Intelligence artificielle et société

Résumé et recommandations

L'intelligence artificielle (IA) est l'une des technologies qui transforme notre société et de nombreux aspects de notre vie quotidienne. L'IA a déjà procuré de nombreux avantages et elle pourrait être une source de prospérité économique considérable. Elle soulève également des questions sur l'emploi, la confidentialité des données, la vie privée, la violation des valeurs éthiques et la confiance dans les résultats. Les décideurs politiques devraient encourager la prise en compte des points suivants qui devraient également mobiliser les scientifiques :

Une gestion prudente est nécessaire pour aider à partager les bénéfices de l'IA dans l'ensemble de la société

. Pour cela, il faudra porter une attention particulière à l'impact de l'IA sur l'emploi, qui sera à son tour influencé par une série de facteurs, notamment des aspects politiques, économiques et culturels, ainsi que par les progrès des technologies d'IA.

Les systèmes et les données d'IA doivent être fiables. Cela devrait être facilité par des mesures portant sur la qualité, l'absence de biais et la traçabilité des données. Bien que cela puisse être encore facilité en rendant les données plus accessibles, les données à caractère personnel ne devraient pas être mises à la disposition de tiers non autorisés.

Les systèmes et les données d'IA doivent être sûrs et sécurisés. Ceci est essentiel dans le cas d'applications qui impliquent une vulnérabilité humaine et qui peuvent nécessiter des systèmes dont il est prouvé qu'ils sont corrects.

Des recherches sont nécessaires pour aider à mettre au point des systèmes d'IA explicables. Lorsque des décisions importantes suggérées par l'IA ont une incidence sur des personnes, les individus concernés devraient recevoir une information suffisante et être autorisés à contester ces décisions (par exemple refuser un traitement ou faire appel d'une décision).

Des connaissances dans de nombreux domaines sont nécessaires pour tirer le maximum de bénéfices sociétaux de l'IA

. La recherche interdisciplinaire devrait porter sur divers domaines tels que les sciences de la nature, les sciences de la vie et les sciences médicales, l'ingénierie, la robotique, les sciences humaines, les sciences économiques et sociales, l'éthique, l'informatique et l'IA elle-même.

Les citoyens doivent être prêts à l'IA. Un éventail de possibilités de formation et d'information sur l'IA devrait être mis à leur disposition et un dialogue bien fondé avec les citoyens devrait être engagé pour démystifier ce domaine.

Introduction

L'IA fait référence à un ensemble de méthodes et de technologies visant à faire fonctionner intelligemment des ordinateurs ou d'autres dispositifs. L'IA consiste essentiellement en un ensemble d'algorithmes fonctionnant sur des données (généralement volumineuses). L'apprentissage machine (ML) est un sous-ensemble de l'IA qui traite des algorithmes d'extraction d'informations utiles, à partir de données complexes. Les applications de l'apprentissage machine ont eu récemment un impact inattendu dans de nombreux domaines de la science et de la technologie. Il existe un large consensus sur la progression régulière de la recherche sur l'IA et sur l'augmentation probable de son impact sur le futur de la société. Le développement de systèmes algorithmiques sophistiqués, combiné à la disponibilité des données et à la puissance de traitement, a conduit à des résultats remarquables pour une série de tâches spécialisées telles que la reconnaissance vocale, la classification d'images, la détection de défauts, les véhicules autonomes, les systèmes d'aide à la décision, la robotique, la traduction automatique, la locomotion de robots humanoïdes, et les systèmes automatiques de réponse aux questions. Certaines de ces applications fournissent des outils de soutien extrêmement précieux pour les personnes handicapées. Grâce à des interfaces cerveau-machine, les individus paralysés peuvent interagir avec leur environnement au moyen d'un ordinateur. Dans le domaine des sciences de la nature et dans celui des sciences sociales, les algorithmes d'apprentissage machine permettent des progrès et fournissent de nouveaux outils pour le traitement et la modélisation de données et de processus complexes, avec d'énormes avantages potentiels. Étant donné qu'une grande partie de ce que la civilisation a à offrir est issue de l'intelligence humaine, nous ne pouvons qu'imaginer ce qui pourrait être accompli lorsque cette intelligence sera amplifiée par les outils que l'IA peut fournir. Il y a toutefois un certain nombre de questions et d'inquiétudes sur des écueils potentiels qui méritent un examen plus approfondi. Les progrès 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 bénéfices pour la société tout en respectant les valeurs éthiques. Le déploiement et l'évolution technique de l'IA devraient donc être guidés par des considérations éthiques. On craint de plus en plus que des biais puissent être générés par les systèmes d'IA fondés sur l'analyse de données statistiques et l'apprentissage automatique. Dans ce contexte général, on traitera en premier lieu les problèmes posés par l'impact économique transformatif de l'IA. Puis, dans un deuxième temps des propriétés générales dont les systèmes d'IA devraient disposer pour interagir de façon satisfaisante et éthique avec les humains. On abordera ensuite des questions plus spécifiques liées à l'utilisation des systèmes d'IA dans le domaine de la santé, des questions soulevées par d'éventuelles applications de l'IA à des systèmes d'armes autonomes, et on considèrera le potentiel de l'IA intégrée dans les systèmes robotiques. Cette analyse conduit à un ensemble de recommandations rassemblées dans le résumé de ce document.

