Intelligence artificielle et société

Sommaire et recommandations

Sommaire et recommandations

L'intelligence artificielle (IA) est l'une des technologies qui transforment notre société et de nombreux aspects de notre vie quotidienne. L'IA a déjà permis de nombreuses avancées et pourrait être une source de prospérité économique considérable. Elle soulève également des questions quant à l'emploi, la confidentialité des données, la protection de la vie privée, la violation de valeurs éthiques et la confiance dans les résultats.

Les décideurs politiques devraient encourager, et les scientifiques s'engager à :

- **Une gérance prudente pour contribuer au partage des bénéfices de l'IA au sein de la société**. Pour ce faire, une attention particulière devra être portée quant à l'impact de l'IA sur l'emploi, qui sera à son tour façonné par un éventail de facteurs, y compris les éléments politiques, économiques et culturels, ainsi que les progrès des technologies de l'IA.
- **Une fiabilité des systèmes et des données de l'IA**. Elle devrait être facilitée par des mesures traitant les problèmes de qualité, de parti pris et de traçabilité des données. Bien que ceci soit possible en améliorant l'accessibilité des données, les données personnelles ne devraient pas être mises à la disposition de parties non autorisées.
- **Une sûreté et une sécurité des systèmes et des données de l'IA**, essentielles dans le cas des applications qui impliquent une vulnérabilité humaine et nécessitent des systèmes assurément corrects.
- **Des recherches plus poussées pour contribuer au développement de systèmes d'IA permettant des explications**. Lorsque des décisions importantes ayant un impact sur les personnes sont suggérées par l'IA, les personnes concernées devraient recevoir suffisamment d'informations et pouvoir contester les décisions (par exemple refuser un traitement ou faire appel d'une décision).
- **Des connaissances issues de nombreux domaines afin de maximiser les avantages sociétaux de l'IA**. La recherche interdisciplinaire devrait inclure divers domaines tels que les sciences naturelles, les sciences de la vie et les sciences médicales, le génie, la robotique, les lettres et sciences humaines, les sciences économiques et sociales, l'éthique, l'informatique et l'IA elle-même.
- **Préparer les citoyens pour l'IA**. Un éventail de possibilités éducatives et d'informations sur l'IA devrait être offert ainsi qu'un dialogue construit avec les citoyens pour démystifier ce domaine.
- **Promouvoir le débat de politique publique sur l'utilisation destructrice/militaire de l'IA**. Les projets internationaux limitant les risques liés aux armes autonomes devraient être examinés par l'organisme compétent de l'ONU.
- **Encourager les échanges de talents et la coopération entre la recherche publique et le secteur privé**, pour faciliter le développement sûr et rapide d'applications dans des domaines bénéfiques pour l'être humain. La collaboration est importante pour la collecte à grande échelle de données cruciales pour le développement de systèmes d'IA.

Introduction

L'IA fait référence à un ensemble de méthodes et de technologies visant à faire fonctionner à bon escient des ordinateurs ou d'autres dispositifs. L'IA est essentiellement un ensemble d'algorithmes agissant sur des données (généralement des mégadonnées). L'apprentissage automatique (AA) est un sous-ensemble de l'IA qui traite des algorithmes permettant d'extraire des informations utiles à partir de données complexes. Des applications d'AA ont récemment eu des conséquences inattendues dans de nombreux domaines de la science et de la technologie. Il est largement admis que la recherche sur l'IA progresse de façon constante et que ses répercussions sur la société augmenteront probablement à l'avenir.

Le développement de systèmes algorithmiques sophistiqués, combiné à la disponibilité des données et à la puissance de traitement, a engendré des résultats remarquables dans diverses tâches spécialisées telles que la reconnaissance vocale, la classification d'images, la détection de défauts, les véhicules autonomes, les systèmes d'aide à la prise de décisions, la robotique, la traduction automatique, la locomotion à pattes et les systèmes de réponse aux questions. Certaines de ces applications fournissent des outils de soutien extrêmement précieux pour les personnes en situation de handicap. En utilisant des interfaces cerveau-machine, des personnes paralysées peuvent interagir avec leur environnement par l'intermédiaire d'un ordinateur.

