DÉCODE LES ALGORITHMES ET LES DONNÉES

Résumé du document bref sur les IA dignes de confiance

digital 2030
par digital moment

digital 2030
by digital moment

Résumé du document bref sur les IA dignes de confiance

Ce document est une version abrégée du document bref de Mozilla sur les IA dignes de confiance.

Résumé

Qu'est-ce qu'une IA?

« IA » signifie « intelligence artificielle ». Une IA est une simulation (à savoir une imitation) par des machines de la façon dont les humains réfléchissent et agissent. Les machines recueillent des informations (telles que des données) chaque fois que les humains les utilisent, le but étant de devenir plus intelligentes afin de ressembler davantage aux humains et de mieux faire ce pour quoi elles ont été conçues. Parmi les applications concrètes de l'IA, on trouve notamment les divers systèmes capables de comprendre la parole ou les images, les véhicules autonomes, la reconnaissance faciale, les recommandations personnalisées offertes par YouTube ou Netflix, ou encore les moteurs de recherche sur le web.

Qu'est-ce qu'une IA digne de confiance?

Mozilla utilise le terme « IA digne de confiance » pour désigner les types d'IA dont on sait qu'elles sont suffisamment sûres pour que les humains puissent leur faire confiance. Dire qu'une IA est « digne de confiance » signifie que cette IA :

- Respecte la vie privée des humains qui l'utilisent;
- Indique clairement quelles données elle recueille et comment elle les utilise pour faire des recommandations;

- Prend en compte l'effet qu'elle peut avoir sur les humains et leurs émotions, et adapte ses actions pour éviter de leur faire du mal (ou de les manipuler injustement);
- Peut être arrêtée si elle fait du mal à des humains.

Pourquoi est-il important de créer des IA dignes de confiance?

L'IA peut contribuer à améliorer notre vie, mais elle est tellement puissante qu'elle peut aussi faire beaucoup de dégâts. Les recherches de Mozilla ont démontré que la manière dont les IA sont actuellement développées peut parfois nuire au bien-être des humains. Mozilla souhaite recueillir l'avis d'un grand nombre de personnes différentes au sujet de l'IA afin de pouvoir émettre des recommandations visant à rendre les IA dignes de confiance et leur utilisation sécuritaire. Ces résultats permettront d'aider les grosses entreprises et les gouvernements à collaborer afin de créer de meilleures technologies aussi utiles que sûres.

Introduction

Comment l'IA fonctionne-t-elle à l'heure actuelle et qu'est-ce qui pourrait être fait différemment?

À l'instar d'un·e enfant qui va à l'école, l'IA « apprend ». Ses enseignant·e·s, à savoir les programmeur·euse·s, lui donnent des cours. Cependant, au lieu de s'appuyer sur des manuels, ils et elles enseignent à l'IA en utilisant des données. Les données sont des informations, par exemple des noms et des chiffres, ou encore des idées telles que « les enfants qui aiment regarder des vidéos sur Roblox aiment aussi regarder des vidéos sur Minecraft ». L'IA utilise ensuite les informations qu'elle a apprises pour résoudre des problèmes tels que : « Quelle vidéo dois-je recommander ensuite à cette personne qui est en train de regarder une vidéo sur Roblox? »

Cela peut poser un problème, car nous ne savons pas toujours quelles données sont recueillies ou comment celles-ci sont utilisées. De plus, ce n'est pas une méthode infaillible. Par exemple, certains enfants regardent des vidéos sur Minecraft, mais n'aiment pas les vidéos sur Roblox. D'ailleurs, les enfants n'ont peut-être pas envie que l'IA sache quelles vidéos ils et elles aiment!

Alors, à quoi pourrait bien ressembler une IA digne de confiance? Les grands comme les petits pourraient-ils décider de son fonctionnement? Serions-nous en mesure de choisir les informations que nous souhaitons divulguer (pour obtenir des recommandations pertinentes)? Comment pouvons-nous garantir la confidentialité des informations que nous ne souhaitons pas partager?

Quels sont les acteur·rice·s concernés par l'émergence de l'IA?

Les consommateur·rice·s

Adultes, enfants, enseignant·e·s, etc. : il s'agit de toutes les personnes qui utilisent une IA ou un service reposant sur l'IA (comme un moteur de recherche). Les consommateur·rice·s ne savent pas toujours précisément comment leurs informations sont recueillies ou utilisées, ce qui peut être inquiétant, surtout lorsqu'on commence à avoir l'impression que les machines écoutent nos conversations ou lisent nos courriels. Lorsque les consommateur·rice·s se tournent vers les entreprises pour en savoir plus, les réponses qu'ils et elles obtiennent sont souvent longues et complexes, ce qui crée encore plus de confusion.

Le secteur de l'IA

Il s'agit des personnes et des entreprises qui développent des produits reposant sur l'IA. Si vous souhaitez que votre IA soit performante, vous avez besoin de beaucoup de données. Parfois, les membres du secteur de l'IA prennent des raccourcis pour obtenir ces données, car ils souhaitent avoir la meilleure IA possible. Mais cette pratique peut être néfaste pour les consommateur·rice·s, qui ne savent pas forcément que l'IA recueille leurs données.

