



Le Forum économique mondial propose un plan mondial pour contrôler les contenus en ligne grâce à l'intelligence artificielle

Le Forum économique mondial a publié ce mois-ci un article appelant à la mise en place d'un système de censure en ligne alimenté par une combinaison d'intelligence artificielle et humaine qui, selon un critique, « mondialiserait » la « recherche de la mauvaise pensée ».

Par [Children's Health Defense](#)

Mondialisation.ca, 24 août 2022

[The Defender](#) 19 août 2022

Mettant en garde contre un « monde obscur de préjugés en ligne » auquel il faut s'attaquer, le Forum économique mondial (WEF) [a publié ce mois-ci un article](#) appelant à une « solution » aux « abus en ligne » qui serait alimentée par l'intelligence artificielle (IA) et l'intelligence humaine.

La proposition [prévoit](#) la mise en place d'un système, basé sur l'IA, qui automatiserait la censure des « fausses informations » et des « discours de haine » et s'efforcerait de [venir à bout de](#) la propagation des « abus sur les enfants, de l'extrémisme, de la désinformation, des discours de haine et de la fraude » en ligne.

Selon l'auteur de l'article, [Inbal Goldberger](#), les « équipes de confiance et de sécurité » humaines ne sont pas entièrement capables, à elles seules, de contrôler ces contenus en ligne.

Mme Goldberger est vice-présidente d'[ActiveFence Trust & Safety](#), une société technologique basée à New York et à Tel-Aviv qui [affirme](#) « collecter automatiquement des données provenant de millions de sources et appliquer l'intelligence artificielle contextuelle pour renforcer les opérations de confiance et de sécurité de toute taille ».

Au lieu de s'appuyer uniquement sur des équipes de modération humaines, Mme Goldberger [propose](#) un système basé sur « l'intelligence hors plateforme, multilingue et alimentée par l'homme » – en d'autres termes, des données fournies par des sources humaines « expertes » qui créeraient ensuite des « ensembles d'apprentissage » qui entraîneraient l'IA à reconnaître des contenus prétendument préjudiciables ou dangereux.

Cette « intelligence hors plateforme » – qui relève plus de l'apprentissage automatique que de l'intelligence artificielle à proprement parler, [selon Didi Rankovic](#) de ReclaimTheNet.org – serait collectée à partir de « millions de sources », puis regroupée et fusionnée avant d'être [utilisée](#) pour des « décisions de retrait de contenu » de la part des « plateformes Internet ».

Selon Mme Goldberger, le [système](#) compléterait « une détection automatisée plus

intelligente par une expertise humaine » et permettrait de créer « une IA avec une intelligence humaine intégrée ».

Cela permettrait de se protéger contre « des acteurs de plus en plus avancés qui utilisent les plates-formes à mauvais escient de manière unique ».

« Un modérateur humain expert en suprématie blanche européenne ne sera pas nécessairement capable de reconnaître un contenu nuisible en Inde ou des récits de désinformation au Kenya », [a expliqué Mme Goldberger](#).

Toutefois, « en combinant de manière unique la puissance d'une technologie innovante, la collecte de renseignements hors plateforme et les prouesses d'experts en la matière qui comprennent comment les acteurs de la menace opèrent, la détection à grande échelle des abus en ligne peut atteindre une précision quasi parfaite », car ces ensembles d'apprentissage sont « intégrés » à l'IA au fil du temps, a déclaré Mme Goldberger.

Cela permettrait ensuite aux « équipes chargées de la confiance et de la sécurité » d'« arrêter les menaces qui montent en ligne avant qu'elles n'atteignent les utilisateurs », a-t-elle ajouté.

Dans son [analyse](#) de ce à quoi la proposition de Mme Goldberger pourrait ressembler en pratique, le blogueur Igor Chudov a expliqué comment la police du contenu sur les médias sociaux se fait aujourd'hui plate-forme par plate-forme.

Par exemple, les modérateurs de contenu de Twitter ne regardent que le contenu publié sur cette plateforme particulière, mais pas le contenu d'un utilisateur publié en dehors de Twitter.

Selon M. Chudov, c'est la raison pour laquelle le WEF semble soutenir une proposition visant à « aller au-delà des principales plateformes Internet, afin de collecter des renseignements sur les personnes et les idées partout ailleurs ».

« Une telle approche, écrit M. Chudov, leur permettrait de mieux savoir quelle personne ou idée censurer – sur toutes les grandes plateformes à la fois. »

Les « renseignements » recueillis par le système auprès de ses « millions de sources » permettraient, selon M. Chudov, de « détecter les pensées qu'ils n'aiment pas », ce qui entraînerait des « décisions de suppression de contenu transmises à des sites comme Twitter, [Facebook](#), etc... un changement majeur par rapport au statu quo où chaque plateforme décide de ce qu'il faut faire en fonction des messages postés sur cette plateforme spécifique uniquement ».

