la sophistication des outils malveillants disponibles et accroîtra la rapidité et l'efficacité des criminels. Les logiciels malveillants améliorés grâce à l'IA seront en mesure de percer les remparts informatiques plus rapidement qu'un humain ne pourrait le faire, et d'autres outils d'IA seront en mesure de s'infiltrer automatiquement et profondément dans les millions d'enregistrements de bases de données de renseignements personnels afin de faciliter le vol d'identité. En revanche, les logiciels de sécurité améliorés grâce à l'IA seront encore meilleurs pour détecter les attaques, y réagir et en tirer des leçons. Google travaille au développement d'un détecteur de logiciels malveillants fondé sur l'AA pour son système d'exploitation Android⁹⁹.

Ce qui est susceptible de se produire, c'est que les capacités évoluées des cybercriminels qui manient des outils d'IA exposeront les individus, les organisations et les gouvernements qui n'ont pas accès à des mesures de sécurité à jour à des risques encore plus grands d'une attaque réussie. De plus, au fil du temps ces outils vont probablement réduire les coûts de l'engagement dans des cyberattaques à grande échelle, ce qui devrait probablement entraîner une expansion à la fois du groupe d'acteurs qui entreprennent ces attaques et du groupe de cibles (Brundage et coll., 2018).

Photo par Patryk Gradys sur Unsplash



Ces prévisions s'additionnent
pour produire un scénario assez
désastreux marqué par des
inégalités accrues jumelées à
des perturbations économiques
et à l'agitation sociale

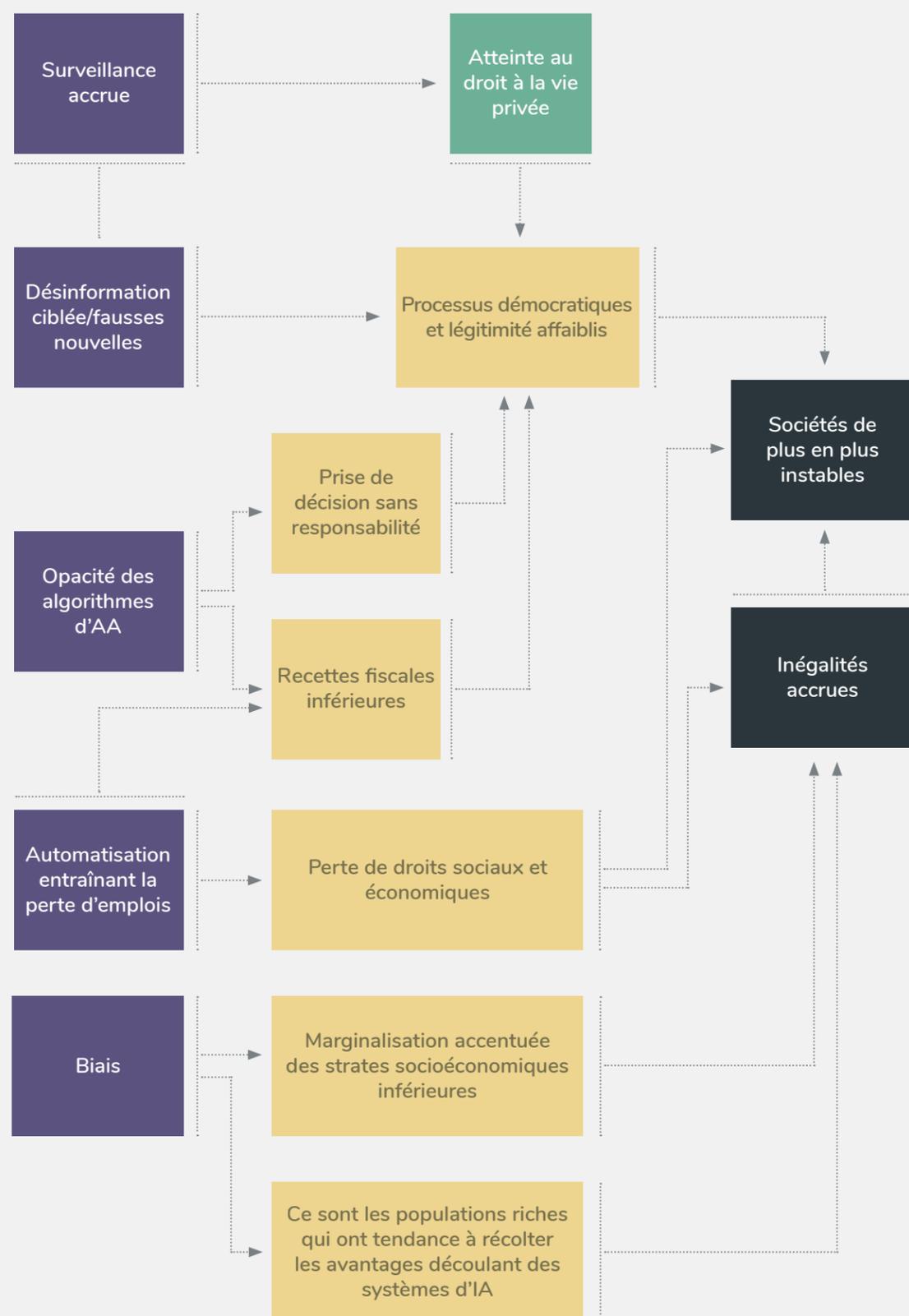
EXTRAPOLATION DANS LE FUTUR: RISQUES LIÉS À L'INTELLIGENCE ARTIFICIELLE DANS LES PAYS DU SUD

Malgré le potentiel de l'intelligence artificielle d'aider à relever de pressants défis en matière de développement, les risques qui lui sont associés (dont il a été question ci-dessus) et les leçons tirées de plus de 20 ans de recherche sur les technologies numériques dans les pays du Sud fournissent amplement de sujets d'inquiétude au sujet de l'application de l'IA. Tous les deux ou trois ans, une nouvelle technologie voit le jour que l'on qualifie de la « prochaine révolution » en matière de développement : ordinateurs, Internet, téléphones mobiles, mégadonnées et chaînes de blocs, pour n'en nommer que quelques-unes. En extrapolant à partir d'expériences passées¹⁰⁰, nous pouvons imaginer un avenir pour l'intelligence artificielle en matière de développement qui se déroulera un peu de la même manière que d'autres technologies l'ont fait :

- ▶ Les applications d'IA entraîneront des avantages spectaculaires sur le plan social, plus précisément des progrès dans les soins de santé, l'éducation et l'efficacité économique. Cependant, **ces avantages seront probablement répartis inégalement.**
- ▶ **Le fossé de l'IA continuera d'être un obstacle** au développement et au déploiement plus inclusifs de l'IA pendant encore un certain temps, limitant les avantages pour les collectivités habituellement mal desservies ou marginalisées.
- ▶ Certaines applications d'IA seront construites avec des **biais** ignorés ou non reconnus qui **reproduiront et aggraveront la marginalisation sociale.**



FIGURE 3 Les risques liés à l'IA et comment ils pourraient entraîner des répercussions sociales négatives.



- ▶ L'intelligence artificielle entraînera **l'accroissement de la surveillance et de l'atteinte au droit à la vie privée** en raison des activités d'intervenants des secteurs tant public que privé. Cela risque d'influer de façon disproportionnée sur les populations marginalisées et économiquement défavorisées.
- ▶ **Des intervenants malveillants utiliseront les techniques d'IA avec une sophistication accrue pour faciliter la criminalité, la mésentente sociale et l'agitation politique.**

Dans l'ensemble, ces prévisions s'additionnent pour produire un scénario assez désastreux marqué par des inégalités accrues jumelées à des perturbations économiques, à l'agitation sociale et, dans certains cas, à l'instabilité politique (voir la Figure 3) à moins que des mesures ne soient prises. En effet, c'est la nature double de l'intelligence artificielle – selon laquelle certains de ses avantages directs (par exemple l'efficacité économique) peuvent simultanément freiner les progrès dans d'autres domaines (par exemple l'emploi et la stabilité sociale) – qui en fait une technologie tellement perturbatrice. Une thèse un peu dystopique reposerait sur deux suppositions: 1) le caractère fondamental de l'IA et son intégration imminente à d'autres technologies et dispositifs font en sorte que les dommages potentiels de l'IA devraient probablement être ressentis beaucoup plus rapidement et massivement que ce que nous avons vu dans le passé; 2) le contexte actuel dans les pays du Sud fait en sorte que les dommages sont plus susceptibles de se produire.

Heureusement, probabilité n'équivaut pas à inévitabilité. Si nous agissons pour nous prémunir contre les défis posés par l'intelligence artificielle, nous pourrions éviter ou atténuer des résultats négatifs on ne peut plus prévisibles en habilitant les pays du Sud à tirer pleinement parti de son potentiel positif. L'avenir de l'intelligence artificielle est déjà à nos portes. La question est de savoir si nous serons prêts à l'accueillir.

Photo par Ben Neale sur Unsplash



Conclusions et recommandations

Même si dans l'ensemble le présent livre blanc semble adopter une attitude pessimiste face à l'avenir de l'intelligence artificielle dans les pays du Sud (ou à l'IA pour le développement), il a pour but de se montrer ni pessimiste, ni optimiste, mais plutôt prudent et encourageant. Il ne fait aucun doute que les technologies d'IA seront d'ordre transformationnel. Des avancées spectaculaires seront réalisées, une richesse extraordinaire sera créée et bon nombre de nos structures sociales et institutionnelles seront transformées. Cependant, nous devons poser la question : des existences seront améliorées, en effet, mais pour qui au juste ? Les libertés politiques et économiques connaîtront des avancées, mais pour qui ? Pour qui la privation systématique de ces libertés sera-t-elle adoucie ? La conclusion de la présente communication est que, si nous continuons à aller aveuglément de l'avant, nous devons nous attendre à voir les inégalités s'accroître, de concert avec la perturbation économique, l'agitation sociale et, dans certains cas, l'instabilité politique, et ce sont les gens défavorisés et sous-représentés sur le plan technologique qui subiront les pires effets.

