

> Il peut s'agir de *nudges* malveillants, que Richard Thaler appelle *sludges* (littéralement « boues », en anglais). L'un des défis consiste à évaluer la menace qu'ils représentent. Ainsi, le projet *Bad Nudge-Bad Robot* que je dirige au CNRS à l'université Paris-Saclay vise à mettre en évidence le danger des techniques de *nudging* pour les personnes vulnérables telles que les enfants ou les personnes âgées.

Ce projet entre dans le cadre de la chaire d'intelligence artificielle HUMAINE (Human-Machine Affective Interaction & Ethics, <https://humaaine-chaireia.fr/>) au CNRS, qui repose sur une forte collaboration interdisciplinaire entre l'informatique affective et la robotique émotionnelle, l'économie comportementale, la linguistique et le traitement du langage naturel. Ce groupe de recherche a pour objectif d'étudier le comportement des agents conversationnels, ou « chatbots », dans leurs interactions avec des interlocuteurs humains, afin d'évaluer l'influence potentielle des *nudges* sur les humains.

Par exemple, en 2019, dans une expérience réalisée dans une école primaire, nous avons montré que les *nudges* des machines (robots, chatbots) étaient plus efficaces que les *nudges* d'humains dans un jeu du dictateur. Dans cette expérience, la machine ou l'humain donne le même nombre de billes à chaque enfant. La première question posée à l'enfant est combien il en garde pour lui et combien il en donne aux autres enfants. La deuxième question est un *nudge* pour le faire changer d'avis, par exemple en lui disant que les autres en donnent plus (ce qui active le biais de conformisme). Plus les enfants étaient jeunes, plus ils étaient facilement manipulés par les machines. Cet exemple met en lumière le fort potentiel d'incitation associé à des bots capables de déployer un *nudge*. À n'en pas douter, de tels bots accroîtront les moyens de manipulation sur le web.

L'intelligence artificielle fournit ainsi des armes diverses, les unes permettant de traquer les propos haineux et repérer les trolls et les bots, les autres de propager des fausses informations et de manipuler les internautes,



Le *nudge* le plus célèbre est celui des urinoirs dont le bas est orné d'une image de mouche, qui incite les usagers à viser cette cible et ainsi à maintenir la propreté des toilettes. Ce concept d'incitation/manipulation douce se répand aussi dans le monde numérique.

BIBLIOGRAPHIE

H. Ali Mehenni, L. Devillers et al., **Nudges with conversational agents and social robots : a first experiment with children at a primary school**, *Conversational Dialogue Systems for the Next Decade*, Springer, pp. 257-270, 2021.

L. Devillers, **Les robots émotionnels**, L'Observatoire, 2020.

C. Schneider et al., **Digital nudging : guiding online user choices through interface design**, *Communications of the ACM*, vol. 61(7), pp. 67-73, 2018.

D. Kahneman, **Thinking, Fast and Slow**, Farrar, Straus and Giroux, 2011.

R. H. Thaler et C. R. Sunstein, **Nudge : Improving Decisions About Health, Wealth and Happiness**, Yale University Press, 2008.

notamment grâce aux *nudges*. Ceux-ci sont aussi susceptibles d'être utilisés de façon positive. Imaginons par exemple qu'à la fin de chaque mois on vous attribue des points en fonction de la qualité de ce que vous avez tweeté, et que vous puissiez comparer avec vos amis ou vos collègues. Cela constituerait un *nudge* qui vous inciterait à moins relayer des informations non vérifiées et peut-être à adopter des pratiques plus respectueuses des autres.

L'utilisation de l'intelligence artificielle pour combattre la désinformation et la manipulation sur internet n'en est certainement qu'à ses débuts. Un autre aspect de cette lutte que l'on n'a pas évoqué ici porte sur l'équité et la transparence des systèmes d'intelligence artificielle, questions éthiques qui suscitent depuis quelques années un fort intérêt de la communauté scientifique.

