

Article invité : Les dangers de la manipulation subliminale de l'IA

Afrodille

Au niveau mondial, en novembre 2021, les 193 États membres de l'UNESCO ont adopté une recommandation sur l'éthique de l'intelligence artificielle. La recommandation sur l'IA reconnaît explicitement les "impacts positifs et négatifs profonds et dynamiques de l'intelligence artificielle (IA) sur les sociétés, les environnements, les écosystèmes et les vies humaines, y compris *de l'esprit humain*" sur. L'accent mis sur les effets sur l'esprit humain est particulièrement remarquable, car l'utilisation de l'IA affecte non seulement l'environnement, mais aussi la pensée, l'interaction et la prise de décision humaines.

A propos de l'auteur:

Rostam J. Neuwirth est professeur de droit et chef du département d'études juridiques mondiales à la faculté de droit de l'Université de Macao, Macao (Chine). Son livre *The EU Artificial Intelligence Act* a récemment été publié par Routledge.

La proposition de loi sur l'intelligence artificielle (IA) de la Commission européenne, publiée en avril 2021, attire encore plus l'attention sur les graves dangers que les systèmes d'IA font peser sur les pensées et le comportement humains. La proposition vise généralement à garantir la fiabilité et la sécurité des systèmes d'IA.

Plus précisément, selon l'article 5, la proposition prévoit une interdiction des pratiques d'IA particulièrement nocives qui violent les valeurs de l'Union. Cette catégorie de systèmes d'IA malveillants comprend ceux qui 1) exploitent une faiblesse d'un groupe spécifique de personnes, 2) évaluent la fiabilité des individus (*systèmes de notation sociale*), 3) utilisé pour l'identification biométrique à distance en temps réel, et 4) les systèmes d'IA utilisés pour l'influence subliminale en dehors de la conscience d'une personne.

Combinaison avec des interfaces cerveau-ordinateur

Cette dernière catégorie de "systèmes d'IA subliminale" en particulier doit se voir attribuer un rôle clé dans la présente proposition législative. La principale raison en est que l'IA, en combinaison avec un nombre toujours croissant d'autres technologies de pointe, telles que les interfaces cerveau-ordinateur (BCI), l'imagerie par résonance magnétique fonctionnelle, le big data, l'Internet des objets, la blockchain, la robotique ou Eye technologies de suivi et bien d'autres, est capable non seulement de manipuler l'esprit des utilisateurs, mais aussi d'influencer de manière significative leur comportement. Ces possibilités réelles de manipulation des pensées et des comportements ne feront qu'augmenter à l'avenir. La programmation des ordinateurs a déjà été élargie linguistiquement et techniquement pour inclure la programmation des personnes.

Techniques de manipulation connues depuis 100 ans

Que les dangers de la manipulation subliminale soient réels était déjà prouvé par des études menées avant 1917 par le neurologue et psychiatre autrichien Otto Pötzl. Plus tard, les mêmes techniques ont été utilisées aux États-Unis pour accroître l'efficacité de la publicité, comme l'utilisation d'images subliminales pour vendre du pop-corn et certaines boissons gazeuses. Dans un livre de 1957 ("*Hidden Persuaders*"), l'auteur Vance Packard a averti que l'utilisation de techniques subliminales est et sera utile non seulement pour la publicité commerciale mais aussi pour la manipulation de processus politiques tels que les élections.

En Europe et ailleurs, ces préoccupations ont finalement conduit à une interdiction générale de la publicité subliminale à la télévision à partir de 1989. Dans les années 1990, les questions de manipulation subliminale ont de nouveau défrayé la chronique aux États-Unis dans le cadre de plusieurs procès contre des musiciens de rock dont les chansons ont été remplacées par des messages inversés (*masquage arrière*) seraient responsables du suicide de plusieurs jeunes porteurs de messages subliminaux.

De l'espionnage cérébral aux deep fakes

Aujourd'hui scientifiquement il n'y a plus aucun doute sur l'efficacité et les effets réels des techniques subliminales. Dans le domaine du neuromarketing, la question n'est pas formulée *simais* seulement plus *comment* efficaces ces méthodes subliminales sont. L'état actuel de la technique se caractérise le mieux par ce que l'on appelle les logiciels d'espionnage du cerveau (*cerveau logiciel espion*) illustré. En utilisant des techniques subliminales et un modèle d'apprentissage automatique, cela permet d'accéder à des données privées telles que des informations bancaires, des codes PIN, un lieu de résidence ou une date de naissance dans le cerveau.