Cependant, au fil du temps, le secteur de l'IA essaie de prendre davantage en compte le bien-être des consommateur·rice·s lorsqu'il développe des systèmes d'IA. À votre avis, quels éléments devrait-il prendre en compte?

Les autorités de réglementation

Il s'agit là du gouvernement et, à l'occasion, d'autres groupes professionnels chargés de surveiller ce que fait le secteur de l'IA. Ces entités s'interrogent souvent sur la manière dont l'IA peut être utilisée et elles établissent des règles encadrant cette technologie. Cependant, l'IA évolue tellement vite que plusieurs gouvernements ne réussissent pas à faire passer de nouvelles lois à temps.

Quels sont les défis liés à l'IA?

Si l'IA apprend des informations erronées, ou trop d'informations d'un type et pas assez de l'autre, elle peut devenir **biaisée**. Cela signifie qu'elle prend des décisions injustes. Elle peut également s'immiscer dans la vie privée des consommateur·rice·s, car elle ne fait que suivre les instructions qui lui ont été données lors de sa programmation. L'IA ne comprend pas les conséquences de ses actes ou ce que nous pouvons ressentir.

Voici certains des problèmes majeurs liés à l'IA :

- **Monopole et centralisation :** Dans la mesure où une bonne IA nécessite de grandes quantités de données, l'entreprise qui peut en recueillir le plus est donc celle qui possède la meilleure IA et qui gagne le plus d'argent. La concurrence est donc rude pour les entreprises de petite taille ou qui débutent sur le marché. Cela signifie également qu'un petit nombre d'entreprises en sait beaucoup sur nous et que, si l'on n'établit pas des règles, nous ne pourrons rien faire pour protéger nos données.
- **Confidentialité des données et réglementation :** Puisque les IA ont besoin de beaucoup de données, les entreprises recueillent parfois ces dernières à notre

insu via Internet afin de perfectionner leur IA. Parfois, elles cachent leurs questions dans des conditions d'utilisation aussi difficiles à lire qu'à comprendre, pour ne pas dire indigestes et obscures. Cela dit, même si vous parveniez à comprendre les conditions d'utilisation, vous ne pourriez pas y faire grand-chose, car vous ne pouvez pas les modifier et vous aurez peut-être tout de même besoin d'utiliser le service en question (pour l'école ou le travail, par exemple).

- **Biais et discrimination :** Une IA ne peut apprendre que ce qu'on lui enseigne. Donc, si ce qu'elle apprend est erroné (par exemple, que la Lune est constituée de fromage) ou biaisé (par exemple, que les personnes portant une chemise bleue sont plus drôles que celles portant une chemise rouge), alors l'IA prendra des décisions en se fondant sur ces mauvaises informations, ce qui peut l'inciter à prendre de mauvaises décisions.
- **Responsabilité et transparence :** Comme plusieurs entreprises développent leur propre IA, elles ont tendance à ne rien divulguer pour éviter que les autres ne les copient. Cependant, cela signifie que personne ne sait comment leurs IA fonctionnent, ni si elles respectent la loi. Il est donc plus difficile pour tout le monde de prendre une décision éclairée concernant l'utilisation de la machine ou du service en question, car il est impossible d'obtenir les informations dont nous avons besoin pour prendre la bonne décision.
- **Normes du secteur :** Quelquefois, lorsque tout le monde agit de la même façon, il peut s'avérer difficile de prendre du recul sur ce que l'on fait et d'agir différemment, ou encore de prendre le temps de réfléchir aux conséquences de l'IA que l'on développe. Les normes servent à faire en sorte que tous suivent un certain nombre de règles communes à toutes les entreprises d'un même secteur.
- **Nécessité d'employer des humains :** Certaines IA doivent analyser de grandes quantités d'images pour apprendre à reconnaître certains objets, et ce sont des humains qui doivent trouver et trier toutes ces images. Bien souvent, ce type d'emploi est mal payé. Cela signifie que les personnes qui entraînent les IA font partie de groupes bien particuliers et ne représentent que des opinions limitées, et non pas l'intégralité de la population. Cela peut créer des IA biaisées et moins fiables.

- **Sécurité :** L'IA étant une technologie récente, les cybercriminel·le·s apprennent de nouvelles manières de s'en servir à des fins malhonnêtes. Comme les IA excellent dans les tâches répétitives, les cybercriminel·le·s peuvent aussi s'en servir pour envoyer de grandes quantités de messages à la fois ou pour recueillir des données qui peuvent être utilisées contre vous (par exemple, en dérobant votre identité ou en vous fournissant de fausses nouvelles afin que vous réalisiez l'action qui les intéresse).