Cette sombre prévision tire son origine de l'entrelacement de deux éléments : la nature des applications d'IA, et les prévisions des incidences des applications d'IA dans le contexte mondial actuel. Ce qui est préoccupant, c'est la dynamique avec laquelle nos ensembles actuels d'institutions et de cultures façonnent l'évolution des technologies, et comment, à leur tour, ces technologies façonnent ces institutions et ces cultures.

Photo par Robert Haverly sur Unsplash



Comme nous l'avons mentionné, l'intelligence artificielle fait référence à une large classe de technologies différentes qui, pour la plupart, fonctionnent en arrière-plan. De cette manière, c'est comparable à Internet et à d'autres technologies fondamentales qui traversent et influencent nos existences sur le plan social, politique et économique. La principale différence tient au fait que l'intelligence artificielle table sur l'infrastructure existante (Internet, grands ensembles de données, capacité de s'alimenter à partir d'un nombre croissant de capteurs numériques et de sources de données) pour réduire radicalement les coûts des activités (à la fois les nouvelles et les anciennes, les bonnes et les mauvaises) sur une grande échelle: par exemple, apprentissage personnalisé ou diagnostic en matière de santé, soins automatisés, traduction automatique, robots conversationnels intelligents, chaînes d'approvisionnement optimisées, cybercriminalité encore plus sophistiquée, publicités ciblées et surveillance grâce à la reconnaissance faciale. Ces puissantes nouvelles technologies d'IA sont mises au point et déployées dans un contexte d'inégalités flagrantes et de noeuds du pouvoir fortement consolidés. Pourquoi, alors, devrions-nous nous attendre à ce que les résultats – pris globalement – soient équitables ou éthiques ? Ajoutez à cela les risques largement accrus d'atteinte au droit à la vie privée, le potentiel de surveillance et les possibilités de mauvaise utilisation par les intervenants malveillants, et il devient facile de se montrer pessimiste.

Ceci dit, nous aimerions conclure sur une note plus encourageante et constructive. L'argument présenté ci-dessus suppose que nous poursuivions sur la même trajectoire; que les sociétés seront incapables d'adapter et de réglementer de façon appropriée les applications d'IA, de protéger la vie privée et de serrer la bride aux abus; et que l'intelligence artificielle ne sera pas mise au point d'une manière plus inclusive axée sur la résolution de problèmes sociaux, économiques et environnementaux pressants. Cependant, si le passé est souvent un mauvais indicateur prévisionnel de l'avenir, il reste que nous sommes en présence d'indications d'une préoccupation et d'un intérêt croissants en regard de la construction d'un résultat global plus positif.

L'un de ces indicateurs est la manière dont l'émergence de l'intelligence artificielle dans le discours mondial semble avoir déclenché une discussion publique de plus grande portée au sujet du genre de société que nous voulons bâtir à l'avenir. À bien des égards, c'est le potentiel de l'intelligence artificielle de transformer nos institutions économiques, politiques et sociales qui force cette introspection alors qu'il attire l'attention sur des aspects de nos sociétés (comme les biais sociaux inhérents) qui avaient été auparavant ignorés ou

passés sous silence. C'est seulement lorsque ces éléments sont révélés que nous pouvons examiner des solutions novatrices, qu'elles soient techniques, culturelles, économiques ou politiques. En outre, l'impératif d'assurer une participation inclusive au développement et au déploiement de l'intelligence artificielle et d'engager une conception inclusive des applications d'IA, tout en travaillant très fort en vue d'atténuer les risques potentiels, offre une occasion de changer le discours et les pratiques en ce qui a trait à l'utilisation des technologies dans le cadre de projets de changements sociaux.

Les exemples sont nombreux.

- ▶ Le codage de biais sociaux dans les algorithmes d'IA, et leur exposition subséquente, peuvent ouvrir un espace pour la discussion publique au sujet de ces biais. Effectivement, des exemples de biais dans les algorithmes d'AA (de même que leur opacité) ont fait l'objet de nombreux articles dans la presse. Cependant, cela pourrait se révéler un problème critique relativement mineur, et pourrait même être à l'origine de changements sociaux plus positifs. Premièrement, cette conversation a incité des développeurs à prendre des mesures pour éviter d'introduire des stéréotypes sexistes dans leurs applications d'IA¹⁰¹ et le nouveau domaine de recherche s'est attaché à repérer et à éliminer les biais dans les algorithmes d'AA¹⁰². Éliminer les biais et l'opacité pourrait s'effectuer principalement par des correctifs techniques. Deuxièmement, découvrir ces biais sociaux offre une possibilité unique de tenir une discussion sociale à plus grande échelle sur les biais en tant que tels, leur histoire et leur nature, comment ils ont pu être institutionnalisés, et ce qui pourrait être fait pour les corriger. L'intelligence artificielle peut coder les biais sociaux, mais elle peut aussi être utilisée comme un miroir pour les refléter à la société.
- ▶ La reconnaissance croissante du potentiel des plateformes Internet, comme Facebook et Google, d'exercer une influence sociale et politique importante a suscité des discussions sur les mesures antitrust et la surveillance réglementaire. Des conversations ouvertes ont aussi été tenues sur les moyens que pourraient prendre les grandes plateformes elles-mêmes pour s'attaquer à ces préoccupations, et certaines ont même déjà commencé à le faire¹⁰³.
- ▶ Les prévisions de pertes d'emplois massives en raison de l'automatisation grâce à l'IA sont à l'origine de nouvelles innovations et expérimentations en matière de politiques dans les domaines de la protection sociale, de la fiscalité et de l'éducation.
- ▶ Les gouvernements commencent eux aussi à expérimenter avec la réglementation. Par exemple, comme nous l'avons déjà mentionné, l'UE



inclut le langage autour de la transparence des algorithmes d'AA dans son Règlement général sur la protection des données. En France, le président Emmanuel Macron propose de déposer un projet de loi pour interdire les fausses nouvelles¹⁰⁴.

Si personne ne peut prévoir quelles seront ces mesures, ce qui est clair en revanche, c'est que le statu quo n'a pas beaucoup de chances d'être maintenu.

Même si des efforts visant à affronter les défis posés par l'intelligence artificielle voient le jour, il reste beaucoup à faire, surtout en ce qui concerne l'IA dans les pays du Sud, où le développement de l'IA, les capacités de développement et d'élaboration de politiques et les ressources sont comparativement bien minces, et les bénéfices et les risques potentiels de l'IA sont amplifiés. Nous avons besoin de recherches et de données probantes pour éclairer et façonner le développement et l'utilisation des applications d'IA, et de politiques et réglementations gouvernementales afin d'assurer une utilisation équitable et appropriée. Il nous faut une plus grande capacité dans les pays du Sud pour stimuler ce programme de recherche, élaborer des solutions adaptées à l'échelle locale et élaborer des politiques et une réglementation qui sont efficaces dans les contextes institutionnels locaux.

Photo par Matus Kovacovsky sur Unsplash



L'avenir de l'intelligence artificielle est déjà à nos portes.
La question est de savoir si nous serons prêts à l'accueillir.

Renforcer les capacités dans le Sud n'est que le début; il faut que l'effort soit d'envergure mondiale. Le caractère transnational d'Internet, des plateformes en ligne, des flux de données, et ainsi de suite, fait en sorte que l'application éthique et équitable de l'intelligence artificielle est un défi collectif qu'aucun pays n'est en mesure de relever seul. Nous devons établir les valeurs et les principes à l'échelle mondiale et locale eu égard à l'intelligence artificielle qui accordent la priorité à l'inclusion, à l'éthique et à la transparence. Nous devons mettre en place des collaborations interdisciplinaires internationales dans le cadre desquelles des chercheurs de l'IA se pencheront sur les problèmes du monde réel. Et il est en outre essentiel d'étendre l'autodétermination des collectivités afin qu'elles puissent activement tenir des discussions sur le développement et le déploiement des applications d'IA qui ont une incidence sur leurs vies.

Un point de départ possible pourrait être l'affectation de ressources importantes par l'entremise d'un fonds mondial réservé à l'intelligence artificielle pour le développement axé sur le renforcement des capacités de « centres d'excellence de l'IA » dans les pays du Sud. Ces centres entreprendraient des recherches et formuleraient aux gouvernements des conseils empiriquement étayés en matière de politiques et de réglementations à l'échelle régionale et nationale. Ils travailleraient non seulement localement, en mobilisant les collectivités pertinentes et touchées, mais collaboreraient aussi avec les centres de recherche et d'élaboration de politiques de partout dans le monde pour s'attaquer aux défis communs et apporter une plus grande diversité de perspectives et de voix à l'échelle mondiale.

Le futur est inconnu, mais il est clair que le moment est venu d'agir, avant que l'intelligence artificielle ne perturbe radicalement les sociétés des pays du Sud. Dans la dernière section suivante, le document présente une série de recommandations de mesures et de recherches conçues pour contribuer à l'application éthique et égalitaire de l'intelligence artificielle dans les pays du Sud.



RECOMMANDATIONS

En nous fondant sur les conclusions du présent document et de la documentation consultée, nous avons cerné trois principaux domaines dans lesquels des mesures peuvent être prises : politiques et réglementations, applications inclusives et éthiques de l'IA et infrastructure et compétences. Dans chaque domaine, nous formulons une série de recommandations sur des recherches qui sont nécessaires pour obtenir des progrès concrets¹⁰⁵. Veuillez prendre note que cette liste n'est pas exhaustive, et qu'elle vise plutôt à mettre en lumière les interventions les plus pressantes.

Voici les caractéristiques de la recherche:

Interdisciplinaire: Bon nombre de questions de recherche portent sur l'intersection entre les facteurs sociaux et techniques. Pour les aborder efficacement, il faudra établir des collaborations multidisciplinaires entre les sciences sociales (y compris l'économie et la politique), les sciences humaines et l'informatique.

Effectuée localement: Afin d'assurer la pertinence et l'utilité, la recherche doit être dirigée et effectuée par des chercheurs dans les pays du Sud.

Conçue pour appuyer les pratiques et les politiques: Pour exercer un réel impact, il faut que la recherche soit rigoureuse et qu'elle donne lieu à des constatations concrètes qui facilitent la mise en oeuvre de programmes et de politiques.