UN JEU PERMANENT DU CHAT ET DE LA SOURIS

Lutter contre la désinformation à l'aide des techniques de l'intelligence artificielle est un défi permanent, un jeu du chat et de la souris où il faudra sans cesse améliorer les algorithmes pour garder une longueur d'avance. Mais, à l'évidence, se doter de tels outils pour traquer les fausses informations, les propos toxiques ou les bots malveillants ne suffit pas.

Pour protéger les citoyens et les démocraties, il est en effet nécessaire de réguler la jungle que constitue le web avec des lois et des garde-fous éthiques. Le silence de la grande majorité des ingénieurs des plateformes œuvrant sur la Toile, telles que Google, Facebook ou Twitter, est assourdissant, alors même que les techniques qu'ils mettent en œuvre façonnent la société de demain. Par ailleurs, il n'est pas souhaitable que ces plateformes aient un véritable pouvoir de censure.

La régulation des plateformes du web est assurément une question difficile, mais aussi cruciale pour l'avenir de nos sociétés. Elle suscite aujourd'hui de nombreuses réflexions. Notamment, le Comité national pilote d'éthique du numérique a publié en juillet 2020 un document portant sur les enjeux dans la lutte contre la désinformation et la mésinformation et qui identifie les tensions éthiques relatives à la mise en œuvre, par les plateformes, d'outils de modération de contenus et aux mécanismes de lutte contre la désinformation. Il souligne aussi certains questionnements éthiques relatifs à l'autorité acquise par ces plateformes et aux rapports qu'elles entretiennent avec l'État et la justice. En résumé, les innovations de l'intelligence artificielle ne sont pas tout : il est aussi nécessaire de redéfinir la responsabilité des plateformes aux niveaux national et européen, tout en protégeant la liberté d'expression. ■

1 AN

3 FORMULES AU CHOIX

BULLETIN D'ABONNEMENT

PAG21STD



OUI, je m'abonne
pour 1 an à POUR LA **SCIENCE**

☐ M. ☐ Mme

Nom: Prénom:

Adresse:

Code postal | | | | | Ville:

Téléphone

Email: (indispensable pour la formule intégrale)

.....@

J'accepte de recevoir les offres de Pour la Science

☐ OUI ☐ NON

☐ **Par chèque** à l'ordre de *Pour la Science* en nous retournant ce bulletin d'abonnement complété



**Pour un paiement par carte bancaire,
rendez-vous sur boutique.groupepourlascience.fr**

* Réduction par rapport au prix de vente en kiosque et l'accès aux archives numériques.
Durée d'abonnement: 1 an. Délai de livraison: dans le mois suivant l'enregistrement de votre règlement. Offre valable jusqu'au 31/03/2022 en France métropolitaine uniquement. Pour un abonnement à l'étranger, merci de consulter notre site boutique.groupepousscience.fr. Photos non contractuelles. Vous pouvez acheter séparément les numéros de *Pour la Science* pour € 90 et les hors-séries pour 790 €.

Les informations que nous collectons dans ce bulletin d'abonnement nous aident à personnaliser et à améliorer les services que nous vous proposons. Nous les utiliserons pour gérer votre accès à l'intégralité de nos services, traiter vos commandes et paiements, et vous faire part notamment par newsletters de nos offres commerciales moyennant le respect de vos choix en la matière. Le responsable du traitement est la société *Pour la Science*. Vos données personnelles ne seront pas conservées au-delà de la durée nécessaire à la finalité de leur traitement. *Pour la Science* ne commercialise ni ne loue vos données à caractère personnel à des tiers. Les données collectées sont exclusivement destinées à *Pour la Science*. Nous vous invitons à prendre connaissance de notre charte de protection des données personnelles à l'adresse suivante : <https://rebrand.ly/charte-donnees-pls>. Conformément à la réglementation applicable (et notamment au Règlement 2016/679/UE dit « RGPD ») vous disposez des droits d'accès, de rectification, d'opposition, d'effacement, à la portabilité et à la limitation de vos données personnelles. Pour exercer ces droits (ou nous poser toute question concernant le traitement de vos données personnelles), vous pouvez nous contacter par courriel à l'adresse protection-donnees@pourlascience.fr.