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

Les économistes et les informaticiens s'accordent généralement pour dire qu'il faut faire de la recherche afin de maximiser les bénéfices économiques de l'IA tout en en atténuant les effets négatifs. A ce stade, il est important de considérer l'impact possible de l'IA en termes d'accroissement des inégalités, de chômage et de comportements non éthiques. Ces questions en suspens sont examinées plus en détail dans ce qui suit.

1.1 Prévisions du marché du travail

L'IA pourrait apporter des avantages économiques importants : dans tous les secteurs, les technologies de l'IA offrent la promesse d'accroître la productivité et de créer de nouveaux produits et services. Ce potentiel soulève des questions sur l'impact de l'IA sur l'emploi et la vie professionnelle. L'IA aura probablement un effet perturbateur considérable sur le travail, certains emplois seront perdus, d'autres seront créés et d'autres enfin seront en mutation. Les études de projections sur l'impact de l'IA sur l'emploi comportent un degré élevé d'incertitude quant à la vitesse des changements et à la proportion des tâches ou des emplois susceptibles d'être automatisés. A plus long terme, les technologies contribueront à accroître la productivité et la richesse de la population. Toutefois, ces avantages peuvent prendre du temps à se manifester, et on peut vivre des périodes au cours desquelles une partie de la population n'éprouvera que les inconvénients. Cela donne à penser que des effets transitoires importants pourraient apparaître et entrainer des perturbations pour certaines personnes ou certains lieux, et potentiellement aggraver les inégalités sociales à court terme. Il est clairement nécessaire de mener des recherches pour anticiper l'impact économique et sociétal d'une telle disparité, en tenant compte de la vulnérabilité des emplois à l'automatisation. Il sera plus facile d'analyser l'impact des systèmes d'IA sur divers types d'emplois, ceux qui nécessitent des travailleurs peu qualifiés et ceux qui ont besoin de professionnels hautement qualifiés, que de prévoir les emplois qui pourraient être créés à l'avenir dans le cadre de politiques diverses. Il existe un certain nombre de pistes plausibles pour le développement futur des technologies de l'IA. Une série de facteurs joueront un rôle dans la détermination de l'impact de l'IA sur l'emploi, y compris des éléments politiques, économiques et culturels, ainsi que les capacités des technologies de l'IA. L'utilisation des meilleures données de recherche disponibles dans toutes les disciplines peut aider à élaborer des politiques qui feront partager les avantages de ces changements technologiques à l'ensemble de la société.

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

L'IA aura un impact important sur toute une série de secteurs de la société, en augmentant ou en 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 d'IA est nécessaire à tous les âges, non seulement pour les producteurs et les utilisateurs professionnels de l'IA, mais aussi pour tous les citoyens. L'introduction de concepts clés dans les écoles peut aider à y parvenir. L'adoption d'un programme d'études large et équilibré pour l'éducation des jeunes dans les domaines des sciences, des mathématiques, de l'informatique, des arts et des sciences humaines pourrait leur permettre d'acquérir un large éventail de compétences et fournir une base plus solide pour l'apprentissage tout au long de la vie. Il y a aussi une forte demande de recrutement de personnes hautement qualifiées. De nombreux secteurs et de professions nécessiteront des compétences pour utiliser de l'IA d'une manière qui leur soit utile. De nouvelles initiatives peuvent aider à créer un ensemble d'utilisateurs avertis des systèmes d'IA. Il est également nécessaire de soutenir de nouvelles filières d'apprentissage et des infrastructures pour développer des compétences avancées en IA qui permettront de nouvelles applications et la création en nombres de nouveaux emplois. Ces questions faisaient déjà partie de la déclaration d'Ottawa rédigée lors du dernier sommet du G7 « Réaliser notre avenir numérique et façonner son impact sur le savoir, l'industrie et la main-d'œuvre ». Les gouvernements sont encouragés à mettre en œuvre des politiques qui seront inclusives et capables de fournir à chaque citoyen un accès équitable aux prestations de l'IA. Cela suppose que la qualité, la sécurité et la résilience de l'information soient également garanties, de même que la transparence, l'ouverture et l'interopérabilité des systèmes d'IA.

