Deepfake

Source
Usbek & Rica
Fabien Benoit
9 mai 2019

1. Fausses vidéos : extension du domaine du fake

Les avancées en matière d'intelligence artificielle et de deep learning permettent aujourd'hui de retoucher facilement des vidéos pour leur faire dire ce qu'on veut. Demain, pourrons-nous encore croire ce que nous voyons ? Ou bien sommes-nous en train de basculer dans une nouvelle ère de l'image, celle du fake intégral ?

Peut-être vous souvenez-vous de cette vidéo où Barack Obama, face caméra, traitait Donald Trump de « *pauvre con* ». Une sortie plutôt cavalière pour un homme réputé jusqu'alors pour son élégance et sa mesure. L'ancien Président des États-Unis n'a évidemment jamais tenu ces propos. Il s'agit d'un faux, d'une vidéo retouchée et doublée par le comédien américain Jordan Peel. Et conçue, en quelques clics, grâce à l'application FakeApp.

Ce happening, diffusé sur le site BuzzFeed l'an passé, voulait nous alerter sur l'essor du *deepfake*, une technique permettant à un algorithme de superposer un visage sur une vidéo pré-existante et de modifier son contenu.

Le fake à la portée de tous

Le deepfake, contraction de « deep » pour deep learning et de « fake » pour faux, a fait son apparition en 2017 sur le site communautaire Reddit lorsqu'un de ses utilisateurs a commencé à remplacer les visages d'actrices pornographiques par ceux de stars hollywoodiennes. S'en est suivie la diffusion d'une application gratuite et relativement simple d'utilisation, FakeApp, dont se sont servi d'autres internautes afin de bricoler à leur tour des vidéos.

Cette application a pu tirer partie des avancées réalisées en matière de deep learning - ou « apprentissage profond » - et tout particulièrement d'une technique que l'on nomme GAN (pour « generative adversarial networks » ou « réseaux adverses génératifs »). En somme, la possibilité, pour un algorithme, de générer de nouvelles données à partir de données déjà existantes. Par exemple, il est possible d'analyser des milliers de photos de Barack Obama pour en créer une nouvelle, différente de celles existantes mais totalement crédible. La démocratisation de cette technique, par le truchement d'outils diffusés en open source comme TensorFlow de Google, a pu permettre la conception d'un logiciel tel que FakeApp.

Si la retouche vidéo n'est pas à proprement parler une nouveauté - pensons aux effets spéciaux au cinéma - c'est sa démocratisation qui constitue un phénomène en pleine émergence. Aux États-Unis, l'essor du *deepfake* a suscité une sorte de panique morale. L'apparition de cette technique a été décrite comme une nouvelle étape dans l'ère des fake news et de la post-vérité. *« Que se passerait-il si la veille d'une élection au Texas, un deep fake de Beto O'Rourke* - le sénateur démocrate du Texas, ndlr - *en train de coucher avec une prostituée était publié ? »,* s'est ainsi interrogée Danielle Citron, professeure de droit à l'université du Maryland, et co-autrice en juillet 2018 d'un rapport sur le *deepfake*. *« Je sais bien qu'il pourrait être prouvé qu'il s'agit d'un faux, mais si la publication intervient la veille de l'élection, le temps de le faire il serait déjà trop tard,* poursuit-elle. *Un deep fake peut tout à fait venir perturber le processus démocratique ».*

Danielle Citron et son co-auteur Bobby Chesney, lui aussi professeur de droit, voient dans l'essor du *deepfake* une menace pour la démocratie et un phénomène qui a le potentiel pour éroder profondément la confiance des citoyens et diviser encore davantage la société. C'est ce qu'affirme également Solange Ghernaouti, professeure à l'université de Lausanne et spécialiste en cybersécurité : *« On pourrait assister à une abolition de la frontière entre le réel et le virtuel, entre ce qui est vrai et ce qui est faux*, souligne-t-elle. *On peut imaginer qu'à l'avenir on fabrique du deepfake à la demande, adapté à différents publics, pour nous orienter dans nos choix et nos votes. Dans ce cas, si on fabrique des informations taillées sur mesure pour différents publics, on n'a plus de références communes, on ne parle plus de la même chose et il devient impossible de faire société, de débattre ensemble. C'est très inquiétant ».*

Des sénateurs américains ont pris les devants et proposé une loi interdisant la production et la diffusion de *deepfakes*. Une loi dont on peut légitimement douter des effets, au même titre que les différents textes concernant les fake news adoptés ces derniers mois en Europe.