Nous répartissons les recommandations dans les trois niveaux de l'écosystème d'IA : 1) structures politiques et réglementaires; 2) applications; et 3) infrastructure et compétences. Il convient de mentionner que ces trois couches sont interreliées et se chevauchent, et de ce fait, elles sont fournies à titre d'heuristique utile plutôt que de catégories distinctes.

Structures politiques et réglementaires

Applications

Infrastructure et compétences



STRUCTURES POLITIQUES ET RÉGLEMENTAIRES

Favoriser la conception de politiques et de règlements qui habilitent une IA inclusive et axée sur les droits

Mener des recherches de base sur la prévalence des applications d'IA et de politiques dans les pays du Sud. Malgré des poches d'activités liées à l'IA dans les pays du Sud, il n'existe aucune synthèse systématique du niveau de cette activité. La collecte de données de base doit inclure les ensembles de politiques, de règlements, d'applications, d'ensembles de données ouvertes existants et de niveaux de compétences en matière d'intelligence artificielle. Cette recherche doit être effectuée chaque année, ou à tout le moins, tous les deux ans, afin d'appuyer les activités continues, l'élaboration de politiques et le programme de recherche.

Se renseigner au sujet de modèles réglementaires efficaces. Documenter et évaluer les modèles réglementaires de l'intelligence artificielle élaborés pour composer avec l'émergence de nouvelles activités qui utilisent l'IA comme les services de police prédictifs, les véhicules autonomes et les robots conversationnels. Déterminer si les risques potentiels des applications d'IA sont pris en compte adéquatement dans la réglementation existante, ou si la réglementation existante doit être adaptée ou encore si une nouvelle

réglementation doit être élaborée. Déterminer les réponses réglementaires à des cas d'utilisation de l'intelligence artificielle donnés, ainsi que les niveaux de risque qui sont appropriés pour les contextes associés à une capacité institutionnelle réduite. Même si les leçons apprises dans les pays du Nord sont utiles, il est essentiel de ne pas importer directement les approches institutionnelles et réglementaires dans les contextes du Sud où les formes institutionnelles et culturelles sont différentes.

Surveiller l'incidence de l'IA sur l'emploi et le travail. Mener des recherches sur la politique sociale et économique afin de comprendre les effets de l'IA sur l'emploi, la nature du travail et les marchés du travail. Dans quelle mesure l'automatisation grâce à l'IA modifie-t-elle les modèles d'emplois et transforme-t-elle le milieu du travail ? Quels sont les autres modèles de revenus et de distribution des ressources, d'éducation et de recyclage professionnel dans les différents contextes ?

Examiner les approches pour composer avec la responsabilité, et les mécanismes de réparation eu égard à la prise de décisions fondées sur l'IA. Concevoir des systèmes et des cadres réglementaires qui déterminent la responsabilité eu égard à la prise de décisions fondées sur l'IA qui sont erronées, biaisées ou discriminatoires, et établir des mécanismes de réparation. Ces mesures peuvent comprendre notamment des politiques qui stipulent la transparence eu égard au processus décisionnel automatisé, des procédures d'évaluation en vue de déterminer la compétence des systèmes d'IA, et la certification des systèmes d'IA qui exécutent des tâches nécessitant un certain degré de compétence ou de formation. La nécessité d'agir est particulièrement urgente dans le cas de systèmes de décision qui influent sur le bien-être ou la liberté des personnes, comme ceux qui font appel à l'usage de la force ou à l'incarcération. La recherche est essentielle dans ce cas pour découvrir et documenter quels sont les systèmes efficaces, et dans quels contextes, en matière de responsabilité et de réparation.

Étudier l'incidence de l'IA sur les droits de la personne¹⁰⁶. De manière générale, les Nations Unies reconnaissent que les droits qui s'appliquent hors ligne doivent s'appliquer en ligne, témoignant de la pertinence des droits analogiques dans des environnements régis numériquement. Des organismes professionnels demandent spécifiquement que les droits de la personne soient pris en considération dans le contexte de la conception et de l'exploitation de l'IA¹⁰⁷. Calquer les évaluations d'impact sur les risques liés à l'IA contribuerait à encourager les programmes de développement à intégrer la technologie d'IA de manière à respecter et à promouvoir les droits de la personne, y compris le droit à la vie privée, à l'égalité et à la liberté d'expression.



APPLICATIONS

Catalyser le développement d'applications d'IA inclusives et éthiques

Appuyer le développement et le déploiement d'applications d'IA innovatrices au service du bien social. Investir dans le développement, le déploiement et l'utilisation d'applications pour l'éducation, la santé, l'environnement, la sécurité alimentaire, etc., et faire en sorte que ces applications soient éthiques et inclusives. Comme pour les innovations en matière de réglementation, même s'il est important de s'inspirer d'exemples de partout dans le monde, les applications d'IA nécessiteront souvent des solutions locales pour être efficaces.

Faire de la recherche sur les répercussions sociales des innovations utilisant l'IA. Des recherches sont nécessaires afin de mieux comprendre quelles sont les applications d'IA qui fonctionnent (ou celles qui ne fonctionnent pas), pour qui, et dans quels contextes. Nous devons savoir qui profite des applications d'IA, et comment, ainsi que qui est laissé pour compte ou victime de ces applications. On insistera plus particulièrement sur l'examen des répercussions différentielles sur divers groupes, surtout lorsque les différences sont imputables au genre, au statut social et économique, à la race, etc. Cette recherche doit aller au-delà des effets de

premier ordre, comme l'amélioration des gains d'efficacité ou de l'exactitude des diagnostics, pour inclure des effets sociaux plus larges. Il est possible que l'on doive faire appel à de nouvelles méthodologies d'évaluation et d'évaluation d'impact.

Mettre à l'essai et surveiller les biais dans les applications d'IA. Les systèmes d'intelligence artificielle qui prennent des décisions ou qui informent la prise de décisions qui ont une incidence sur le bien-être des humains (par exemple diagnostics médicaux, fournir à un juge une évaluation du potentiel de récidive) devraient être mis à l'essai et surveillés en vue d'y repérer des biais et des erreurs dans les différents contextes et collectivités, avant la diffusion et continuellement par la suite.

Examiner les modèles de conception participative pour l'IA. Mener des recherches sur les pratiques qui appuient le développement d'applications d'IA inclusives. Quelles sont les techniques efficaces pour obtenir des processus véritablement participatifs qui mobilisent des populations diverses dans la conception et le déploiement d'applications d'IA ? Comment et dans quels contextes ces pratiques parviennent-elles à contrer les biais de conception et les biais appris et à faire en sorte que l'IA soit pertinente pour les collectivités marginalisées ? Les intervenants de l'intelligence artificielle sur le terrain devraient publier des données sur la diversité de la participation à la conception et au développement.

Une recherche-action pour mieux comprendre comment mettre à l'échelle efficacement et équitablement des applications d'IA éprouvées. Démontrer la validité du concept est différent de diffuser l'application dans toute une population donnée tout en maintenant la qualité et l'équité des avantages. La recherche sur le processus de mise à l'échelle des applications d'IA, tant à la verticale pour englober les fonctionnalités additionnelles, qu'à l'horizontale pour s'étendre dans de nouveaux endroits, est essentielle pour propager les avantages de ces applications. La généralisation dans divers contextes nécessite la compréhension des raisons pour lesquelles une application a bien fonctionné dans un certain contexte et une appréciation du fait que l'application pourrait avoir besoin de modifications pour réussir dans un nouvel environnement. Nous devons échafauder des théories sous-jacentes à la mise en oeuvre des applications d'IA afin de reproduire de façon fiable la réussite au sein de populations différentes. Dans ce contexte, il faudrait mettre l'accent sur les variations entre les contextes plutôt que s'en tenir strictement aux modalités de l'une ou l'autre des mises en oeuvre. Les difficultés particulières liées aux données comprendront notamment la mise à l'échelle au-delà de la portée des ensembles de données existants et l'élaboration de moyens pour produire rapidement de nouveaux ensembles de données.



INFRASTRUCTURE ET COMPÉTENCES

Construire l'infrastructure et les compétences en vue d'une IA inclusive et éthique

Appuyer des programmes destinés à renforcer l'expertise en matière d'IA dans les gouvernements. Promouvoir l'expertise en matière d'IA dans toutes les branches et à tous les niveaux des gouvernements, y compris les organismes réglementaires et, éventuellement, les nouveaux organismes consultatifs.

Favoriser les capacités locales à la tête de la conception, du développement et du déploiement des applications d'IA. Ces activités pourraient comprendre notamment : soutenir la croissance de centres d'excellence multidisciplinaires en intelligence artificielle dans les pays du Sud en vue de s'engager dans le développement et la recherche à l'échelle locale, et de fournir des données probantes pour contribuer à l'élaboration de la politique nationale et aux décisions réglementaires; établir des liens entre les experts techniques et les communautés marginalisées et à faible revenu dans les pays du Sud; et soutenir les collaborations Sud-Sud.

Élaborer et mettre à l'épreuve des approches rentables pour renforcer des compétences pertinentes en matière d'IA, plus particulièrement chez les femmes et dans les populations marginalisées. Élaborer et soutenir des programmes qui se concentrent sur le développement des capacités



des femmes et d'autres populations marginalisées en vue de les engager à différentes étapes du développement et de l'application des technologies d'IA. La recherche devrait stimuler cette activité au moyen de l'examen de modèles à faibles coûts pour perfectionner les compétences techniques en matière d'IA ainsi que ainsi que mettre à l'essai des programmes et des pédagogies efficaces.

Élargir l'accès aux données et aux ressources informatiques. Dans la mesure du possible, la recherche, les outils et les ensembles de données de formation liés à l'IA doivent être fournis gratuitement. Appuyer le développement et le partage d'ensembles de données diversifiées et inclusives qui sont nécessaires pour les applications d'IA dans différents contextes.