Dans le domaine des sciences naturelles et sociales, les algorithmes d'apprentissage automatique permettent des progrès et fournissent de nouveaux outils pour le traitement et la modélisation de données et de processus complexes, offrant d'énormes avantages potentiels. Étant donné qu'une grande partie de ce que la civilisation a à offrir est un produit de l'intelligence humaine, imaginez ce qui pourrait être réalisé si cette intelligence était amplifiée par les outils offerts par l'IA.

Certaines questions et préoccupations quant aux pièges potentiels requièrent cependant une réflexion plus profonde.

Les progrès actuels de la recherche sur l'IA permettent de concentrer les efforts non seulement sur l'amélioration des capacités de l'IA, mais aussi sur la maximisation de ses avantages pour la société tout en respectant les valeurs éthiques. L'évolution et les développements techniques de l'IA devraient donc suivre les considérations éthiques. Des préoccupations émergent par rapport aux partis pris qui pourraient découler des systèmes d'IA s'appuyant sur l'analyse de données statistiques et l'apprentissage automatique.

Dans ce contexte général, nous abordons tout d'abord les problèmes posés par l'impact économique profond de l'IA. Deuxièmement, nous abordons les propriétés générales que devraient posséder les systèmes d'IA en vue d'interagir de manière satisfaisante et éthique avec les humains. Nous abordons ensuite des problématiques plus précises liées à l'utilisation des systèmes d'IA dans le domaine de la santé, des questions soulevées par les applications possibles de l'IA aux armes autonomes et nous penchons sur le potentiel de l'intelligence artificielle intégrée aux systèmes robotiques. Cette analyse résulte en un ensemble de recommandations regroupées dans le sommaire.

1. Gérer et optimiser l'impact de l'IA sur nos sociétés

Les économistes et les informaticiens s'entendent généralement pour dire que la recherche doit être menée de manière à maximiser les avantages économiques de l'IA tout en atténuant ses effets négatifs. En ce moment, il est important de prendre en considération l'impact possible de l'IA en termes d'inégalité accrue, de chômage et de comportements contraires à l'éthique. Ces questions en suspens sont examinées plus en détail ci-dessous.

1.1. Prévisions du marché du travail

L'IA pourrait engranger d'importants avantages économiques. Dans tous les secteurs, les technologies de l'IA font la promesse d'accroître la productivité et de créer de nouveaux produits et services. Ce potentiel soulève des questions par rapport aux répercussions de l'IA sur l'emploi et sur la vie professionnelle.

L'IA provoquera vraisemblablement des perturbations considérables sur le travail, en supprimant, créant ou modifiant des emplois. Les études établissant des projections sur l'impact de l'IA sur l'emploi présentent des degrés élevés d'incertitude quant au rythme des transformations et à la proportion de tâches ou d'emplois qui pourraient être automatisés.

À plus long terme, les technologies contribuent à accroître la productivité et la richesse de la population. Toutefois, ces avantages peuvent prendre du temps à se manifester, et il peut y avoir des périodes où certaines parties de la population n'en profitent pas. Cet état de fait indique qu'il pourrait y avoir d'importants effets transitoires causant des perturbations pour certaines personnes ou certains lieux, et aggravant potentiellement les inégalités sociétales à court terme. Des recherches visant à anticiper l'impact économique et sociétal d'une telle disparité, en tenant compte de la vulnérabilité de certains emplois par rapport à l'automatisation, sont clairement nécessaires. Il sera plus facile d'analyser l'impact des systèmes d'IA sur divers types d'emplois (ceux requérant des travailleurs peu qualifiés et ceux exigeant des professionnels hautement qualifiés) que de prédire quels emplois seront créés à l'avenir dans le cadre de diverses mesures. Les technologies d'IA pourraient suivre un certain nombre de pistes plausibles pour se développer. Plusieurs facteurs joueront un rôle dans la détermination de l'impact de l'IA sur l'emploi, y compris les facteurs politiques, économiques, sociaux et culturels, ainsi que les capacités des technologies d'IA. Le recours aux meilleurs résultats de recherches disponibles dans toutes les disciplines peut contribuer à l'élaboration de politiques permettant le partage des bénéfices de ces changements technologiques au sein de la société.