Ceci étant dit, la déferlante de *deepfakes* annoncée ne s'est toujours pas produite pour le moment. *« En vérité, c'est plus compliqué qu'on ne l'imagine de mettre au point des deepfakes vraiment convaincants,* rappelle Patrick Perez, ingénieur en mathématiques appliquées et spécialiste des effets spéciaux. *Toutefois, c'est bien à une démocratisation progressive à laquelle nous assistons ».*

Patrick Perez a fait partie d'une équipe de chercheurs qui a travaillé sur un projet baptisé *Deep Video Portraits.* L'idée était de pouvoir *« générer des vidéos plausibles d'une personne en s'appuyant sur une vidéo réelle - même assez brève - de cette personne là ».* Pour le dire autrement, de pouvoir jouer au ventriloque avec l'image quelqu'un et de lui faire dire ce que l'on veut, de manière plutôt saisissante. La vidéo en question, publiée sur Youtube par le groupe de chercheurs, issus notamment des rangs de Stanford, du Max Planck Institute et de l'entreprise française Technicolor, a été visionnée près de 500'000 fois. *« C'est la première fois dans ma carrière que je vois ça,* commente Patrick Perez, *il y a eu beaucoup d'émoi autour de ce que nous avons fait. Nous avons dû répondre à beaucoup de questions, beaucoup de critiques, beaucoup d'inquiétudes ».*

Le projet, à l'origine, ne vise pourtant qu'à améliorer les techniques de doublage pour le cinéma, notamment en vue de mieux synchroniser les visages des acteurs lors de leurs doublages dans différentes langues. Un perfectionnement de ce que propose déjà une start-up comme Synthesia, qui a produit une vidéo où l'ancienne gloire du foot David Beckham a été doublée en plusieurs langues et les mouvements de son visage synchronisés artificiellement. Mais beaucoup perçoivent déjà les risques de manipulation qui peuvent accompagner ce genre de progrès techniques. *« Pour le moment, il est encore possible de se rendre compte quand une vidéo a été modifiée ou fabriquée,* précise le communiqué de presse du Max Planck Institute, *mais bientôt cela deviendra impossible ».*

Pour Patrick Perez, plusieurs pistes doivent donc être explorées afin de répondre aux craintes de la diffusion de *deepfakes* toujours plus élaborés. La première passe par la vulgarisation. Les gens savent par exemple que la retouche de photo existe, que l'on peut rajeunir quelqu'un sur un cliché, enlever ou rajouter des objets dans une image, et, sachant cela, ils développent un esprit critique plus aiguisé.

Une autre piste réside dans l'amélioration des techniques de détection des vidéos truquées et générées par ordinateurs, un champ dénommé *digital forensics*. Hany Farid, professeur de sciences informatiques à l'université de Berkeley, travaille ainsi sur les très subtiles variations de couleurs de la peau qui se produisent lorsque le sang circule, qui pourraient permettre de vérifier l'authenticité d'une vidéo.

Siwei Lyu, son collègue de l'université d'Albany, tente quant à lui d'analyser le clignement des yeux, tandis que Satya Venneti, de l'université Carnegie Mellon, scrute les pulsations de la circulation sanguine. *« Nous allons sans doute croiser ces recherches basées sur ce que nous connaissons de l'être humain et de son organisme avec le recours à des intelligences artificielles,* pointe Patrick Perez. *Nous entraînerons des IA à reconnaître les fausses vidéos ».* Une dernière piste pourrait également résider dans le fait de « tatouer » les vidéos afin de garantir leur authenticité. C'est une des solutions avancées par Solange Ghernaouti.