Étudier les avantages et les risques liés à l'intelligence artificielle ouverte. Mener des recherches sur les risques et les avantages à court et à moyen terme de l'ouverture dans le domaine de l'IA (c'est-à-dire le partage des ressources d'IA, des ensembles de données, etc)¹⁰⁸. Si possible, cette recherche devrait établir un lien entre les questions du côté de l'approvisionnement (quel serait le meilleur moyen d'offrir un accès ouvert aux algorithmes d'IA, aux outils et aux ensembles de données) en approfondissant la compréhension de l'engagement nécessaire pour s'assurer que les ressources d'IA ouvertes sont disponibles en vue d'une utilisation ou d'une réutilisation et d'une adaptation par diverses populations (et pas seulement par ceux qui sont déjà compétents et dotés de ressources). Il conviendrait d'accorder une attention particulière à la question de l'équilibre entre le partage d'ensembles de données et la protection de la vie privée.

En gardant les yeux bien ouverts et en se renseignant sur le potentiel et les pièges de l'intelligence artificielle, le monde se trouve devant une occasion unique et impérative de se pencher sur les défis posés par cette technologie et de construire un meilleur avenir. Des actions concertées et fondées sur des données probantes, telles que celles qui sont recommandées ci-dessus devraient aider les pays du Sud à récolter les fruits d'un futur propulsé par l'intelligence artificielle, tout en atténuant les risques et les préjudices et en préservant la dignité humaine.

Photo par Samuel Zeller sur Unsplash





ANNEXE

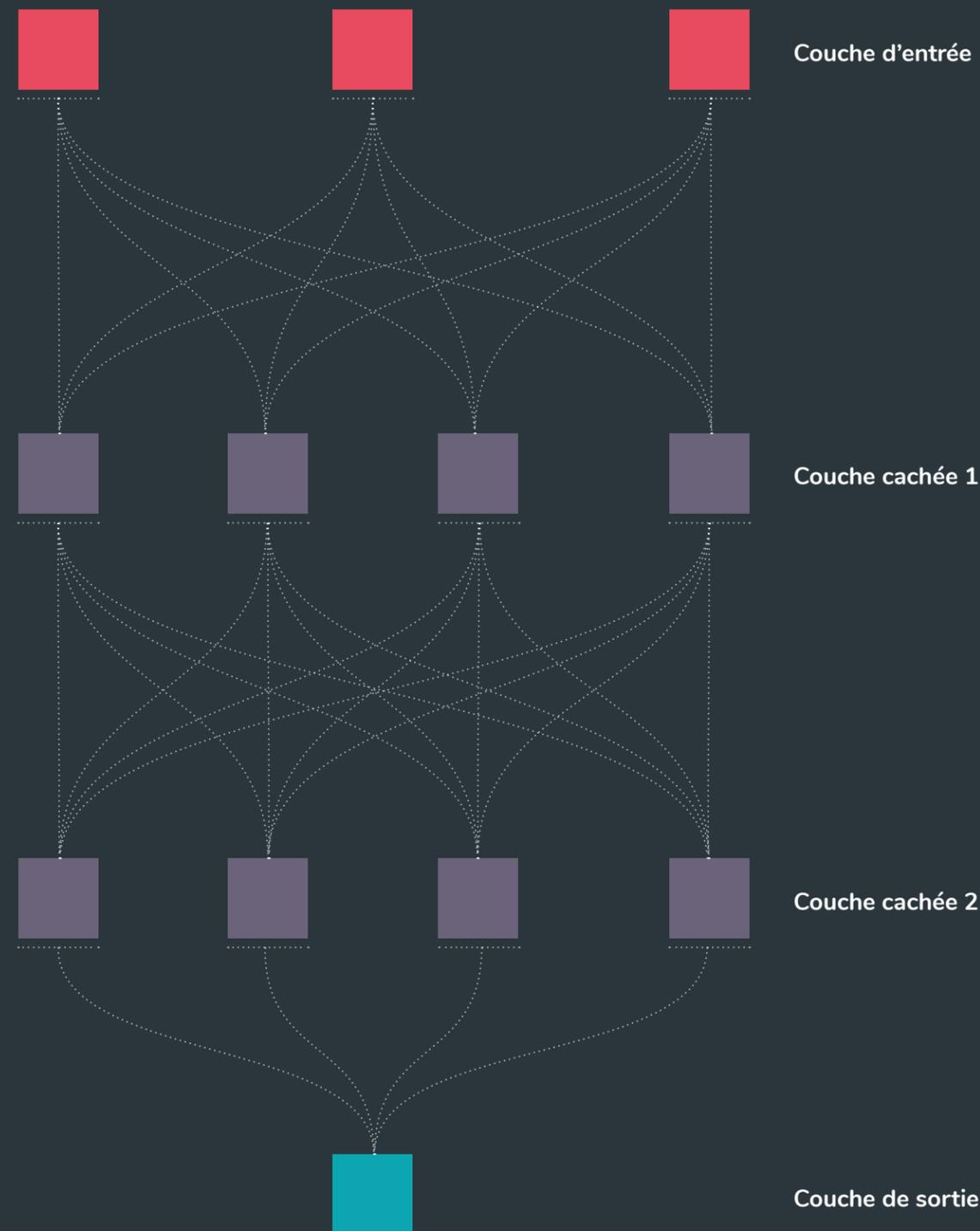
Comment l'IA fonctionne-t-elle?



Même s'il existe de nombreuses techniques pour développer des systèmes d'IA, nous allons nous concentrer sur deux approches largement utilisées. La première est l'apprentissage automatique (AA), avec laquelle le système apprend à trouver une solution. La deuxième approche est celle des systèmes experts, avec laquelle l'IA suit un ensemble prédéterminé et préprogrammé de règles et de logique conçu pour produire des comportements particuliers et qui se répètent. Les systèmes experts sont possibles lorsqu'il existe une base de connaissances antérieure bien développée sur laquelle la machine fonde ses actions.



FIGURE 4 Un réseau neuronal à trois couches



Source: bit.ly/2HZTyna

L'APPRENTISSAGE AUTOMATIQUE

L'AA s'est avéré l'approche la plus fructueuse et influente dans le développement de l'IA. L'augmentation récente de l'utilisation et du succès de l'apprentissage automatique est due à deux facteurs: l'accès à des ordinateurs plus puissants rendu possible grâce à la baisse des coûts et à la capacité croissante, et la disponibilité de vastes ensembles de nouvelles données que l'on appelle les « mégadonnées ». Dans cette section, nous décrivons le fonctionnement de l'apprentissage automatique, et le rôle des ensembles de données dans le fonctionnement de ces systèmes.

Il existe trois types principaux d'algorithmes d'apprentissage automatique: apprentissage supervisé, apprentissage non supervisé et apprentissage par renforcement.

Apprentissage supervisé

Dans l'apprentissage automatique supervisé, les algorithmes apprennent à partir de données (par exemple des images numériques d'oranges et de pommes) étiquetées à l'aide de métadonnées indiquant les bonnes réponses (c'est-à-dire qu'une image donnée représente en fait une orange ou une pomme). Un tel ensemble de données est appelé données d'apprentissage. L'algorithme tente ensuite d'extraire des modèles de classification de cet ensemble de données en vue de classer correctement les nouvelles données non étiquetées. Il extrait ces modèles de classification en modifiant ses propres paramètres jusqu'à ce qu'il ait « appris » les modèles contenus dans les données. Autrement dit, le modèle dans les données devient représenté dans les paramètres de l'algorithme d'AA. On appelle cette approche apprentissage automatique supervisé parce que l'algorithme reçoit au départ des données étiquetées à partir desquelles il peut apprendre.

Un type d'algorithme d'AA est le réseau neuronal artificiel (RNA). Les réseaux neuronaux sont inspirés des neurones biologiques. Les données sont alimentées dans une couche de neurones qui effectuent un calcul en fonction d'un jeu de paramètres. Le résultat de ce calcul est ensuite transmis à la couche suivante ou aux couches suivantes, jusqu'à la production d'une sortie. Un réseau neuronal supervisé comparera ensuite cette sortie avec la sortie souhaitée et calculera l'«erreur». À l'aide de cette erreur, il repasse à travers le réseau en effectuant de petits ajustements aux paramètres pour produire une sortie améliorée pour cette entrée la prochaine fois. Ce processus est répété de nombreuses fois pour toutes les données de l'ensemble d'apprentissage, jusqu'à ce que la somme totale des erreurs soit suffisante pour constituer un succès (voir la Figure 4).



Une fois que le réseau est formé, il peut recevoir de nouvelles entrées et il produira une sortie prévisible. Dans l'exemple des fruits qui précède, l'algorithme « apprend » à représenter les caractéristiques utiles des fruits, comme la taille et la couleur, en procédant à l'ajustement de nombreux paramètres. Une fois qu'il a appris, il devient effectivement un détecteur capable de faire la distinction entre « une orange/pomme » qui fournira sa meilleure réponse pour indiquer si une image représente une orange ou une pomme.

Apprentissage non supervisé

Les algorithmes d'apprentissage non supervisé apprennent des modèles dans les données sans avoir reçu de données de sortie étiquetées. Les algorithmes non supervisés ne reçoivent pas la bonne « réponse »; ils tentent plutôt d'extraire les diverses caractéristiques de l'ensemble de données et construisent des classes de points de données possédant des caractéristiques comparables. Ceux-ci peuvent servir à découvrir des groupes, comme un segment de clients, ou à former des associations (ou règles) dans les données – par exemple, « si X se produit, alors Y devient ». Dans l'exemple des fruits, un algorithme non supervisé pourrait classer les gros fruits et les petits fruits dans deux classes. Il pourrait aussi avoir la capacité de créer des classes multidimensionnelles fondées sur d'autres caractéristiques comme la forme et la couleur.

