Publié le 5 juillet 2019

Deepfakes : une technique avancée de manipulation de l'information

Devons-nous toujours croire ce que nous voyons ? Pouvons-nous toujours faire confiance aux vidéos qui circulent sur internet ? Avec le développement de nouvelles technologies de pointe, on peut en effet aujourd'hui truquer une vidéo de manière aussi réaliste qu'indétectable : il s'agit d'une deepfake, un phénomène nouveau qui risque de remettre en cause le rapport de l'homme à la vidéo, avec de multiples conséquences dans la vie sociale.

Devant le Centre américain pour le Progrès, le président de la chambre des représentants et grande figure du parti démocrate aux États-Unis, Nancy Pelosi, termine laborieusement son discours. Elle hésite, bute sur les syllabes, et ses phrases s'enchaînent difficilement. La vidéo le prouve : elle n'est pas en état de défendre quoi que ce soit. Une aubaine pour les membres du parti républicain, car elle est une des figures les plus virulentes contre Donald Trump : elle est «ivre» ou sujette «à des addictions», commente-t-on. La vidéo a été vue par plus de deux millions de personnes aux États-Unis, mettant en jeu à la fois la réputation et la crédibilité politique de Nancy Pelosi.

Le *Washington Post*, pourtant, s'étonne. Lors de la conférence, sur place, Nancy Pelosi parlait à un bon rythme, avec un débit clair et soutenu. Il découvre que la vidéo a été trafiquée, sa vitesse réduite de 75 %, et le débit de voix adapté à cette lenteur.

Voilà une *deepfake* !

Le développement des *systèmes algorithmiques*

Ces dernières années, ce que l'on appelle à tort et par abus de langage des «intelligences artificielles[1]» (I. A.), a été développé par l'homme afin que ces procédés techniques exécutent des exercices de plus en plus complexes (d'un simple algorithme pour créer un adversaire dans un jeu vidéo jusqu'à des algorithmes complexes pour battre un champion du monde au jeu de go). En 2018, par exemple, un système algorithmique perfectionné a pu réaliser, en analysant tous les tableaux de Rembrandt, une peinture originale simulant le coup de pinceau de l'artiste[2]. Est-ce pour autant une œuvre du grand peintre ?

La technologie qui a été utilisée s'appelle «réseaux antagonistes génératifs» (GAN, en anglais). Il s'agit d'un système numérique composé de deux réseaux de «neurones» indépendants, dont le premier va réaliser une œuvre en se fondant sur des tableaux de maître existants et le second va

1 - «Intelligence artificielle» (I. A.) est le nom donné à un procédé technique (un système d'algorithmes qui traite les images et les sons) qui permet de simuler l'intelligence. Ce n'est pas une «intelligence» à proprement parler, même si ce terme d'I. A. est utilisé actuellement pour désigner ce procédé technique.

2 - *Huffington Post* du 22 octobre 2018.

tenter de déterminer si cette œuvre est le fruit d'un travail humain ou a été réalisée par une machine. Ce système de calcul fonctionne par répétition, et l'algorithme s'améliore au fur et à mesure de ses échecs : c'est du *deep learning*.

SI l'I. A. réalise un produit fini que son concepteur ne prévoyait pas, ce dernier a quand même prévu, dans le programme et dans l'algorithme, qu'elle réalise un produit de ce type. Dans le cas du tableau ressemblant à un Rembrandt, le concepteur de l'algorithme a prévu que l'algorithme ferait une toile ressemblant à un Rembrandt, même s'il n'a lui-même aucune idée de l'apparence finale de la toile.

N'oublions jamais que l'I. A. est un algorithme programmé. Elle est un processus de calcul qui imite le raisonnement humain. Il ne s'agit pas d'une intelligence, mais d'un programme qui tente de reproduire certaines fonctions de l'intelligence humaine.

Deepfake : de quoi s'agit-il ?

La *deepfake* est une fausse vidéo, réalisée par une I. A. programmée, et s'améliorant grâce à la technologie de «réseaux antagonistes génératifs» (technologie GAN). La grande force de la *deepfake* par rapport aux *fake news*, c'est qu'il est extrêmement difficile pour un internaute qui la regarde de discerner si la vidéo a été manipulée ou non, si elle est fausse ou vraie. Il s'agit d'une véritable vidéo dans laquelle plusieurs éléments sont assemblés de telle manière qu'elle donne une impression de vérité. Elle est destinée à tromper ceux qui la regardent.

L'humoriste Jordan Peel a diffusé une vidéo dans laquelle il explique le risque de manipulation. Dans cette vidéo, on peut voir Barack Obama, assis et souriant, discourant sur la politique américaine. Tout d'un coup, l'ancien président des États-Unis traite son successeur, Donald Trump, d'«abruti». C'est à ce moment que l'humoriste explique qu'il s'agit d'une fausse vidéo, montée de toutes pièces grâce à un algorithme GAN, qui reproduit les traits, les mouvements du visage, les mouvements de la tête, la voix, l'intonation, le jeu du corps et des mains. Sans l'intervention de l'humoriste, il aurait été très difficile pour une personne seule devant son ordinateur de ne pas croire ce qu'elle venait d'entendre et voir.