1.2. Politiques de gestion et d'intégration du développement de l'IA dans la société

L'IA aura des conséquences importantes sur de nombreux secteurs de la société, augmentant ou remplaçant le travail humain. Le défi consiste à anticiper ces changements et à élaborer des politiques qui limiteront les effets négatifs et permettront une meilleure intégration de l'IA. L'éducation est essentielle à la fois pour favoriser l'adoption de l'IA et pour lutter contre les inégalités.

Une compréhension de base de l'utilisation des données et des technologies de l'IA est nécessaire à tous les âges, non seulement chez les producteurs et les utilisateurs professionnels de l'IA, mais aussi chez tous les citoyens. L'introduction de concepts clés

dans les écoles peut garantir cette compréhension. L'adoption d'un programme d'études équilibré pour les jeunes dans le domaine des sciences, des mathématiques, de l'informatique, des arts, des lettres et des sciences humaines pourrait leur permettre d'acquérir un large éventail de compétences et leur fournir une base plus solide pour l'apprentissage tout au long de la vie.

La demande en employés hautement qualifiés est également élevée. Plusieurs secteurs et professions auront besoin de compétences pour utiliser efficacement l'IA. De nouvelles initiatives peuvent aider à créer un bassin d'utilisateurs expérimentés en systèmes d'IA. Il est également nécessaire de soutenir les nouvelles filières d'apprentissage et les infrastructures pour développer des compétences avancées en IA qui permettront de nouvelles applications et la création de nombreux nouveaux emplois.

Ces questions faisaient déjà partie intégrante de la déclaration d'Ottawa sur « Notre avenir numérique et son impact sur la connaissance, l'économie et la main-d'œuvre », publiée lors du dernier Sommet du G7. Les gouvernements sont encouragés à mettre en œuvre des politiques qui seront inclusives et qui donneront à chaque citoyen un accès équitable aux prestations d'IA. Pour ce faire, la qualité, la sécurité et la résilience de l'information doivent être garanties, ainsi que la transparence, l'ouverture et l'interopérabilité des systèmes d'IA.

Dans les domaines où les capacités d'IA ont surpassé la réglementation actuelle, il pourrait être nécessaire d'établir de nouvelles approches de gouvernance tenant compte des questions éthiques découlant de l'interaction humaine avec les machines intelligentes. Il convient de souligner le rôle des lettres, des sciences humaines et des sciences sociales en général, en partenariat avec les développeurs et les utilisateurs, pour explorer les manières dont l'IA pourrait remettre en question les normes éthiques actuelles ou révéler les nouveaux défis éthiques qu'elle pose.

2. Caractéristiques des systèmes d'IA qui devraient être encouragées

2.1. Les données

Notre capacité à tirer pleinement parti de la synergie entre l'IA et les mégadonnées dépendra partiellement de notre capacité à acquérir, évaluer de façon critique et gérer les données. La plupart des technologies actuelles d'IA requièrent un accès à d'énormes volumes de données. Pour tirer pleinement parti de la technologie, de nouveaux cadres pourraient être nécessaires pour disposer de données. C'est notamment le cas pour les données ouvertes et les données privées d'intérêt public, pour lesquelles de nouvelles normes pourraient s'avérer nécessaires afin de garantir une utilisation efficace des données. Par exemple, il faudra veiller à rendre explicite la signification des données, à décrire le contexte dans lequel elles ont été obtenues et à donner des informations sur leur origine et leur traitement. Toutes ces questions peuvent être traitées par des techniques d'IA, qui peuvent donc être importantes pour tenir les nombreuses promesses véhiculées par les données ouvertes et assurer l'interopérabilité entre les différents types de données (sociales, économiques, organisationnelles et techniques).