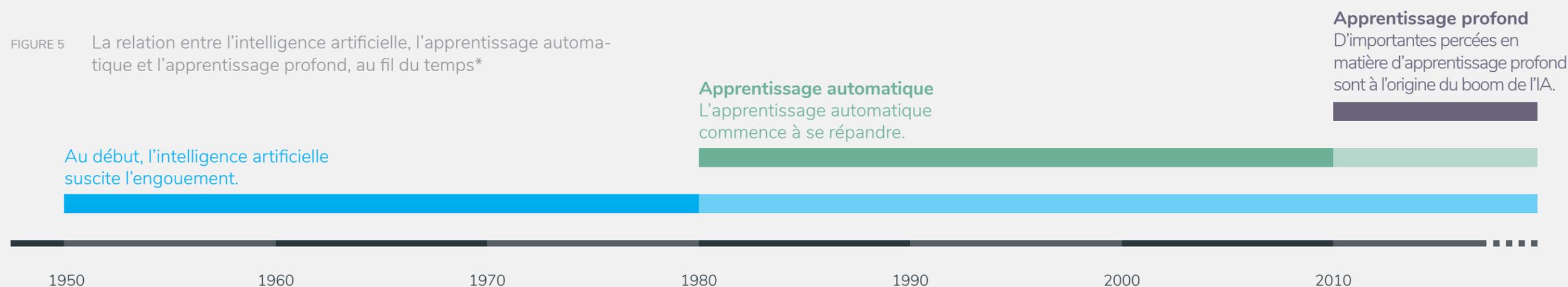
Apprentissage semi supervisé

Cette approche sert à former l'algorithme à la fois avec les données étiquetées et les données non étiquetées. Le coût de l'étiquetage des ensembles de données peut être très élevé, et un ensemble de données entièrement étiqueté peut être irréalisable. L'inclusion de données étiquetées, toutefois, aide l'algorithme à identifier des classes ou des règles potentielles dans les données, tandis que les données non étiquetées permettent à l'algorithme de façonner les limites de ces classes ou règles et même d'en trouver de nouvelles. L'apprentissage semi-supervisé donne de bons résultats avec des tâches comme la reconnaissance vocale et des images.

L'apprentissage profond est un type de réseau neuronal artificiel qui utilise de nombreuses couches cachées. Habituellement, il mélange l'apprentissage supervisé et non supervisé, et à l'occasion, l'apprentissage semi-supervisé. Même s'il n'existe aucun nombre exact de couches cachées qui correspond à l'apprentissage « profond », des réseaux neuronaux artificiels (RNA) d'apprentissage profond ont été créés qui comportaient plus d'un milliard de connexions au total¹⁰⁹. La capacité de mettre en oeuvre des techniques d'apprentissage profond est le facteur qui explique le regain d'intérêt pour l'intelligence artificielle et les avancées réalisées (voir la Figure 5).

La puissance de l'apprentissage profond provient de la capacité des nombreuses couches d'extraire des caractéristiques complexes à partir des données d'apprentissage.

FIGURE 5 La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond, au fil du temps*

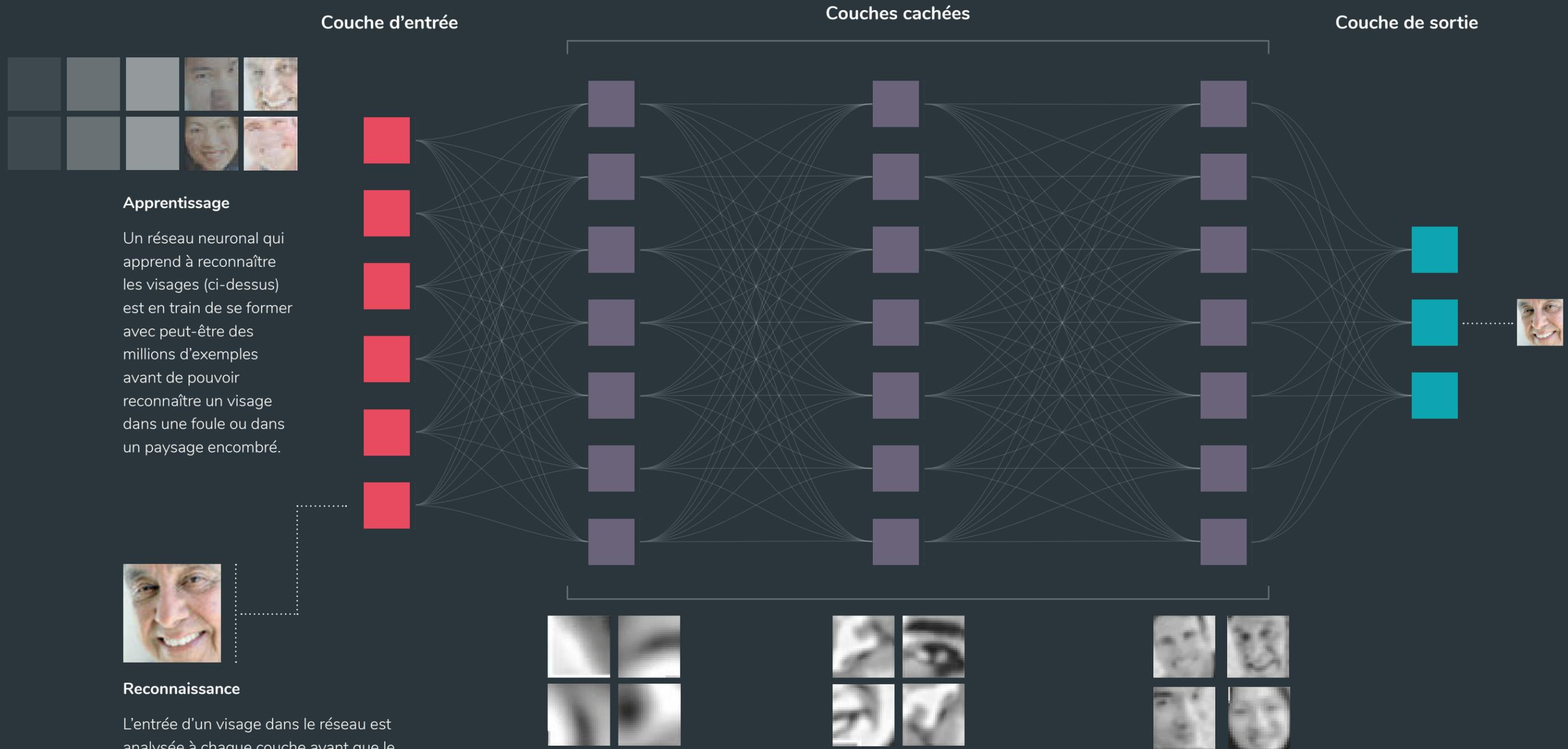


Depuis une première poussée d'optimisme dans les années 1950, de plus petits sous-ensembles d'intelligence artificielle – d'abord, l'apprentissage automatique, puis l'apprentissage profond, un sous-ensemble de l'apprentissage automatique – ont entraîné des perturbations de plus en plus grandes.

* bit.ly/2HXaIVZ

FIGURE 6

Dévoilement des couches cachées de l'apprentissage profond



Apprentissage

Un réseau neuronal qui apprend à reconnaître les visages (ci-dessus) est en train de se former avec peut-être des millions d'exemples avant de pouvoir reconnaître un visage dans une foule ou dans un paysage encombré.



Reconnaissance

L'entrée d'un visage dans le réseau est analysée à chaque couche avant que le réseau ne puisse deviner correctement son identité.

Chaque couche identifie progressivement des caractéristiques plus complexes.

Source: <http://bit.ly/2J6xqZq>

Dans l'exemple de la classification visuelle des fruits, des caractéristiques comme la couleur et la dimension se verraient attribuer une priorité plus élevée que le goût. Dans le cortex visuel du cerveau d'un mammifère, la donnée qui entre par l'œil traverse une série de couches successives de traitement hiérarchisé qui extraient des caractéristiques comme les arêtes, les contours, et les formes des objets dans de nombreuses zones visuelles (Hubel et Wiesel, 1962; Felleman et Van Essen, 1991)¹¹⁰, (Voir la Figure 6).

Apprentissage par renforcement

Si les algorithmes d'AA comme les réseaux neuronaux apprennent à partir d'ensemble de données, les algorithmes d'apprentissage par renforcement (AR) apprennent à partir d'un renforcement positif et négatif fourni par l'environnement. Lorsqu'un agent d'IA avec un algorithme d'AR prend contact avec le monde, il reçoit une rétroaction positive ou négative fondée sur sa plus récente action. Cette rétroaction est récupérée dans l'algorithme, lequel apporte des corrections en fonction de l'action qu'il a choisie à cet état particulier (voir la Figure 7).

L'AR est très puissant pour les tâches axées sur un but, comme apprendre un jeu ou pour enseigner à un robot une compétence. Par exemple, à l'aide

de l'AR, un bras robotisé a appris comment retourner une crêpe¹¹¹ et à jouer au ping-pong¹¹² et une araignée autonome a appris à marcher¹¹³. La capacité de manipuler des objets est essentielle pour la robotique dans le secteur manufacturier. L'AR a aussi été utilisé dans le secteur financier pour élaborer des stratégies visant les opérations sur actions¹¹⁴. L'AR a aussi habilité l'application d'IA de jeu AlphaGo à élaborer des stratégies nouvelles et uniques pour vaincre le champion mondial au jeu de go¹¹⁵.

Réflexions finales sur les algorithmes d'AA

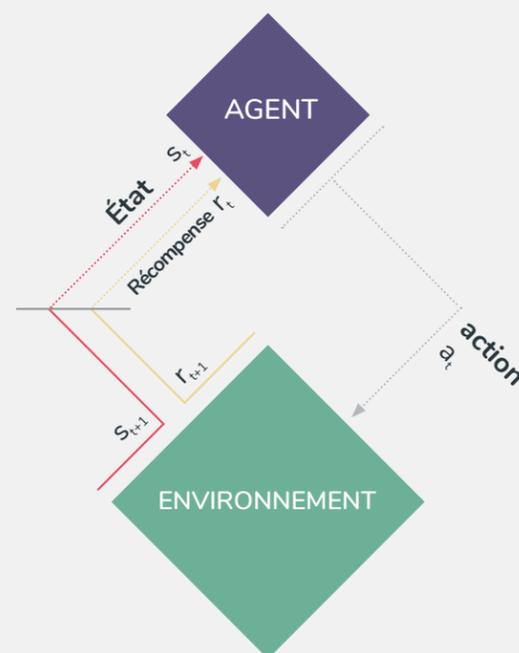
Il existe aussi des algorithmes d'apprentissage plus « simples » qui n'ont pas besoin de gros ensembles de données ni de puissance de calcul importante. Par exemple, une simple régression linéaire – une technique mathématique qui élabore une équation qui « correspond » à un ensemble de données particulier constitué d'entrées et d'une sortie (des variables indépendantes et une variable dépendante) – est considéré comme un algorithme d'AA¹¹⁶. L'algorithme peut ensuite « prévoir » la sortie la plus probable à partir d'un nouveau point de données entré. Une application de cette technique dans le contexte de l'enseignement supérieur pourrait consister à « prévoir » le niveau de risque d'un étudiant à partir de données telles que l'assiduité et les taux de participation à des systèmes de gestion de l'apprentissage. Cet algorithme particulier ne nécessite pas de gros ensembles de données, de puissance de calcul importante, ni de algorithmes d'apprentissage complexes.

