

Propos haineux, théories du complot et fausses nouvelles déferlent sur la Toile grâce, notamment, au renfort de robots logiciels. Contre ces fléaux issus de l'IA, il est heureusement un recours à fort potentiel : l'IA elle-même.

42

Communication

Les deux visages de l'intelligence artificielle

Laurence Devillers



En bref

> Sur internet, des outils d'intelligence artificielle sont utilisés de façon malveillante pour propager des informations fausses ou pour influencer les internautes.	> Ces outils exploitent nos biais cognitifs et nos réactions émotionnelles. Les <i>nudges</i> , qui consistent à inciter de façon imperceptible, renforcent leurs capacités de nuisance.	> Mais les <i>nudges</i> pourraient aussi aider les internautes à mieux résister à la désinformation.	> Des outils logiciels sont aussi développés pour traquer les contenus toxiques et repérer les robots malveillants œuvrant sur la Toile.
---	--	---	--

44

Avec la pandémie de Covid-19, nous avons vécu une période d'incertitude et d'anxiété qui a amplifié, en particulier sur internet, la propagation d'informations incomplètes ou fausses. Il y a eu d'abord la désinformation, c'est-à-dire la diffusion massive et délibérée d'informations visant à tromper ou à manipuler via les réseaux sociaux, les sites web, les forums, les messageries ou les *trolls* (internautes qui créent volontairement des polémiques). À laquelle s'est ajoutée la mésinformation, définie comme la propagation d'informations non vérifiées, incomplètes ou incertaines par des internautes qui n'ont bien souvent pas conscience des conséquences délétères de leurs actions.

Ainsi, compte tenu de l'arrivée massive d'informations provenant de sources multiples, contrôler la crédibilité et la fiabilité des données circulant en ligne devient une tâche urgente. D'autant que les techniques d'intelligence artificielle se développent et que les manipulateurs en tous genres n'hésitent pas à s'en servir.

Citons par exemple les techniques du *deep-fake*, qui permettent de truquer des vidéos en leur superposant d'autres bandes-son ou images. Par ailleurs, sur les réseaux sociaux, l'utilisation des *bots*, ou robots logiciels, devient de plus en plus fréquente. Il existe ainsi des comptes automatisés, gérés par un algorithme plutôt que par une personne réelle, qui peuvent influencer le comportement des internautes, émettre des avis et renforcer des opinions. Ces faux profils alimentent des réseaux complexes en administrant des groupes, en augmentant le nombre de leurs membres et en « likant » les messages sur les pages. On estime

que jusqu'à 15% des comptes Twitter actifs sont de faux comptes pilotés par des *bots*.

La menace va grandissant. Demain, ces *bots* seront capables d'interagir personnellement avec les internautes en détectant leurs émotions et leurs intentions. En exploitant nos biais cognitifs (voir le schéma page ci-contre), ils pourront nous manipuler de façon très efficace et nous inciter par exemple à adhérer à un parti politique ou à acheter des produits censés nous apporter du bien-être. Les défis portent donc, d'une part, sur la vérification de la crédibilité des informations émises ou relayées par les internautes et par les *bots*, et, d'autre part, sur la vérification des « intentions » de ces machines.

Les développements dans le domaine de l'intelligence artificielle devraient conduire à des outils de plus en plus efficaces pour traquer sur internet les propos haineux, offensants ou menaçants (voir l'encadré pages 48 et 49). Comme le préconisent la Commission européenne et le rapport 2019 de la mission française « Régulation des réseaux sociaux – Expérimentation Facebook », il est nécessaire de mener une réflexion d'ampleur en vue de constituer des bases de données communes pour améliorer les outils numériques de lutte contre la désinformation.

Des outils sont également disponibles pour détecter les faux profils et cartographier la diffusion de fausses nouvelles. Les ingénieurs de Facebook ont par exemple mis au point un outil d'apprentissage automatique qui a déjà permis de supprimer des milliards de faux comptes en 2020. Mais dès qu'un faux profil disparaît, un autre surgit. Pour mieux enrayer les dérives