Parallèlement, l'accès à des ensembles de données de haute qualité devrait respecter la confidentialité des données à caractère personnel et répondre aux préoccupations relatives aux préjugés et aux droits individuels. Des efforts intensifs devraient être déployés pour réglementer l'accès aux données confidentielles par des tiers tels que les banques, les

compagnies d'assurance et les employeurs potentiels. Les ensembles de données doivent être protégés contre les attaques malveillantes. Des politiques régissant la collecte, le partage et l'accès aux données devraient être mises en place non seulement pour les grandes entreprises, mais également pour les codes sources libres.

2.2. Rendement et possibilité d'explication

Certaines des avancées les plus réussies et les plus populaires de l'IA - notamment l'apprentissage approfondi – permettent peu d'explications pour le moment, et différentes méthodes d'IA prennent en charge différentes possibilités d'explication. Dans certains cas, cela pourrait diminuer la confiance des utilisateurs dans de tels outils. Certains domaines ne peuvent exister sans explications : dans les applications médicales, un diagnostic rendu sans explication aurait peu de chances d'être accepté. Les compromis entre la performance et les possibilités d'explication devraient être explicités pendant le développement de nouveaux modèles. Les limites des algorithmes mis en place doivent être décrites pour permettre aux utilisateurs de comprendre les motifs des décisions proposées par les systèmes d'IA. L'amélioration des possibilités d'explication de l'IA peut contribuer au respect de l'objectivité des systèmes d'IA. L'effet disparate est le concept juridique et théorique prédominant utilisé pour désigner la discrimination involontaire produite par l'application d'algorithmes lorsqu'un attribut personnel (comme l'origine ethnique, les origines sociales, le sexe et l'âge) a un effet direct sur les décisions prises par l'algorithme. Les systèmes d'IA utilisés pour la prise de décisions ayant un impact profond sur la vie quotidienne de la population ne devraient pas générer d'effets disparates indésirables.

2.3. Vérification et validation des systèmes évolutifs en ligne

Les systèmes en ligne évoluent dans le temps en fonction des données qu'ils récoltent en permanence. Il est apparu récemment qu'un système d'IA pouvait s'écarter de son état initial d'une manière indésirable, par exemple en ce qui concerne le sexe et la race. Il est donc nécessaire de contrôler les résultats des systèmes évolutifs en ligne afin de détecter des évolutions indésirables.

3. Domaines d'application exemplaires et conséquences sociétales

3.1. Applications dans le domaine des soins de santé

L'IA offre d'importants bénéfices potentiels dans les systèmes de soutien à la prise de décisions en matière de santé et de soins. Les problèmes structurels dans ce domaine peuvent conduire à des erreurs de diagnostic, à d'éventuels manquements de l'expertise et à une inefficacité de la communication de l'information entre les mondes de la recherche, du génie et de la clinique. L'IA peut contribuer à évaluer d'énormes quantités de publications scientifiques, à repérer les corrélations faibles et peu plausibles dans d'énormes quantités de données, à analyser des images et d'autres données produites par les systèmes de soins de santé et à élaborer de nouvelles technologies. En raison de l'importance vitale de l'amélioration des systèmes d'aide à la prise de décision clinique, l'IA peut grandement aider les cliniciens grâce à une gamme d'outils et de dispositifs d'aide à la prise de décision en matière de diagnostic et d'options thérapeutiques. L'objectif est l'amélioration de l'interprétation des observations et des mesures, de l'établissement des diagnostics et de la précision, de l'efficacité et de l'accessibilité des soins de santé. Pour ce faire, il faut concevoir le système prudemment, en tenant compte des manières dont l'IA peut se

développer aux côtés des utilisateurs humains, du caractère interprétable qui pourrait être nécessaire dans différents contextes, et des moyens de vérifier et de valider ces systèmes. Les médecins et les patients doivent avoir confiance dans de tels systèmes et ces systèmes doivent fonctionner de manière pertinente pour divers groupes d'utilisateurs.