SYSTÈMES EXPERTS

Les algorithmes d'AA sont un type d'IA « non symbolique ». Autrement dit, ils ne sont pas explicitement programmés pour résoudre un problème à l'aide d'instructions logiques (par exemple, « si ceci, faire cela ») ou d'un raisonnement symbolique; au contraire, les algorithmes d'AA traitent les données à grande vitesse et apprennent des solutions en ajustant les paramètres de l'algorithme. Ces techniques sont inspirées largement du fonctionnement du cerveau humain, où l'on croyait que les connaissances étaient stockées dans une masse de connexions neuronales qui sont constamment ajustées, renforcées ou rompues.

La naissance de l'intelligence artificielle, toutefois, est ancrée dans des approches qui utilisaient des symboles pour représenter un problème, un peu comme le font les humains qui réfléchissent à un problème. Une approche de ce genre est à l'origine du développement des « systèmes experts » (Harmon et King, 1985) qui

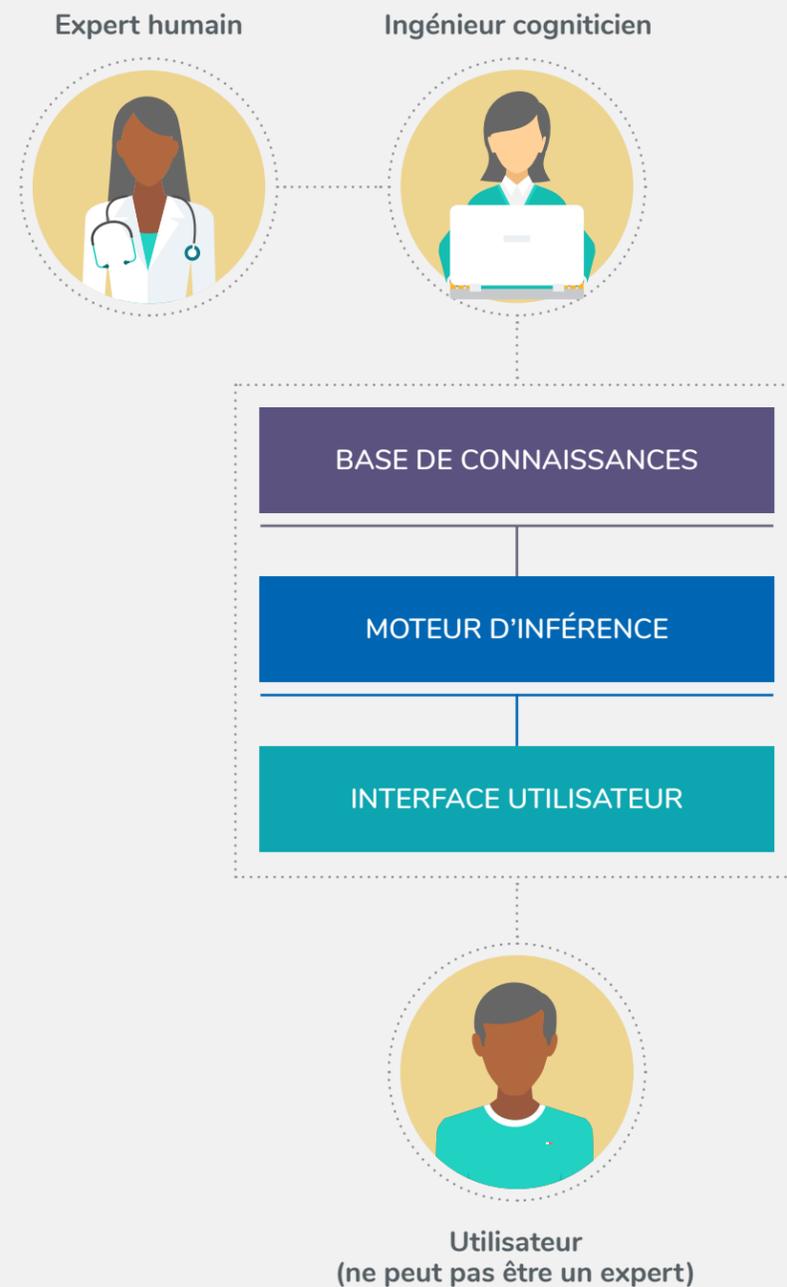
FIGURE 7 L'apprentissage par renforcement et sa mise en œuvre



Source: bit.ly/2uepiTw



FIGURE 8 Composants des systèmes experts



Source: bit.ly/2DNtkS9

tentent d'émuler les aptitudes à la résolution de problèmes d'un expert humain (Durkin, 1990 : 171). Edward Feigenbaum définit un système expert comme « un programme informatique intelligent qui utilise des connaissances et des méthodes de déduction pour résoudre des problèmes qui sont suffisamment difficiles pour nécessiter l'expertise humaine pour trouver leur solution¹¹⁷ ». Les systèmes experts étaient populaires à une époque où la puissance de calcul et la disponibilité des données étaient plus limitées. Par exemple, en 1987, plus des deux tiers des entreprises appartenant au groupe Fortune 1000 travaillaient au développement de projets de systèmes experts (Durkin, 1990).

Les principaux éléments d'un système expert sont:

- ▶ Une base de connaissances constituée de connaissances spécialisées codées de manière explicite (expertise), souvent codées selon les règles « si... alors ».
- ▶ Mémoire de travail : Les données entrées par les utilisateurs ou tirées d'autres sources de données et les faits déduits par le système.
- ▶ Moteur d'inférence : Dérivation de nouveaux renseignements au sujet du problème à partir des renseignements disponibles dans la mémoire de travail et la base de connaissances. Le mode de fonctionnement habituel du système consiste a) à établir un but ou une hypothèse ou à tenter de la vérifier au moyen de l'analyse de données, ou b) à recueillir des renseignements au sujet d'un problème, puis à déduire d'autres renseignements (Durkin, 1990).

La Figure 8 montre la relation entre l'expert, le système expert et l'utilisateur.

Il existe de nombreux usages pour les systèmes experts de nos jours, ils peuvent en effet servir d'outils de diagnostic médical ou de systèmes tutoriels intelligents, ils peuvent faire voler des avions en pilotage automatique et faire fonctionner des ascenseurs, ils peuvent aussi être utilisés dans un cadre de planification écologique, pour effectuer le diagnostic de pannes de production et dans l'ordonnancement (Liao, 2005). La représentation symbolique des connaissances et les techniques des systèmes experts sont aussi utilisées simultanément avec les algorithmes d'AA pour obtenir des comportements plus complexes.



RÉFÉRENCES

Ahmad, T. et coll., « Characterizing dengue spread and severity using internet media sources ». Actes du 3e Symposium on Computing for Development de l'ACM, ACM, 2013, p. 18. Sur Internet: bit.ly/2Fe1y1W

AI Now, The AI Now 2017 Report, 2017. Sur Internet: bit.ly/2FewZcy

AI Now, The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term. Sommaire du symposium public AI Now organisé par la Maison-Blanche et l'Information Law Institute de l'Université de New York et tenu le 7 juillet 2016. Sur Internet: bit.ly/2qX0kDW

Arntz, M., T. Gregory et U. Zierahn, The Risk of Automation for Jobs in OECD Countries, documents de travail de l'OCDE sur les questions sociales, l'emploi et les migrations, no 189, 2016. Sur Internet: bit.ly/2qT0qM6

Banque mondiale, World Development Report 2016: Digital Dividends, Banque mondiale, 2016. Sur Internet: bit.ly/2JjR8Ql

Banavar, G., Learning to Trust Artificial Intelligence Systems: Accountability, Compliance and Ethics in the Age of Smart Machines, Somers (New York), IBM, 2016. Sur Internet: ibm.co/2qXGoJE

Benkler, Y., « Networks of power, degrees of freedom », International Journal of Communication, vol. 5, no 39, 2011

Bolsover, G. et P. Howard, « Computational propaganda and political big data: Moving toward a more critical research agenda », Big Data, vol. 5, no 4, 2017, p. 273-276. Sur Internet: bit.ly/2JqBLGb

Bostrom, N., « Strategic implications of openness in AI development », Global Policy, vol. 8, no 2, 2017, p. 135-148. DOI: 10.1111/1758-5899.12403.

Brundage, M. et coll., « The malicious use of artificial intelligence: Forecasting, prevention, and mitigation », ArXiv Preprint ArXiv:1802.07228, 2018.

Caliskan, A., J.J. Bryson et A. Narayanan, « Semantics derived automatically from language corpora contain human-like biases », Science, vol. 356, no 6334, 2017, p. 183-186.