De cette manière, « la recherche de la mauvaise pensée se mondialise », conclut M. Chudov.

En réponse à la proposition du WEF, ReclaimTheNet.org a [souligné](#) que « l'on peut commencer à discerner l'argument ici ... comme faisant simplement pression sur les réseaux sociaux pour qu'ils commencent à évoluer vers une censure préventive ».

[Selon](#) M. Chudov, le WEF promeut la proposition parce qu'il « commence à s'inquiéter », car « les opinions non approuvées sont de plus en plus populaires, et les censeurs en ligne ne peuvent pas suivre les millions de personnes qui deviennent plus conscientes et plus

vocales ».

[Selon le Daily Caller](#), « le document du WEF ne précise pas comment les membres de l'équipe de formation de l'IA seraient décidés, comment ils seraient tenus responsables ou si les pays pourraient exercer des contrôles sur l'IA ».

Dans un [avertissement](#) accompagnant l'article de Mme Goldberger, le WEF a rassuré le public en précisant que le contenu exprimé dans l'article « est l'opinion de l'auteur et non celle du Forum économique mondial », ajoutant que « cet article a été partagé sur des sites web qui ont l'habitude de déformer le contenu et de diffuser des informations erronées ».

Cependant, le WEF semble être ouvert à des propositions comme celle de Mme Goldberger. Par exemple, un [article](#) publié en mai 2022 sur le site du WEF propose le « [conseil de surveillance](#) » de Facebook comme exemple de « modèle de gouvernance du monde réel » pouvant être appliqué à la gouvernance dans le [métavers](#).

Et, comme l'[a fait remarquer](#) M. Chudov, « la modération du contenu par l'IA s'intègre directement dans le [système de notation sociale de](#) l'IA ».

[Download for Free: Robert F. Kennedy's New Book — 'A Letter to Liberals'](#)

L'ONU, soutenue par la Fondation Gates, vise également à « briser la chaîne de la désinformation ».

Le Forum économique mondial n'est pas la seule entité à réclamer une surveillance plus stricte des contenus en ligne et de la « désinformation ».

Par exemple, l'UNESCO a récemment [annoncé](#) un partenariat avec Twitter, la Commission européenne et le Congrès juif mondial, qui a débouché sur le lancement de la campagne #ThinkBeforeSharing, destinée à « stopper la propagation des théories du complot ».

Selon l'UNESCO :

« La pandémie de [COVID-19](#) a suscité une augmentation inquiétante de la désinformation et des théories du complot.

« Les théories du complot peuvent être dangereuses : elles ciblent souvent des groupes vulnérables et les discriminent, ignorent les preuves scientifiques et polarisent la société avec de graves conséquences. Il faut que cela cesse. »

La [directrice générale](#) de l'UNESCO, Audrey Azoulay, [a déclaré](#) :

« Les théories du complot causent un réel préjudice aux personnes, à leur santé, mais aussi à leur sécurité physique. Elles amplifient et légitiment les idées fausses sur la pandémie, et renforcent les stéréotypes qui peuvent alimenter la violence et les idéologies extrémistes violentes. »

L'UNESCO a déclaré que le partenariat avec Twitter [permet d'informer les gens](#) que les événements qui se produisent dans le monde ne sont pas « secrètement manipulés dans les coulisses par des forces puissantes aux intentions négatives ».

L'UNESCO a publié des conseils sur ce qu'il faut faire si l'on rencontre un « théoricien de la

conspiration » en ligne : Il faut « réagir » immédiatement en publiant dans les commentaires un lien pertinent vers un « site web de vérification des faits ».

L'UNESCO fournit également des conseils au public au cas où quelqu'un rencontrerait un « théoricien du complot » en chair et en os. Dans ce cas, l'individu doit éviter d'argumenter, car « tout argument peut être pris comme une preuve que vous faites partie de la conspiration et renforcer cette croyance ».

La campagne #ThinkBeforeSharing propose une multitude d'infographies et de documents d'accompagnement destinés à expliquer ce que sont les « théories du complot », comment les identifier, comment en rendre compte et comment y réagir de manière plus générale.

Selon ces documents, les théories du complot ont six choses en commun, notamment :

- Un « prétendu complot secret ».
- Un « groupe de conspirateurs ».
- « Des preuves qui semblent soutenir la théorie de la conspiration ».
- Des suggestions qui affirment « faussement » que « rien n'arrive par accident et qu'il n'y a pas de coïncidences », et que « rien n'est comme il apparaît et tout est lié ».
- Ils divisent le monde en « bon ou mauvais ».
- Ils désignent des personnes et des groupes comme boucs émissaires.

L'UNESCO [ne rejette pas entièrement](#) l'existence des « théories du complot », mais admet que « de véritables conspirations, grandes et petites, existent ».

