



GETTY IMAGES

Une machine à textes... et à fake news

Elle serait capable de manier la plume comme Proust ou Camus. Elle, c'est une intelligence artificielle répondant au nom de GPT-2. Née au sein du centre de recherche privé OpenAI, ses capacités d'écriture seraient si phénoménales, autrement dit ses faux textes si convaincants, qu'elle en arrive à faire peur. Spécialisée en intelligence artificielle et financée par des milliardaires américains comme Elon Musk, PDG de SpaceX et de Tesla, ou encore Peter Thiel, cofondateur de PayPal, l'organisation à but non lucratif OpenAI a annoncé repousser la publication intégrale du programme GPT-2 pour laisser le temps à la communauté humaine d'en mesurer les possibles conséquences éthiques et sociétales... Prédire le mot suivant en se basant sur tous les mots précédents d'un texte, voilà le but de GPT-2. Il suffit donc de lui donner une phrase pour qu'elle en rédige automatiquement la suite. L'intelligence artificielle s'adapte au style et au contenu des informations qu'on lui fournit, ce qui permet, selon OpenAI, de « générer une suite réaliste et cohérente sur le sujet choisi ». Le programme peut ainsi rédiger une recette de cuisine, une chronique humoristique, un pamphlet politique, un roman à l'eau de rose ou un article de presse.

Le *Guardian* a pu tester la machine. La première phrase d'un article sur les conséquences économiques du Brexit lui a été soumise. La suite du

texte, écrit par GPT-2, est impressionnante par son contenu chiffré et crédible ainsi que par le style employé. « Une IA écrit exactement comme moi. Préparez-vous à l'apocalypse des robots », a mis en garde une chroniqueuse du journal britannique.

GPT-2 est un programme informatique composé de 1,5 milliard de paramètres s'entraînant à partir de 40 gigabits de textes, issus de huit millions de pages Web. Il n'est pas tout-puissant pour autant. « Il lui faut plusieurs essais pour produire un bon texte. Quand on lui donne des sujets qui sont énormément représentés dans les données (le Brexit, Miley Cyrus, *Le Seigneur des anneaux*, etc.), il est capable de produire un texte correct dans 50 % des cas, précisent les chercheurs d'OpenAI sur leur blog. Mais l'inverse est aussi vrai : sur des contenus très techniques ou ésotériques, les résultats sont mauvais. »

Néanmoins, craignant de voir son modèle de génération automatisée de texte être utilisé à mauvais escient – par exemple pour créer des fake news à tour de bras – l'organisation a indiqué qu'elle ne proposerait pas son moteur en open source, contrairement à son habitude. Ses chercheurs redoutent également que GPT-2 soit utilisée pour usurper l'identité d'internautes. « Le grand public doit être davantage méfiant face aux textes publiés en ligne, comme les phénomènes de deepfakes doivent les rendre plus méfiants concernant les images », insistent les chercheurs, en référence aux vidéos manipulées qui reproduisent la gestuelle et le ton de voix d'une personne, en lui faisant tenir un faux discours. **V** LAETITIA THEUNIS