Le Temps Publié le jeudi 6 avril 2023 à 05:35

Fausses photos, vidéos et voix : une terrible ère du doute commence

Technologie

L’intelligence artificielle, à portée de tous, permet de créer des images très facilement. Désormais, il est aussi possible de fabriquer des vidéos à partir de textes. Démêler le vrai du faux s’annonce très difficile...

Anouch Seydtaghia

L’histoire serait-elle en train de mal tourner? En ce mois d’avril 2023, le doute devient immense. Ces dernières années, l’intelligence artificielle (IA) annonçait de folles promesses: une aide si précieuse pour la médecine, une recherche en médicaments accélérée, une automatisation de travaux fastidieux… Mais, aujourd’hui, c’est le côté sombre de l’IA qui s’impose. Ou plutôt son utilisation à des fins malveillantes. Photos artificielles qui semblent si vraies, voix de synthèse troublantes, fausses vidéos conçues à partir de simples textes: le pire de la technologie s’affiche, faisant craindre une déferlante de manipulations en tous genres.

1. Des photos artificielles ultra-réalistes

Des fausses images, rien de nouveau à cela. Celle d’un requin prétendument en train de nager dans une autoroute inondée, créée avec Photoshop, date de 2012. Mais il y a plus de dix ans, les réseaux sociaux étaient embryonnaires, Photoshop coûtait une fortune et exigeait de solides connaissances. Cette semaine, il n’a sans doute fallu que quelques secondes à un grand fan de Donald Trump – son pseudo est Brick Suit sur Twitter – pour créer une photo de l’ancien président paradant dans les rues de New York, une foule immense le suivant. L’image, republiée par son fils Eric Trump, a été vue plus de 6 millions de fois.

D’Elon Musk main dans la main avec la directrice de General Motors, Mary Barra, à Emmanuel Macron au milieu d’une montagne d’ordures, les fausses images se multiplient grâce à la démocratisation des outils pour les créer. Midjourney, Dall-E, Stable Diffusion sont de plus en plus accessibles, Microsoft venant d’ailleurs d’intégrer Dall-E à son moteur de recherche Bing.

Selon Sébastien Marcel, responsable du groupe de recherche en sécurité biométrique et protection de la vie privée de l’institut Idiap de Martigny, cette démocratisation n’est pas une surprise. «C’est une

technologie que nous suivons depuis plusieurs années, c'était inévitable.» Le spécialiste prédit: «De même que la qualité des images synthétiques va s'améliorer, les méthodes de détection vont s'adapter si les financements de recherche dans ce domaine suivent.»

Un point de vue que partage Geovani Rizk, postdoctorant spécialisé en IA, travaillant dans le Laboratoire de calcul distribué de l'EPFL, dirigé par Rachid Guerraoui: «Ces neuf dernières années, des milliers de travaux de recherche ont repris un modèle de génération d'images et ont proposé des améliorations. Pour beaucoup de ces travaux, les codes sources utilisés sont disponibles en open source et donc à la disposition de n'importe qui. Il devient donc plus simple pour quiconque de reprendre les modèles déjà construits et de lancer l'apprentissage pour obtenir un résultat semblable à ceux proposés dans les papiers de recherche sans en avoir une compréhension poussée.»

Selon Geovani Rizk, «la mise à disposition de nouveaux outils comme les modèles génératifs engendrera indéniablement une augmentation de ces images sur les différents réseaux. Il faudra donc rester vigilant sur ce que l'on peut voir sur une image et toujours essayer de croiser les informations: par exemple, d'où provient l'image? Qui la publie?, etc.» Si l'on revient à l'image de Donald Trump à New York, on remarque que plusieurs personnes l'accompagnant ont des déformations au niveau du visage et des mains. La génération d'images se perfectionne. Mais elle n'est pas parfaite.

2. L'explosion des fausses vidéos

Créer des vidéos à partir de quelques lignes de texte, c'est possible. En 2022, Meta, maison mère de Facebook, avait dévoilé un outil de ce type appelé Make-A-Video. Ces derniers jours, la start-up Runway permet de créer des mini-clips de quelques secondes: les scènes sont chaotiques, les personnages étranges, mais cela fonctionne. De son côté, ModelScope a lancé un système que des internautes ont par exemple utilisé pour créer des vidéos de Will Smith ou Scarlett Johansson se goinfrant de spaghettis à la main. On a aussi vu Emmanuel Macron – encore lui – courir au milieu de déchets.

Selon Sébastien Marcel, «on peut s'attendre également à une évolution dans la génération de vidéos synthétiques, mais cela ne sera pas aussi rapide car il est beaucoup plus complexe de générer une séquence d'images synthétiques cohérentes». Pour Geovani Rizk, «pour la vidéo, prendre l'apparence d'une autre personne demande un grand nombre de données pour que cela puisse être réaliste – en tout cas pour l'instant. La plupart des *deepfakes* générés portent sur des célébrités ou des personnes politiques, puisqu'il est aisé de récolter plusieurs dizaines d'heures de vidéo avec leur visage. Prendre l'apparence d'une personne moins exposée médiatiquement donnera un résultat moins réaliste.» A noter que la menace de voir de fausses vidéos pornographiques devient de plus en plus élevée, avec le risque de dégâts considérables.

3. Des voix déjà usurpées

C'est sans doute le domaine où l'IA a atteint une quasi-perfection. «Il est possible de synthétiser de la voix à partir d'un texte ou d'une autre voix», assure Sébastien Marcel. Le chercheur entrevoit un autre phénomène: «Nous sommes très près de pouvoir observer des *deepfakes* audiovisuels générés en temps réel pour, par exemple, convertir le visage et la voix d'une personne dans le visage et la voix d'une autre personne. L'enjeu est l'usurpation d'identité pour de la fraude, du chantage ou de la désinformation.»

Geovani Rizk estime lui aussi qu'«il est déjà possible d'imiter de manière convaincante la voix de quelqu'un lorsque vous avez assez de données. Comme pour le point précédent sur la vidéo, en l'absence de ces données, il sera bien plus compliqué de produire un résultat convaincant.»

Au final, détecter de fausses images, vidéos, voix (ou autre chose) deviendra-t-il très rapidement impossible? «Pour les images et les vidéos, je pense que cela deviendra très rapidement impossible de détecter cela à l'œil nu», estime Geovani Rizk. Sébastien Marcel est plus optimiste: «Les technologies de détection suivront. Le plus important, c'est de rester dans la course, de continuer à analyser les nouvelles techniques de génération pour les comprendre et pouvoir anticiper.»

Liens

Deepfakes : SwissInfo - https://www.swissinfo.ch/fre/sci-tech/comment-les-deepfakes-changent-notre-vision-de-la-r%C3%A9alit%C3%A9/46865490

Fausses photos : Le Temps - https://www.letemps.ch/economie/cyber/fausses-photos-videos-voix-une-terrible-ere-doute-commence

Le clonage vocal : RTS - https://www.rts.ch/info/sciences-tech/technologies/13850084-deepfakes-le-clonage-vocal-la-nouvelle-menace-pour-lidentification-en-ligne.html

Deep Fakes : EPFL - https://epfl-pavilions.ch/fr/exhibitions/deep-fakes-art-and-its-double

What is a deepfake : Tech Accelerator - https://www.techtarget.com/whatis/definition/deepfake

Tout savoir sur le deepfake : OMPI - technique de synthèse reposant sur l'intelligence artificielle - https://www.wipo.int/wipo_magazine/fr/2022/02/article_0003.html