Vers des IA plus dignes de confiance

Puisque l'IA est très récente, et que l'idée d'une IA « digne de confiance » l'est bien plus encore, il est extrêmement difficile de savoir quelle est la « bonne » manière de régler ces problèmes. En instaurant des principes fondamentaux plutôt qu'en mettant en place des règles strictes, nous pouvons essayer d'enseigner à tous et à toutes des valeurs qui contribueront au développement d'IA sûres et dignes de confiance.

Voici les deux principes fondamentaux à respecter pour qu'une IA soit digne de confiance :

1. AGENTIVITÉ (contrôle humain)

Toutes les IA doivent être conçues en prenant en compte la vie privée et le bien-être des êtres humains.

2. RESPONSABILITÉ (responsabilité des entreprises)

Tou·te·s les créateur·rice·s d'IA doivent endosser la responsabilité et les conséquences, quelles qu'elles soient, des actions de leur IA.

En plus de proposer ces deux principes, Mozilla a également fait certaines suggestions.

digital 2030
by digital moment

1. Modifier les normes du secteur pour se concentrer davantage sur les IA dignes de confiance

Nous devons nous assurer que les personnes qui développent les IA savent comment faire pour qu'elles soient dignes de confiance. Pour cela, voici quelques pistes à explorer :

- Créer des règles et des directives claires concernant les IA dignes de confiance.
- Enseigner aux personnes qui développent les IA comment rendre ces dernières dignes de confiance.
- S'assurer que des personnes de différents horizons sont impliquées dans le processus de conception, de développement et d'apprentissage des IA.
- Encourager les investisseur·euse·s et les très grosses entreprises à investir dans des produits et des entreprises reposant sur des IA dignes de confiance.

2. Encourager les utilisateur·rice·s à se tourner vers des IA dignes de confiance plutôt que d'autres IA

Au bout du compte, les produits sont conçus par les entreprises dans le but d'être utilisés. Il appartient donc également aux utilisateur·rice·s de dire aux entreprises que les IA dignes de confiance sont importantes à leurs yeux. Pour cela, voici quelques pistes à explorer :

- La protection de la vie privée doit être considérée comme un véritable pilier dans le développement des IA et comme le point de départ de la conception de tout nouveau produit.
- La mise en place de guides de produits ou de « scores de confiance » (à l'instar des étiquettes relatives à la santé sur les emballages de nourriture) permettra d'aider les utilisateur·rice·s à choisir les produits en fonction de leur degré de fiabilité.
- Les utilisateur·rice·s réclament plus de transparence sur le fonctionnement des IA, et les entreprises doivent inclure ces informations.

- Lorsque des entrepreneur·euse·s trouvent de nouvelles idées pour développer des IA dignes de confiance, les investisseur·euse·s devraient leur apporter un soutien financier.
- Les artistes, les journalistes et les éducateur·rice·s informent les utilisateur·rice·s sur le fonctionnement des IA et sur l'impact qu'elles peuvent avoir sur leur vie, et aident ces utilisateur·rice·s à trouver des idées pour améliorer ces technologies.
- Les citoyen·ne·s expriment leur mécontentement envers les entreprises à l'aide de pétitions et en demandant à leurs élu·e·s d'agir.
- Les groupes qui luttent pour la défense des droits de la personne doivent également savoir ce qu'est une IA digne de confiance.

3. Mettre en place des règles plus strictes pour faire en sorte que toutes les IA soient dignes de confiance

Dans la mesure où la technologie évolue à toute vitesse, les lois ne peuvent pas toujours suivre le rythme. Les gouvernements doivent à tout prix comprendre les enjeux liés aux IA dignes de confiance afin de pouvoir rédiger des lois pertinentes qui pourront être appliquées. Pour cela, voici quelques pistes à explorer :

- Les personnes travaillant pour le gouvernement doivent apprendre comment fonctionnent les IA.
- Les règles existantes relatives à la protection des personnes doivent être mises à jour de manière à inclure les IA dignes de confiance. Les gouvernements doivent également être plus sévères envers ceux et celles qui ne respectent pas ces règles.
- Les entreprises doivent faire preuve de transparence envers le gouvernement et les autorités de réglementation concernant le fonctionnement de leur IA.
- Les gouvernements doivent commencer à investir dans les IA dignes de confiance et à les utiliser.

digital 2030
by digital moment

Prochaines étapes

Certes, il y a beaucoup de travail à faire, mais il s'agit d'une tâche très importante qu'il faut entreprendre dès maintenant. Nous avons besoin de l'opinion d'un maximum de personnes de tous les horizons concernant les IA dignes de confiance afin de pouvoir avoir notre mot à dire sur la manière dont les IA sont développées. Tout le monde doit comprendre le fonctionnement des IA et les raisons pour lesquelles il est essentiel qu'elles soient dignes de confiance.

N'hésitez pas à partager votre opinion avec nous ou avec Mozilla. Il est temps de faire savoir au secteur de l'IA et aux autorités de réglementation à quel point les IA dignes de confiance sont importantes pour les enfants!