Calo, R., Artificial Intelligence Policy: A Primer and Roadmap, 2017. Sur Internet (SSRN): bit.ly/2qVDegW

Photo par Todd Diemer sur Unsplash



- Christensen, G. et coll., *The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why?*, 2013. Sur Internet (SSRN): bit.ly/2KcBEQ6
- Danks, D. et A.J. London, *Algorithmic Bias in Autonomous Systems*. Présenté à la 26e International Joint Conference on Artificial Intelligence (IJCAI), 2017. Sur Internet: bit.ly/2Kb0s11
- Deibert, R.J., *Black Code: Inside the Battle for Cyberspace*, Toronto, Signal, 2013.
- Durkin, J., « Research review: Application of expert systems in the sciences », *Ohio Journal of Science*, vol. 90, no 5, décembre 1990, p. 171-179. Sur Internet: bit.ly/2HnKMU9
- Felleman, D.J. et D.C. Van Essen, « Distributed hierarchical processing in the primate cerebral cortex », *Cerebral Cortex*, vol. 1, no 1, janvier 1991, p. 1-47. DOI : 10.1093/cercor/1.1.1.
- Forum économique mondial, *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution, Global Challenge Insight Report*, 2016.
- Frey, C.B. et M.A. Osborne, « The future of employment: How susceptible are jobs to computerisation? » *Technological Forecasting and Social Change*, vol. 114, no C, 2013, p. 254-280. Sur Internet: bit.ly/2kP5AGv
- Friedman, B. et coll., « Minimizing bias in computer systems », *ACM SIGCHI Bulletin*, vol. 28, no 1, 1996, p. 48-51.
- Friedman, B. et H. Nissenbaum, « Bias in computer systems », *ACM Transactions on Information Systems (TOIS)*, vol. 14, no 3, 1996, p. 330-347.
- Garrido, M. et coll., *An Examination of MOOC Usage for Professional Workforce Development Outcomes in Colombia, the Philippines, and South Africa*, Technology and Social Change Group, Université de Washington, 2016.
- Hansen, J.D. et J. Reich, « Democratizing education? Examining access and usage patterns in massive open online courses », *Science*, vol. 350, no 6265, 2015, p. 1245-1247.
- Harmon, P. et D. King, *Expert Systems: Artificial Intelligence in Business*, New York, John Wiley & Sons, 1985. Sur Internet: bit.ly/2HrQ4Ke
- Hilbert, M., « The bad news is that the digital access divide is here to stay: Domestically installed bandwidths among 172 countries for 1986–2014 », *Telecommunications Policy*, vol. 40, no 6, 2016, p. 567-581.
- Hossain, M.S. et coll., « Belief-rule-based expert systems for evaluation of e-government: a case study », *Expert Systems*, vol. 32, no 5, 2015, p. 563-577.
- Hubel, D. et T. Wiesel, « Receptive fields, binocular interaction and functional architecture in the cat's visual cortex », *The Journal of Physiology*, vol. 160, no 1, janvier 1962. DOI : 10.1113/jphysiol.1962.sp006837.
- Institute of Electrical and Electronics Engineers (IEEE), *Ethically Aligned Design, version 2. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, 2017. Sur Internet: ethicsinaction.ieee.org/
- International Federation of Robotics, *The Impact of Robots on Productivity, Employment and Jobs*, 2017. Sur Internet: bit.ly/2HIPR8Z
- Kulik, J.A. et J.D. Fletcher, « Effectiveness of intelligent tutoring systems: a meta-analytic review », *Review of Educational Research*, vol. 86, no 1, 2016, p. 42-78.
- Laurillard, D., E. Kennedy et T. Wang, *How Could Digital Learning at Scale Address the Issue of Equity in Education?*, Quezon City (Philippines), Foundation for Information Technology Education and Development, Inc., 2018.
- Liao, S. H., « Expert system methodologies and applications – a decade review from 1995 to 2004 », *Expert Systems with Applications*, vol. 28, no 1, 2005, p. 93-103. DOI : 10.1016/j.eswa.2004.08.003.
- Lim, C.P. et V.L. Tinio (dir.), *Learning analytics for the Global South*, Quezon City (Philippines), Foundation for Information Technology Education and Development, 2018. Sur Internet: bit.ly/2HVlKr9
- Long, P. et G. Siemens, « Penetrating the fog: Analytics in learning and education », *Educause Review*, 2011. Sur Internet: bit.ly/2JoqoyR
- Lyon, D., « Surveillance, Power and Everyday Life », dans P. Kalantzis-Cope et K. Gherab-Martin (dir.), *Emerging Digital Spaces in Contemporary Society: Properties of Technology*, Springer, 2010, p. 107-120.
- Magoutas, B. et G. Mentzas, « SALT: A semantic adaptive framework for monitoring citizen satisfaction from e-government services », *Expert Systems with Applications*, vol. 37, no 6, 2010, p. 4292-4300.
- Manyika, J. et coll., *A Future that Works: Automation, Employment, and Productivity*, McKinsey Global Institute, 2017.



Marcus, G., « Deep Learning: A Critical Appraisal », ArXiv Preprint ArXiv:1801.00631, 2018. Sur Internet: bit.ly/2HJsBrd

Muralidharan, K., A. Singh et A.J. Ganimian, Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India, CESifo Working Paper No. 6328, Center for Economic Studies and Ifo Institute, 2017.

Mwalumbwe, I. et J.S. Mtebe, « Using learning analytics to predict students' performance in Moodle Learning Management Systems: A case of Mbeya University of Science and Technology », The Electronic Journal of Information Systems in Developing Countries, vol. 79, 2017, p. 1-13. DOI : 10.1002/j.1681-4835.2017.tb00577.x.

Nilsson, N.J., The Quest for Artificial Intelligence: A History of Ideas and Achievements, New York, Cambridge University Press, 2009.

Nissenbaum, H., « How computer systems embody values », Computer, vol. 34, no 3, 2001, p. 120-119. DOI : 10.1109/2.910905.

Nye, B.D., « Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context », International Journal of Artificial Intelligence in Education, vol. 25, no 2, 2015, p. 177-203. DOI : 10.1007/s40593-014-0028-6.

Osoba, O.A. et W. Welser IV, An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence, RAND Corporation, 2017. DOI : 10.7249/RR1744.

Penney, J., « Internet surveillance, regulation, and chilling effects online: A comparative case study », Internet Policy Review, vol. 6, no 2, 2017.

Petrovica, S., A. Anohina-Naumecca et H.K. Ekenel, « Emotion recognition in affective tutoring systems: Collection of ground-truth data », Procedia Computer Science, vol. 104, 2017, p. 437-444. DOI : 10.1016/j.procs.2017.01.157.

Phobun, P. et J. Vicheanpanya, « Adaptive intelligent tutoring systems for e-learning systems », Procedia – Social and Behavioral Sciences, vol. 2, no 2, 2010, p. 4064-4069. DOI : 10.1016/j.sbspro.2010.03.641.

Press, G., « Robot Overlords: AI at Facebook, Amazon, Disney and Digital Transformation at GE, DBS, BNY Mellon », Forbes Online, 28 mai 2017. Sur Internet: bit.ly/2KaZ9sB

Privacy International, « Algorithms, Intelligence, and Learning Oh My » [billet de blogue], 1er décembre 2016. Sur Internet: bit.ly/2HuzjOk (2017).

Idem, « Submission of evidence to the House of Lords Select Committee on Artificial Intelligence », Londres (Grande-Bretagne), 6 septembre 2017.

PwC, Sizing the Prize: What's the Real Value of AI for Your Business and How can you Capitalise?, 2017. Sur Internet: pwc.to/2vHve8j

Quinn, J., « Computational Techniques for Crop Disease Monitoring in the Developing World », dans A. Tucker et coll. (dir.), Advances in Intelligent Data Analysis XII, IDA 2013. Série Lecture Notes in Computer Science, Springer, vol. 8207, 2013. DOI : 10.1007/978-3-642-41398-8_2.

Quinn, J., V. Frias-Martinez et L. Subramanian, « Computational sustainability and artificial intelligence in the developing world », AI Magazine, vol. 35, no 3, 2014. Sur Internet: bit.ly/2HJQA9Q

R. Winthrop et E. McGivney, Can We Leapfrog? The Potential of Education Innovations to Rapidly Accelerate Progress, Center for Universal Education at Brookings, 2017.

Reyes, C.T., Developing a student support system through learning analytics for undergraduates at the University of the Philippines Open University, Quezon City (Philippines), Foundation for Information Technology Education and Development, 2018.

Singh, A. et coll., « Disguised Face identification (DFI) with Facial Keypoints Using Spatial Fusion Convolutional Network », ArXiv Preprint ArXiv:1708.09317, 2017.

Singh, A. et coll., « Disguised Face identification (DFI) with Facial Keypoints Using Spatial Fusion Convolutional Network », ArXiv Preprint ArXiv:1708.09317, 2017.

Stone, P. et coll., Artificial intelligence and life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015 Study Panel, Stanford (Californie), University Stanford, septembre 2016. Sur Internet: stanford.io/2HYuZrU

Suwajanakorn, S., S.M. Seitz et I. Kemelmacher-Shlizerman, « Synthesizing Obama: learning lip sync from audio », ACM Transactions on Graphics, vol. 36, no 4, 2017, p. 1-13. Sur Internet: bit.ly/2HGwCgj



Assemblée générale des Nations Unies, Document final de la réunion de haut niveau de l'Assemblée générale sur l'examen d'ensemble de la mise en oeuvre des textes issus du Sommet mondial sur la société de l'information, A/RES/70/125, 70e session, 16 décembre 2015. Sur Internet: bit.ly/2vGPtTA

Conseil des droits de l'homme des Nations Unies, La promotion, la protection et l'exercice des droits de l'homme sur Internet, A/HRC/RES/32/13, 32e session, 1er juillet 2016. Sur Internet: bit.ly/2vKayMZ

Van de Poel, I. et P. Kroes, « Can Technology Embody Values? », dans P. Kroes et P.-P. Verbeek (dir.), *The Moral Status of Technical Artefacts*, Dordrecht, Springer Netherlands, vol. 17, 2014, p. 103-124. DOI : 10.1007/978-94-007-7914-3_7.

Wachter, S., B. Mittelstadt et L. Floridi, « Why a right to explanation of automated decision-making does not exist in the general data protection regulation », *International Data Privacy Law*, 2017. Sur Internet (SSRN): bit.ly/2vHMiuW

World Wide Web Foundation, *Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries*, 2017. Sur Internet: bit.ly/2JnldyU

Yang, H.-H. et coll., « Improvement of e-government service process via a grey relation agent mechanism », *Expert Systems with Applications*, vol. 39, no 10, 2012, p. 9755-9763.