Une gouvernance prudente des données est également nécessaire. Les collaborations internationales visant à accélérer les avancées dans le domaine de l'IA servent les intérêts des citoyens de tous les pays.

3.2. Armes autonomes

L'IA ouvre de nouvelles possibilités d'applications militaires, en particulier en ce qui concerne les systèmes d'armes autonomes utilisées pour la sélection et l'attaque de cibles. Ces armes autonomes pourraient conduire à une nouvelle course aux armements, abaisser le seuil définissant la notion de guerre ou devenir un outil pour les oppresseurs ou les terroristes. Certaines organisations demandent l'interdiction des armes autonomes, à l'instar de conventions dans le domaine des armes chimiques ou biologiques. Une telle interdiction nécessiterait une définition précise des termes « armes » et « autonomie ». En l'absence d'une interdiction des systèmes d'armes létales autonomes (SALA), tout système d'armes devrait être conforme au droit international humanitaire. Ces armes devraient être intégrées dans les structures existantes de commandement et de contrôle de telle sorte que la responsabilité légale continue d'incomber à des acteurs humains spécifiques. La transparence et le débat public sont clairement nécessaires sur les questions soulevées dans ce domaine.

3.3. Robotique

Les robots détectent et déplacent des machines, incarnant l'IA. Le contact physique de la machine avec l'environnement, y compris les humains, est un défi. Les robots doivent être sûrs, fiables et sécurisés. Jusqu'à récemment, les robots étaient principalement utilisés dans l'industrie manufacturière et dédiés à des situations spécifiques sans partager l'espace avec l'humain. Aujourd'hui, à la suite de la deuxième vague de développement de la robotique, les robots partagent davantage l'espace et interagissent avec les humains. Alors que les applications d'IA se concentrent sur les technologies de traitement de données pour en tirer des connaissances utiles à la prise de décision, le but ultime de la robotique est de créer des systèmes techniques capables d'interagir avec le monde physique.

Outre l'utilisation d'algorithmes d'apprentissage automatique, la robotique subit des contraintes fondamentales en termes de sécurité physique. La conception de la robotique nécessite une certification d'un logiciel et une vérification formelle afin de maximiser la tolérance aux pannes, la fiabilité et la capacité de survie.

Malgré les progrès récents, les attentes en matière de progrès surestiment souvent le rythme de l'évolution technologique.

Enfin, dans l'imaginaire populaire, la robotique, et plus généralement l'IA, est influencée par les récits fantastiques plutôt que par des éléments probants. Il est important de démystifier et de diffuser la robotique et la science de l'IA en s'engageant dans l'éducation publique, la discussion et le débat avec tous les citoyens.

Royal Society
Canada

Chad Gaffield

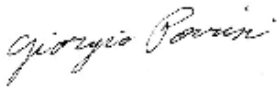
Académie des sciences
France

Pierre Corvol

Deutsche Akademie der Naturforscher Leopoldina
Germany

Jörg Hacker

Accademia Nazionale dei Lincei
Italy



Giorgio Parisi

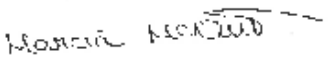
Science Council
Japan

Juichi Yamagiwa

Royal Society
United Kingdom

Venkatraman «Venki» Ramakrishnan

National Academy of Sciences
United States of America



Marcia McNutt