



l'engagement de la génération Greta

tables rondes

— samedi 22 mai

cité

sciences
et industrie

© Gettyimages

La paralysie des sociétés pendant la pandémie favorise la réflexion sur l'évolution de nos modes de vie. Face à la menace d'une récession économique et aux périls climatiques, des voix s'élèvent pour tenter des transformations radicales vers une société résiliente. Quand allons-nous enfin relever les défis environnementaux pour l'avenir ? Comment redéfinir le travail afin qu'il preserve le monde et les individus ? La situation inédite d'aujourd'hui engendrera-t-elle un monde meilleur ? Comme la jeune militante Greta Thunberg, symbole de la lutte écologiste, les jeunes générations entendent bien dire leur mot sur ces questions !

Des élèves de première et terminale du lycée de La Mare Carré de Moissy-Cramayel (77) présentent leurs réflexions et échangent avec des experts, sociologues et philosophes.

Au programme, deux tables rondes :

- Relever les défis environnementaux
 - Quel travail à l'ère de l'anthropocène ?
- ainsi que des témoignages, des clips vidéos, des podcasts...

**Journée proposée en hommage
au philosophe Bernard Stiegler (1952- 2020)**

Accès gratuit sur réservation obligatoire
ou en ligne suivant la situation sanitaire

AVEC LE SOUTIEN DE **POUR LA SCIENCE**

Informations cite-sciences.fr

L'ESSENTIEL

> 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 « nudges », qui consistent à inciter de façon imperceptible, renforceront leurs capacités de nuisance.

> Mais les « nudges » 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.

L'AUTRICE



LAURENCE DEVILLERS
professeure d'intelligence artificielle à Sorbonne Université, à Paris, et chercheuse au Laboratoire interdisciplinaire des sciences du numérique (LISN, CNRS)

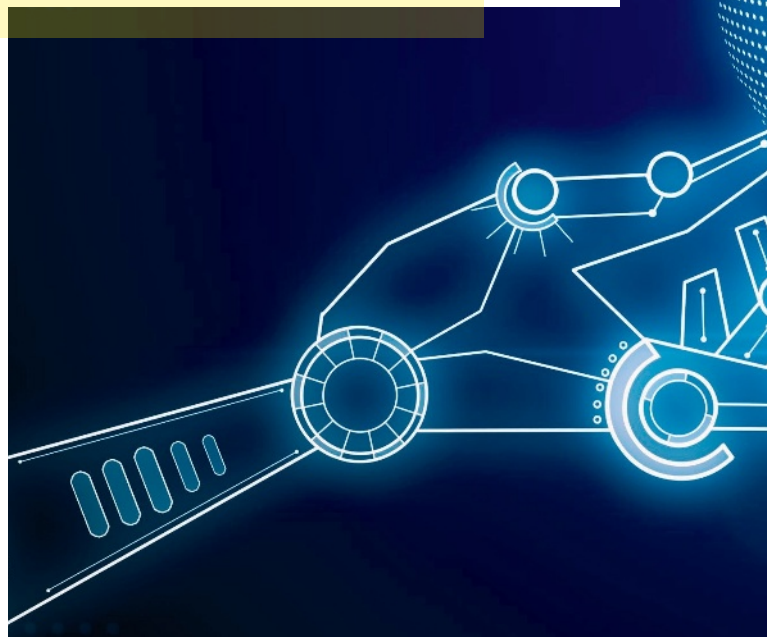
Désinformation Les armes de l'intelligence artificielle