Zhao, J. et coll., « Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-level Constraints », *ArXiv Preprint ArXiv:1707.09457*, 2017. Sur Internet: bit.ly/2HWirKS

NOTES

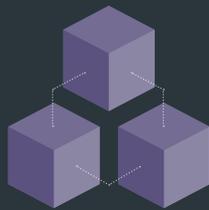
- 1 Ou encore, comme l'a qualifiée Jeff Bezos, « une couche habilitante horizontale » [Traduction]: bit.ly/2FeQ91P
- 2 Les objectifs de développement durable (ODD) sont un ensemble de 17 objectifs mondiaux s'inscrivant dans le cadre du Programme de développement durable des Nations Unies à l'horizon 2030. Voir: bit.ly/2HtQ7cJ
- 3 Tiré de la vidéo de l'UIT: bit.ly/2FhfeJA
- 4 bit.ly/2KfsYZ9
- 5 bit.ly/2HWYQAR
- 6 bit.ly/2HLU6R0
- 7 C'est ce qu'a mis en évidence Eric Horvitz, directeur des laboratoires de recherche de Microsoft et ancien président de l'Association for the Advancement of Artificial Intelligence, dans un éditorial de la revue *Science* (7 juillet 2017): L'excitation entourant l'IA a été tempérée par des préoccupations entourant ses inconvénients... préjugés enfouis profondément dans les ensembles de données, entraînant des déduction inévitables et inexactes... enjeux sur le plan juridique et éthique liés à des décisions prises par des systèmes autonomes, difficultés à expliquer les déductions, menaces aux libertés civiles en raison de nouvelles formes de surveillance, manipulation de précision destinée à la persuasion, utilisations criminelles de l'IA, influences déstabilisatrices d'applications militaires, et le potentiel d'abolir des emplois et d'amplifier les inégalités en matière de richesse. bit.ly/2qZPmh3
- 8 ibm.co/2qZcnzX
- 9 bit.ly/2Ho45wu
- 10 Nilsson 2009.
- 11 bit.ly/2JnZt68
- 12 *AI Now 2016: 2*
- 13 *PwC 2017*.
- 14 bit.ly/2vEKkLL
- 15 bit.ly/2qX1H5Q et bit.ly/2HSnXaz
- 16 Le « traitement du langage naturel (TLN) est la capacité d'un programme informatique de comprendre la parole humaine pendant qu'elle est parlée. » Tiré de: bit.ly/2HsiJCT
- 17 La « traduction automatique (TA) est la tâche consistant à convertir automatiquement une langue naturelle dans une autre, en préservant la signification du texte source, et en produisant un texte fluide dans la langue cible. » Tiré de: nlp.stanford.edu/projects/mt.shtml
- 18 Du point de vue de la philosophie, c'est également un moyen de déterminer l'« intelligence ». « Si un agent se comporte de façon intelligente, il est intelligent. C'est uniquement le comportement externe qui définit l'intelligence; agir de façon intelligente, c'est être intelligent. » bit.ly/2JnZt68
- 19 Voir: bit.ly/2HEacJ2
- 20 Voir: bit.ly/2vEKkLL



- 21 air.ug/microscopy/
- 22 aidr.qcri.org/
- 23 go.nature.com/2vYDy3u
- 24 bit.ly/2ra0iY
- 25 bit.ly/2jguJc3
- 26 bit.ly/2Kp8D3t
- 27 bit.ly/2KnAtwZ
- 28 bit.ly/2r9ZrYM
- 29 Voir, par exemple: bit.ly/2w0YA1C
- 30 bit.ly/2r9joik
- 31 contenttechnologiesinc.com
- 32 bit.ly/2r6Z68w
- 33 mck.co/2HDEwUm
- 34 bit.ly/2HzrwDk
- 35 bit.ly/2HzqXtc
- 36 bit.ly/2FrCkNL
- 37 bit.ly/2vRxIRs
- 38 bit.ly/2vSuSvq
- 39 tcrn.ch/2r7QizT
- 40 bit.ly/2JAM4aP
- 41 bit.ly/2vUDO3s
- 42 nyti.ms/2jeU0n2
- 43 Un rapport publié en 2016 par un consortium d'experts indique clairement qu'il « ne faut pas s'inquiéter et penser que l'IA représente une menace imminente pour l'humanité » (Stone et coll., 2016). Cependant, dans un jeu vidéo, l'IA a permis à un personnage non joueur de créer de super armes et à se mettre à traquer les joueurs, ce qui est un peu préoccupant ! Voir: bit.ly/2KkU0hE
- 44 bit.ly/2HLU6R0
- 45 bit.ly/2KnNx5y
- 46 « Mais des erreurs semblables sont apparues dans le logiciel d'une caméra Nikon qui interprète mal les images de personnes d'origine asiatique et les signale comme en train de faire un clin d'oeil, et dans le logiciel d'une caméra Web de Hewlett-Packard, qui éprouve de la difficulté à reconnaître les personnes à la peau foncée ». Même si le groupe n'avait pas conçu l'algorithme pour traiter les carnations claires comme un signe de beauté, les données entrées ont effectivement incité les robots juges à en venir à cette conclusion: bit.ly/2r7Oh6c
- 47 bit.ly/2jjZQ6V
- 48 bit.ly/2r9Z4gQ
- 49 bit.ly/2HC4dV3
- 50 bit.ly/2KmVisB
- 51 bit.ly/2r8hZYx bit.ly/2r6FD7K
- 52 bit.ly/2HBxL9u
- 53 bit.ly/2JEAueL
- 54 bit.ly/2JskYCW
- 55 bit.ly/2FqRPFQ
- 56 bit.ly/2JqyAyh
- 57 bit.ly/2J6xqZq
- 58 bit.ly/2JskYCW
- 59 bit.ly/2HwosU7
- 60 bit.ly/2HPNAbW
- 61 nyti.ms/2vVoaVR
- 62 bit.ly/2qZnrOd
- 63 Les risques que posent les mégadonnées pour la vie privée font l'objet de discussions et de préoccupations depuis un certain temps, et les sujets correspondants ont été abordés en détail dans un rapport de la Maison-Blanche, Big Data and Privacy, publié en 2014. bit.ly/2JocxbO
- 64 bit.ly/2JtxBh9
- 65 bit.ly/2Ju6CBL
- 66 bit.ly/2qYaDYj
- 67 bit.ly/2IMhrj2
- 68 bit.ly/2HqRPvo
- 69 bit.ly/2Kg4bnM
- 70 Documents sur la vie privée et l'IA: bit.ly/2Kfvm2a, bit.ly/2HrA0bn, bit.ly/2JocxbO, for.tn/2Jqwy17, timreview.ca/article/1067
- 71 En date de janvier 2018, il y avait plus de 2 milliards d'utilisateurs actifs de FB dans le monde. Voir: bit.ly/2HOHJbz
- 72 pewrsr.ch/2KcCcVU
- 73 Voir: bit.ly/2JuEb74
- 74 Voir les arguments de Yochai Benkler à cet effet: bit.ly/2JoeFjJ
- 75 bit.ly/2Ke7b41
- 76 bit.ly/2Kh3pXx
- 77 econ.st/2qYLjI5
- 78 nyti.ms/2vLSB0r
- 79 bit.ly/2Ke7b41
- 80 read.bi/2JqUZLW



- 81 lyrebird.ai/demo
- 82 bit.ly/2HrUosM
- 83 Voir un exemple de vidéo: bit.ly/2Hshtf2
- 84 bit.ly/2rauy6D
- 85 bit.ly/2KbPRMR
- 86 bit.ly/2Fe4oUO
- 87 go.nature.com/2HrH7EQ
- 88 bit.ly/2JmoXkg
- 89 bbc.in/2JoJPaF
- 90 bit.ly/2JmoXkg
- 91 bit.ly/2HoGawZ
- 92 bit.ly/2Kb6M2h
- 93 bit.ly/2Fdb7OI
- 94 bit.ly/2JmpsuC
- 95 bit.ly/2JmoXkg
- 96 bit.ly/2HJRYcy
- 97 bit.ly/2HXoaHb
- 98 bit.ly/2Jq3K97
- 99 bit.ly/2HVJJYn
- 100 Il convient de mentionner que cette perspective est largement influencée par notre position dans un organisme de financement de la recherche dans les pays du Nord, puisque nous avons des antécédents dans le financement de la recherche sur les technologies de l'information et de la communication au service du développement.
- 101 Voir: bit.ly/2IKzPcr
- 102 Par exemple, voir: bit.ly/2xuC2ac et bit.ly/2L9PBxR
- 103 Voir, par exemple: bit.ly/2K7IA1X
- 104 bit.ly/2Kew2EW
- 105 Ces recommandations sont également tirées d'autres importants rapports sur l'IA (AI Now, 2016; 2017; Calo, 2017; Privacy International, 2017; Stone et coll., 2016; Fondation World Wide, Web 2017; Brundage et coll., 2018).
- 106 Tous nos remerciements à Maroussia Levesque pour ce point.
- 107 « les mêmes droits dont les personnes disposent hors ligne doivent être aussi protégés en ligne » (Résolution du Conseil des droits de l'homme des Nations Unies L13, art. 1); Assemblée générale des Nations Unies 2015, art. 43; « [L'IA] devrait être conçue et exploitée de manière à respecter les droits de la personne, les libertés, la dignité humaine et la diversité culturelle » [Traduction] (IEEE, 2017 : 22).
- 108 Pour une discussion sur les avantages et les risques, voir Bostrom (2017).
- 109 bit.ly/2FhONn8
- 110 David Hubel et Torsten Wiesel se sont mérité le Prix Nobel de physiologie ou médecine en 1981, en partie pour leur découverte de neurones visuels sensibles à certaines caractéristiques comme les barres orientées et les grilles. Leurs travaux ont inspiré le développement du néocognitron, le premier réseau neuronal artificiel, proposé par Kunihiro Fukushima en 1982.
- 111 bit.ly/2JtMCPp
- 112 bit.ly/2HqmXep
- 113 bit.ly/2KhudHk
- 114 bit.ly/2Fh4JGf
- 115 bit.ly/2KfukDd
- 116 Voir, par exemple: bit.ly/2xvJTV2 et bit.ly/2shVF0i
- 117 bit.ly/2qYHnku



Intelligence artificielle et développement humain



IDRC | CRDI

International Development Research Centre
Centre de recherches pour le développement international

Canada