Toutefois, selon l'organisation, ces « conspirations » sont « plus souvent centrées sur des événements uniques et autonomes, ou sur un individu, comme un assassinat ou un coup d'État » et ne sont « réelles » que si elles sont « mises au jour par les médias ».

Outre le WEF et l'UNESCO, le Conseil des droits de l'homme des Nations unies (ONU) a [adopté](#) plus tôt cette année « un plan d'action pour lutter contre la désinformation ».

Le « plan d'action », [parrainé](#) par les États-Unis, le Royaume-Uni, l'Ukraine, le Japon, la Lettonie, la Lituanie et la Pologne, souligne « le rôle primordial des gouvernements dans la lutte contre les faux récits », tout en exprimant des [inquiétudes](#) :

« L'impact négatif croissant et de grande portée sur la jouissance et la réalisation des droits de l'homme de la création et de la diffusion délibérées d'informations fausses ou manipulées destinées à tromper et à induire en erreur des publics, soit pour causer un préjudice, soit pour un gain personnel, politique ou financier. »

Même les pays qui n'ont pas officiellement approuvé le plan du Conseil des droits de l'homme ont exprimé leurs [préoccupations](#) concernant la « désinformation » en ligne.

Par exemple, la Chine a identifié cette « désinformation » comme « un ennemi commun de la communauté internationale ».

Une initiative antérieure des Nations unies, en partenariat avec le WEF, a permis de « recruter 110 000 volontaires de l'information » qui, [selon les termes de](#) la directrice de la communication mondiale des Nations unies, Melissa Fleming, agiraient en tant que « premiers répondants numériques » à la « désinformation en ligne ».

L'initiative #PledgeToPause de l'ONU, bien qu'elle ait récemment circulé comme une nouveauté sur les médias sociaux, a été [annoncée](#) en novembre 2020, et a été décrite par l'ONU comme « la première campagne mondiale de changement de comportement en matière de désinformation ».

La campagne fait partie d'une initiative plus large de l'ONU, « [Verified](#) », qui vise à recruter des participants pour diffuser du « contenu vérifié optimisé pour le partage social », émanant directement du service de communication de l'ONU.

M. Fleming a [déclaré](#) à l'époque que l'ONU « travaillait également avec les plateformes de médias sociaux pour recommander des changements » afin de « contribuer à briser la chaîne de désinformation ».

La campagne « [Verified](#) » et la campagne [#PledgeToPause](#) semblent toujours actives au moment où nous écrivons ces lignes.

L'initiative « Verified » est gérée conjointement avec [Purpose](#), un groupe militant qui a [collaboré](#) avec la [Fondation Bill & Melinda Gates](#), la [Fondation Rockefeller](#), Bloomberg Philanthropies, l'Organisation mondiale de la santé, l'Initiative Chan Zuckerberg, Google et Starbucks.

Depuis 2019, l'ONU a conclu un [partenariat stratégique](#) avec le WEF basé sur six « domaines d'intérêt », dont l'un est la « coopération numérique ».

La source originale de cet article est [The Defender](#)

Copyright © [Children's Health Defense](#), [The Defender](#), 2022

Articles Par : [Children's Health Defense](#)

Avis de non-responsabilité : Les opinions exprimées dans cet article n'engagent que le ou les auteurs. Le Centre de recherche sur la mondialisation se dégage de toute responsabilité concernant le contenu de cet article et ne sera pas tenu responsable pour des erreurs ou informations incorrectes ou inexactes.

Le Centre de recherche sur la mondialisation (CRM) accorde la permission de reproduire la version intégrale ou des extraits d'articles du site [Mondialisation.ca](#) sur des sites de médias alternatifs. La source de l'article, l'adresse url ainsi qu'un hyperlien vers l'article original du CRM doivent être indiqués. Une note de droit d'auteur (copyright) doit également être indiquée.

Pour publier des articles de [Mondialisation.ca](#) en format papier ou autre, y compris les sites Internet commerciaux, contactez: media@globalresearch.ca

[Mondialisation.ca](#) contient du matériel protégé par le droit d'auteur, dont le détenteur n'a pas toujours autorisé l'utilisation. Nous mettons ce matériel à la disposition de nos lecteurs en vertu du principe "d'utilisation équitable", dans le but d'améliorer la compréhension des enjeux politiques, économiques et sociaux. Tout le matériel mis en ligne sur ce site est à but non lucratif. Il est mis à la disposition de tous ceux qui s'y intéressent dans le but de faire de la recherche ainsi qu'à des fins éducatives. Si vous désirez utiliser du matériel protégé par le droit d'auteur pour des raisons autres que "l'utilisation équitable", vous devez demander la permission au détenteur du droit d'auteur.

Contact média: media@globalresearch.ca