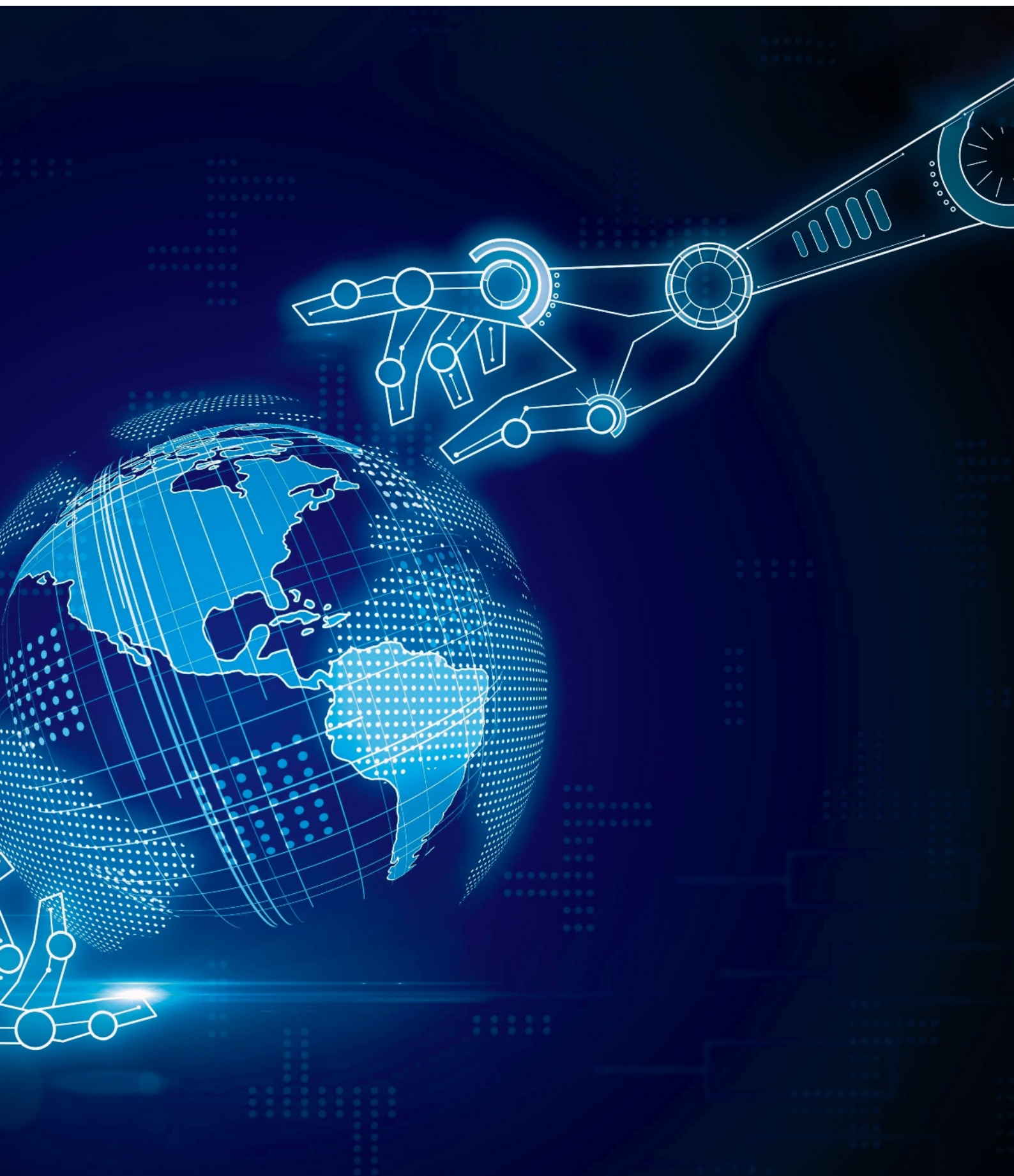
Propos haineux, théories du complot, fausses nouvelles ou manipulations déferlent sur la Toile grâce, notamment, au renfort de robots logiciels. Parallèlement, les chercheurs élaborent d'autres outils d'intelligence artificielle pour lutter contre ces fléaux.

Avec la pandémie de Covid-19, nous vivons une période d'incertitude et d'anxiété qui amplifie la propagation d'informations incomplètes ou fausses. Sur l'origine du virus, sur les traitements, sur les vaccins, sur les mesures de confinement comme sur de nombreux autres sujets, fleurissent informations fausses, théories du complot, propos haineux, jugements à l'emporte-pièce...

La désinformation sur internet est une partie importante de ce phénomène. Il s'agit de la diffusion massive et délibérée d'informations *via* les réseaux sociaux, les sites web, les forums, les messageries ou les trolls (internautes qui créent volontairement des polémiques) dans le but de tromper ou de manipuler. Et il ne faut pas oublier la mésinformation, c'est-à-dire la propagation >

© Shutterstock.com/Pixels Hunter





> 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 (IA) se développent et que les manipulateurs en tous genres n'hésitent pas à s'en servir.

Citons par exemple les techniques du *deepfake*, 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 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 la figure 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 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 constitution 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 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 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 plateformes (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 intelligence artificielle s'ouvrent donc : identifier les fausses informations, les rumeurs et les préjugés, résoudre les conflits entre des sources de données 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 réactions 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 comportements 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