Reste à voir ce que l'essor à grande échelle, fort probable, de vidéos contrefaites ou manipulées peut avoir comme conséquence sur notre rapport aux images et à la vérité. Certains, comme le chercheur américain Aviv Ovadya, craignent un phénomène d'apathie face au réel. En somme, si tout peut être faux, nous ne croirons plus rien. Si tout peut être faux et que je suis coupable, je n'aurai plus qu'à nier la réalité.

D'autres, comme le philosophe français Manuel Cervera-Marzal, auteur d'un ouvrage sur les fake news intitulé *Post-vérité, pourquoi il faut s'en réjouir* (Le Bord de l'eau, 2018), pensent au contraire

que c'est notre esprit critique qui en ressortira grandi. *« Je refuse l'usage même du terme post-vérité, qui suppose qu'on est passé dans une ère d'indifférence ou de haine de la vérité*, précise-t-il. *La démocratisation du trucage d'images viendra sans doute avec un esprit critique plus affûté, nous serons plus à même de comprendre ce qu'il est possible de faire, quelles sont les possibilités. Des faux, dans l'histoire, il y en a toujours eu ».*

Source
lesoir.be
Jennifer Mertens
23 mai 2019

2. Une IA capable de reproduire une voix à la perfection

L'humanisation des intelligences artificielles et autres robots passe une étape importante. Exit les voix robotisées, place aux voix « humaines ».

Jusqu'à présent, la distinction des voix de synthèse n'était pas très compliquée puisqu'elles ne ressemblaient pas du tout à celle d'un être humain. Un côté très robotique, saccadée qui pourrait bien disparaître grâce aux avancées technologiques.

Une équipe de chercheurs vient d'ailleurs de démontrer qu'il était tout à fait possible pour une intelligence artificielle de s'exprimer avec une voix parfaitement humaine. Plus encore, l'intelligence artificielle est capable de copier la voix de quelqu'un.

L'entreprise Dessa a réalisé une petite démonstration en faisant parler un androïde avec la voix de l'ancien commentateur MMA, Joe Rogan. Le choix des chercheurs s'est arrêté sur ce personnage public, car depuis plusieurs années, il enregistre des podcasts. Les chercheurs avaient donc une base de données de 1300 enregistrements de *The Joe Rogan Experience* pour tenter de reproduire la voix du podcasteur avec une intelligence artificielle.

Celle-ci a réussi à générer une imitation très crédible. L'entreprise Dessa a publié une vidéo montrant combien son intelligence artificielle, RealTalk, reproduit parfaitement la voix, l'intonation et les mimiques de Joe Rogan.

Comme toute avancée technologique, reproduire parfaitement une voix d'un humain pourrait amener à des dérives dangereuses et néfastes. Les chercheurs de Dessa soulignent d'ailleurs les risques qu'une telle intelligence artificielle représente, à savoir l'usurpation d'identité ou la diffusion de fake news renforcée par l'idée qu'il s'agit d'une personne importante.

Mais les ingénieurs de Dessa soulignent également que cette technologie pourrait également permettre d'améliorer la technologie existante. On pense évidemment aux assistants personnels, mais ce genre de technologie pourrait également être très utile dans le cadre du double au cinéma ou à la télévision.

Contactée par The Verge, Dessa a indiqué qu'elle ne partagerait pas l'ensemble de ses travaux, afin d'éviter que ces recherches ne soient utilisées à des fins malveillantes.

De son côté, l'homme qui a permis malgré lui de développer l'intelligence artificielle, Joe Rogan, trouve l'expérience et cette technologie terrifiante.

