

> d'informations non vérifiées, incomplètes ou incertaines par des internautes qui n'ont bien souvent pas conscience des conséquences délé-tères de leurs actions.

Ainsi, compte tenu de l'arrivée massive d'in-formations 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 arti-ficielle (IA) se développent et que les manipula-teurs 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 de «bots», des robots logiciels, devient de plus en plus fréquente. Il existe ainsi des comptes automatisés, gérés par un algo-rithme plutôt que par une personne réelle, qui peuvent influencer le comportement des inter-nautes, émettre des avis et renforcer des opi-nions. 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 la figure page ci-contre), ils pour-ront 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 inter-nautes 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 tra-quer sur internet les propos haineux, offen-sants ou menaçants (voir l'encadré pages 30 et 31). Comme le préconisent la Commission européenne et le rapport (2019) de la mission « Régulation des réseaux sociaux – Expérimentation Facebook » commanditée par le gouvernement français, il est nécessaire de mener une réflexion d'ampleur sur la consti-tution de bases de données communes pour améliorer les outils numériques de lutte contre la désinformation et la mésinformation.

Des outils sont également disponibles pour détecter les faux profils et cartographier la dif-fusion 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 est à terre, un autre surgit. Pour mieux enrayer

les dérives informationnelles sur les réseaux sociaux, il est donc nécessaire de comprendre les ressorts de la désinformation sur ces pla-teformes (voir l'article pages 34 à 41), mais aussi les manipulations possibles à partir de la détection de nos comportements.

De nouveaux terrains de recherche en intel-ligence artificielle s'ouvrent donc : identifier les fausses informations, les rumeurs et les préjugés, résoudre les conflits entre des sources de don-nées hétérogènes, mais aussi détecter les trolls ou les bots qui cherchent à nuire en propageant des rumeurs ou en manipulant les opinions.

LES BIAIS COGNITIFS, LES ÉMOTIONS ET LE « NUDGE »

Sur les réseaux sociaux en ligne, la diffusion «virale», c'est-à-dire de façon massive et rapide, de certaines informations tient en grande partie à la capacité d'attention limitée des internautes, à leur appétence pour les réac-tions émotionnelles et aux biais cognitifs qui piègent nos raisonnements. Les plateformes numériques telles que Facebook, Twitter ou Instagram ont ainsi pour effet d'augmenter considérablement la visibilité des usagers malintentionnés ou de ceux qui sont prompts à exprimer librement des opinions. Certes, le principe de liberté d'expression est revendiqué par nos sociétés occidentales, mais les opinions exprimées sur les réseaux sociaux sont souvent des jugements rapides, des idées reçues, des évidences, des préjugés, des lieux communs... Des contenus dont la qualité ne se juge pas au nombre de clics qu'ils suscitent !

Les biais cognitifs jouent un rôle central dans la propagation de tels contenus et de la désinformation. Et les futurs bots malveillants risquent de les exploiter de façon encore plus puissante et plus explicite. Dans les années 1970, les biais cognitifs ont été étudiés dans le cadre de l'économie comportementale, afin de comprendre nos décisions et compor-tements irrationnels. Les psychologues en ont recensé plusieurs dizaines, parfois classés en quatre catégories : les biais qui découlent d'un trop-plein d'informations, ceux dus à un



En 2020, Facebook a supprimé, grâce à un outil d'IA, des milliards de faux comptes

