"On n'est pas obligé de chercher à humaniser un robot conversationnel"

Les performances croissantes des robots conversationnels imposent d'être vigilants face aux risques de dérive, explique Laurence Devillers, professeur en intelligence artificielle, et co-autrice d'un rapport du comité national d'éthique du numérique (CNPEN) sur les enjeux éthiques des chatbots, ou robots conversationnels, remis mardi au Premier ministre.

Temps de lecture : minute

---

10 novembre 2021

Pourquoi s'inquiéter des enjeux éthiques autour des robots conversationnels ? Pour l'instant, leurs performances restent encore souvent limitées...

Les chatbots qui utilisaient des modèles de langue un peu réduits, qui semblaient sans imagination, sont en train de changer. Ils sont capables d'avoir des dialogues beaucoup plus réalistes qui vont engager l'utilisateur de manière plus efficace. Les concepteurs de robots conversationnels cherchent désormais à créer des systèmes personnalisés, émotionnels, des sortes "d'amis virtuels".

Le risque est que les utilisateurs, tout particulièrement les personnes vulnérables - enfants, personnes âgées ...- projettent des capacités humaines sur ces machines, et sur le long terme voient leurs croyances et opinions influencées. La technologie permet par exemple de construire des "*deadbot*" : des machines qui refont parler une personne décédée, après avoir assimilé des montagnes de données sur elles, leurs

conversations, les contenus qu'elles ont postés sur les réseaux sociaux. Quels sont les enjeux de cette postérité numérique? Quels impacts sur la gestion d'un deuil?

Vous soulignez les nouvelles possibilités techniques de ces robots, qui désormais génèrent des phrases inédites.

Les nouvelles générations de chatbots sont de plus en plus performantes grâce à l'évolution des techniques d'apprentissage machine (machine learning). Les "transformeurs", ces modèle de traitement statistiques du langage naturel reposant sur des milliards de données comme le GPT-3 d'Open AI (Elon Musk) ou le LaMDA (Language Model for Dialogue Applications) de Google, permettent de créer des machines capables d'engager une conversation libre sur un nombre potentiellement infini de thèmes. Les machines deviennent capables de produire une parole inédite, qui surprend l'humain, et de manifester une sorte de créativité de langage. Comment fait-on pour qu'elles ne disent pas n'importe quoi? Qui est responsables des paroles qu'elles prononcent et de leurs effets?

À lire aussi

Nous ne sommes pas pour une régulation forte, mais nous recommandons de pousser la recherche scientifique sur le fonctionnement de ces modèles, et de s'engager dans un long chemin de réflexion et de mise en oeuvre de normes ou de bonnes pratiques, avec des comités d'audit notamment. La technologie est extrêmement positive si on la maîtrise, mais il va nous falloir de l'intelligence collective pour comprendre comment l'amadouer et comment ne pas se laisser guider par les Gafam...

Pouvez-vous donner des exemples de recommandation du rapport ?

Nous demandons par exemple de réduire la projection de qualités morales sur un agent conversationnel. Quand on le conçoit, on n'est pas obligé de chercher à l'humaniser pour que l'utilisateur projette dessus des qualités humaines. Il faut que la machine dise qu'elle est une machine, et qu'elle communique sur sa finalité. Nous nous intéressons aussi à la question des biais qui peuvent être perpétués par ces agents conversationnels, et notamment la question des biais de genre.

Si l'on donne systématiquement à des agents serviles des noms et des voix féminines, qu'est ce que cela veut dire des représentations des femmes dans la société ? Il faut s'inquiéter du pouvoir discret de ces choix qui sont faits pour nous... La mission recommande que le choix des noms et des voix (masculin ou féminin) du robot soit effectué de "façon équitable", et que l'utilisateur puisse "modifier ce choix" comme il le souhaite.

À lire aussi
Intelligence artificielle : 2,2 milliards d'euros pour la formation et l'innovation

---

Article écrit par Maddyness avec AFP