PATRICK T. FALLON / AFP

Également appelés « motifs sombres » (*motifs sombres*), qui dénotent une conception manipulatrice de sites Web et de publications sur les réseaux sociaux, utilisent des éléments trompeurs indétectables pour manipuler les utilisateurs par diverses méthodes, telles que B. par la publicité voilée, le déclenchement de la peur, ou « appât et interrupteur » (leurre et interrupteur). De même va *contrefaçons profondes* ou la combinaison de la publicité ciblée avec des algorithmes cachés utilisés par les moteurs de recherche, non seulement pour déterminer le succès des entreprises mais aussi le résultat des élections, comme l'a révélé le scandale Cambridge Analytica.

Les dommages cumulatifs devraient également être inclus dans la loi

Les graves dangers associés à la manipulation subliminale des pensées et du comportement exigent que la législation proposée élargisse la portée juridique des préjudices physiques ou psychologiques pour inclure d'autres formes, telles que les préjudices financiers, économiques, culturels et sociétaux. Étant donné que ces formes de manipulation ont tendance à être subtiles et à se propager à travers une variété de supports, leur impact augmente avec le temps. Il est donc également nécessaire de mieux définir la portée de ces techniques subliminales afin que les dommages cumulés, c'est-à-dire les dommages causés par une consommation continue, puissent également être inclus.

Pas de seuil absolu pour la perception subliminale

En ce qui concerne la manipulation des pensées et de l'esprit, il est également nécessaire de clarifier la question des organes sensoriels correspondants. Cependant, selon la psychophysique (c'est-à-dire la branche de la psychologie qui étudie les interrelations entre les stimuli physiques et la perception de ces stimuli), il n'y a pas de seuil absolu pour la perception subliminale, car les seuils pour un stimulus donné varient à la fois entre différentes personnes et au sein d'une personne varient.



Clément MAHOUEAU contre AFP

Dans le même temps, de nombreux systèmes d'IA utilisent des techniques de manipulation qui fonctionnent à la fois en dessous (subliminal) et au-dessus (supraliminal) du seuil de conscience. Dans ce cas, l'interdiction des systèmes d'IA devrait donc être étendue du "recours aux techniques subliminales" aux "techniques supraliminales", ou être remplacée par le terme "techniques transliminales". La raison en est que les stimuli de manipulation sont dynamiques et opèrent à la fois au-dessus et au-dessous du seuil de perception consciente.

Une définition plus large des techniques d'IA est nécessaire

L'exemple du terme "addiction by design" montre que l'intensité et la quantité des stimuli ne cesseront d'augmenter avec la future intégration des sens dans la soi-disant "réalité augmentée et virtuelle" (réalité augmentée et virtuelle). se propage à travers le concept de métaverse. Pour cette raison, une définition plus large des techniques de manipulation utilisées par l'IA est également justifiée et nécessaire, puisque le processus de perception de la vie inclut tous les sens et représente ainsi une expérience multi-sensorielle. Pour la même raison, l'étude traditionnelle des sens, qui les considérait isolément et limitait leur nombre à cinq, est aujourd'hui clairement obsolète.

Débat mondial stimulé par l'UNESCO

Dans l'ensemble, l'interdiction prévue par l'UE des pratiques d'IA nocives, qui incluent également les systèmes d'IA subliminaux, doit être saluée. Dans le même temps, cependant, diverses questions scientifiques connexes devraient être examinées de plus près et les ajustements appropriés apportés au texte juridique final. Dans le même temps, un débat mondial est également nécessaire, comme celui stimulé avec succès par la recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle, afin non seulement de pouvoir utiliser les opportunités attendues de l'IA, mais aussi de pouvoir contrer les dangers réels et les dommages en conséquence.

En conclusion, il convient de rappeler que déjà en 1948 George Orwell dans son roman *1984* a averti que l'invention de l'imprimerie, du cinéma et de la radio a facilité l'influence sur l'opinion publique, mettant fin à la "vie privée". Traduit dans le présent et l'avenir, cela signifie que l'IA et les technologies connexes, telles que *cerveau logiciel espion* constituent une véritable menace aujourd'hui qui pourrait mettre fin à la « confidentialité de la pensée » à l'avenir.