Dans les domaines où les capacités de l'IA ont dépassé la réglementation actuelle, il pourrait être nécessaire d'adopter de nouvelles approches de gouvernance qui tiennent compte des questions éthiques soulevées par l'interaction humaine avec des machines intelligentes. Il convient de souligner le rôle des sciences humaines et sociales en général et du partenariat avec les concepteurs et les utilisateurs pour explorer les façons dont l'IA peut remettre en question les normes éthiques existantes ou pour identifier les nouveaux défis éthiques de l'intelligence artificielle.

2. Caractéristiques des systèmes d'IA qui devraient être encouragées 2.1 Données

Notre capacité à tirer pleinement parti de la synergie entre l'IA et les données massives dépendra en partie de notre capacité à acquérir, évaluer de façon critique et gérer les données. Une grande partie de la technologie actuelle de l'IA nécessite l'accès à d'énormes volumes de données. Pour tirer pleinement parti de ces technologies, de nouveaux cadres réglementaires peuvent être nécessaires pour que les données soient disponibles. C'est notamment le cas des données ouvertes et des données privées d'intérêt public, pour lesquelles de nouvelles normes pourraient s'avérer nécessaires afin de garantir une utilisation efficace des données. Il faudra par exemple, s'efforcer de rendre explicite la signification des données, ainsi qu'une représentation du contexte dans lequel elles ont été obtenues et des informations sur leur origine et leur traitement. Toutes ces questions peuvent être abordées par des techniques d'IA, qui peuvent donc être importantes pour tenir les multiples promesses des données ouvertes et assurer l'interopérabilité entre différents types, par exemple, sociaux, économiques, organisationnels et techniques. Dans le même temps, l'accès à des ensembles de données de haute qualité devrait respecter la vie privée et la confidentialité des données personnelles et répondre aux préoccupations concernant les biais injustifiés et le respect des droits individuels. Tout doit être mis en œuvre pour que l'accès aux données confidentielles par des tiers tels que les banques, les compagnies d'assurance, les employeurs potentiels soit régi par des réglementations. 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 en place non seulement pour les grandes entreprises, mais aussi pour les initiatives « open source ».

2.2Rendement et possibilité d'explication

Certains des développements les plus réussis et les plus populaires de l'IA - notamment l'apprentissage approfondi - souffrent actuellement de faibles niveaux d'explicabilité et différentes méthodes d'IA demandent divers types d'explicabilité, ce qui pourrait, dans certains cas, réduire la confiance que les utilisateurs accordent à de tels outils. Certains domaines requièrent des explications :dans les applications médicales, un diagnostic sans explication a peu de chances d'être acceptable. Les compromis entre la performance et l'explicabilité devraient être explicités tout en visant à développer des modèles plus explicables. Les limites des algorithmes implémentés doivent être décrites pour permettre aux utilisateurs de comprendre les raisons des décisions proposées par les systèmes d'IA. L'amélioration de l'explicabilité de l'IA peut aider à s'assurer que le système d'IA n'introduit pas de biais. L'impact différencié (« disparate impact ») est apparu comme étant 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, sociale, le sexe et l'âge) a un effet direct sur les décisions prises par l'algorithme. Les systèmes d'IA utilisés pour prendre des décisions qui ont un impact profond sur la vie quotidienne des gens ne devraient pas générer un impact différencié indésirable.

2.3 Vérification et validation des systèmes évolutifs en ligne

Les systèmes en ligne évoluent dans le temps en fonction des données qu'ils traitent en permanence. Il est récemment apparu clairement qu'un système d'IA peut s'éloigner de son état initial d'une manière non souhaitée, par exemple pour le genre et la race. L'évolution des systèmes en ligne nécessite donc une surveillance de leur production pour éventuellement détecter des évolutions indésirables.