C'est tout le danger des *deepfakes* : elles créent une réalité virtuelle presque parfaite dans son imitation du réel.

Les détournements possibles à des fins politiques et diplomatiques

Trafiquer une photo dans un but politique ou diplomatique est une méthode malhonnête, mais efficace, qui a souvent été utilisée pour décrédibiliser un adversaire ou provoquer des conflits dans le monde. Dans la guerre, l'art de la désinformation – c'est-à-dire l'art de tromper l'ennemi avec de fausses informations ou des informations partielles – permet de prendre un avantage certain sur son adversaire. Avec les nouveaux progrès technologiques, manipuler une vidéo pourra servir des objectifs à la fois politiques et diplomatiques : «Les *deepfakes* pourraient être utilisées pour créer des mensonges très réalistes, capable d'encourager la violence, de discréditer les dirigeants politiques et les institutions ou alors de faire basculer les élections», explique le *Foreign Affairs* dans un article du 11 juin 2019.

Selon le même article, le risque premier est de déstabiliser l'organisation politique des pays. Par exemple, la revue imagine le cas d'un homme politique qui serait pris dans une affaire politique. Une vidéo l'incriminerait directement, et la vidéo semblerait authentique. L'homme politique pourrait lancer le discrédit sur cette «preuve» en criant à la *deepfake*. Avec cet exemple, le journal essaie de démontrer que la crise de confiance d'un peuple envers son élite et ses dirigeants pourrait bien être amplifiée par les *deepfakes*.

Le réalisme de vidéos truquées ajouté à la vitesse de propagation des réseaux sociaux constitue le deuxième risque des *deepfakes*. Les réseaux sociaux sont des accélérateurs de nouvelles, qu'elles soient bonnes ou mauvaises, vraies ou fausses. Le risque réside dans le fait qu'une grande partie de l'opinion peut être convaincue par une vidéo fausse et trafiquée, ce qui peut entraîner la possibilité de faire basculer une élection, de provoquer un coup d'État, etc.

Ainsi, la pratique du *Revenge Porn* consiste à publier sur internet des vidéos ou des photos pornographiques d'un ancien amant : il s'agit de se venger d'une relation brisée, et le phénomène se développe chez les jeunes. Sur le même modèle, avec une vidéo *deepfake*, on peut imaginer le plus respectable des hommes politiques se retrouvant sur un site pornographique dans une situation compromettante qui n'aurait pourtant jamais eu lieu. Cela attaque à la fois sa réputation, son intégrité, ses relations personnelles et sa famille. Dans le même domaine, le *Deep Nude* permet de déshabiller une femme en utilisant un algorithme de *deepfakes*[3], à partir d'une photo sur laquelle elle est entièrement habillée. Chantage, extorsion, *Revenge Porn* : les manipulations possibles sont multiples et dangereuses. On peut imaginer une application visant à exercer un chantage sur une femme politique, afin d'obtenir son influence pour enterrer un projet de loi, ou sur un juge afin de changer le cours d'un procès…

Ces manipulations, en plus de s'attaquer à des particuliers, risquent de mettre en péril toute la société, de rendre instables les gouvernements du monde et leur classe politique, mettant en danger, à terme, le Bien commun.

Peut-on encore croire ce que l'on voit dans une vidéo ?

Sur internet, la vidéo et l'enregistrement sonore constituaient jusqu'à présent des «preuves-reines». Il est plus facile de se fier à une vidéo que l'on voit de ses propres yeux ou à un enregistrement que l'on écoute de ses propres oreilles, plutôt qu'à des paroles rapportées dans un article. On s'y fie car il s'agit de supports qui semblent plus proches de la réalité. Chaque spectateur ou auditeur devient une sorte de témoin de l'événement filmé ou enregistré ; il peut dire : «Je l'ai vu, de mes yeux vu !» ou «Je l'ai entendu de mes propres oreilles !» Avec l'avènement des *deepfakes*, il sera de plus en plus difficile de faire confiance, même à une vidéo, car son contenu ne sera plus forcément la réalité.

L'article du *Foreign Affairs* conclut que nous marchons vers un monde où le mensonge et la manipulation auront une place essentielle et déterminante. Malgré la volonté de transparence, il deviendra de plus en plus opaque.

3 - Source : *Numérama*.

Comment arriver à discerner correctement la fiabilité d'une vidéo ou celle d'un enregistrement ? La première des choses est sans doute d'apprendre à lire l'information pour former son intelligence à être assez droite par rapport à la réalité dont on nous informe. Il faut avoir une connaissance vraisemblable de la réalité des faits qui éveille en nous le doute sur l'origine d'une vidéo. C'est ce qu'a fait le *Washington Post* pour Nancy Pelosi : il était douteux qu'une femme aussi qualifiée devienne tout d'un coup aussi hésitante.

Au contraire, si l'intelligence ne guide pas le lecteur, dans un monde où nous serons de moins en moins capables de démêler le vrai du faux, le risque est de créer un climat de suspicion généralisé : l'information est suspecte, l'opinion publique dangereuse, le journalisme partial et partiel…

Le risque est de construire une société sans confiance, sans dialogue et sans liberté, dans laquelle chacun restera enfermé dans ses propres convictions !

Pierre Hardon