Source
Le blog du modérateur
Fabian Ropars
4 septembre 2019

2.1. Deepfake audio : des voleurs imitent la voix du CEO et se font payer 220'000 euros

On connaissait déjà les « arnaques au président » qui consistent à se faire passer pour le CEO d'une entreprise afin de soutirer des fonds. Ces arnaques se faisaient généralement par emails, avec les deepfakes elles se font maintenant par téléphone.*

*CEO : *Chief executive officer* (chef de la direction)

Relayée par The Wall Street Journal, cette histoire pourrait être amenée à se généraliser. Un « Voice Phishing » généré par un algorithme a en effet permis à des voleurs de se faire passer pour un Président d'entreprise allemande, et de recevoir un virement de 220'000 euros. Les voleurs ont utilisé un logiciel de génération de voix par AI pour imiter la voix du Président. Ils ont exigé le transfert immédiat de fonds à un prestataire hongrois. Le Directeur n'a pas tiqué face à la voix familière de son Président, et a donc de son propre gré fait un virement de 220'000 euros à de parfaits inconnus. Les fonds ont évidemment été très vite transférés dans d'autres pays, notamment au Mexique.

Les larcins à base de deepfakes pourraient être le prochain gros levier de développement pour les arnaqueurs et les voleurs. Les possibilités offertes par ces technologies sont en effet très vastes, et permettent de facilement travestir la réalité. Il va en tout cas très vite falloir que les entreprises mettent en places des process pour éviter ces fraudes à base de « social engineering », notamment

des vérifications par email pour toute demande de transferts d'argent.

Source
technikart
Baptiste Manzinali
11 juin 2019

3. On te fake bien profond : le boom du deepfake

Plus vraies que les originales ! Et bidonnées en temps réel… Les deepfakes – des vidéos canulars shootées aux algorithmes et à l'intelligence artificielle – vont perturber les prochaines élections sous l'influence de puissances étrangères. Mytho ou parano ?

Marion Maréchal Le Pen qui fume un joint de weed devant la station Barbès, Benoit Hamon déclarant la guerre aux salauds de gauchistes, ou Emmanuel Macron habillé en black block sur les Champs Elysées, les prochaines échéances électorales devraient livrer leur lot de surprises vidéo-ludiques. Fin janvier, l'expert au Centre pour une nouvelle sécurité américaine, Paul Scharre, spéculait : « *Au cours des deux prochaines années, nous verrons des vidéos truquées jouer un rôle dans les campagnes politiques aux États-Unis ou en Europe (…) pour essayer d'influencer ou de salir les candidats, et ce sera un défi pour les démocraties. »* Car les récents progrès en matière d'intelligence artificielle devraient ravir militants et complotistes acharnés du tweet impulsif, et compromettre au passage nos idéaux démocratiques.

Arme de désinformation massive

Ce qui a mis le feu est poudre est sans doute cette vidéo de Barack Obama publié par BuzzFeed en 2018 et visionnée plus de 5 millions de fois. L'ancien président démocrate s'y adresse sans retenue contre son successeur Donald Trump, le traitant « *d'idiot total et absolu,* avant de reprendre, *nous entrons dans une ère où nos ennemis peuvent nous faire croire que n'importe qui dit n'importe quoi à n'importe quel moment ».* Le rideau tombe enfin lorsque l'image se scinde en deux et fait apparaître le comédien Jordan Peele dont les mouvements faciaux étaient calqués sur le visage de Barack Obama. Une deepfake d'un réalisme total, bluffant, qui pourrait bien se développer au service d'une désinformation de grande ampleur. Florian Silnicki est fondateur de l'agence LaFrenchCom spécialisée dans la gestion de communication de crise.

Sa clientèle : des personnalités publiques, acteurs, politiciens, industriels. « *Il n'y a pas d'activité humaine sans risque. Au théâtre, les rôles sont distribués dès le début, on est le gentil ou le méchant, et c'est difficile de déconstruire un schéma de perception. »* Très fréquemment amené à défendre les intérêts de clients victimes de fake news, l'expert voit s'immiscer des nouvelles technologies redoutables qui rendent obsolètes les moyens actuels de lutte contre la désinformation. « *La rumeur peut être combattue sans difficulté. En revanche, j'ai une inquiétude : l'alliance de l'intelligence artificielle, de moyens colossaux et d'algorithmes de plus en plus performants. J'ai aujourd'hui des sollicitations d'avocats étrangers qui me demandent de contribuer à la diffusion d'informations dont ils savent qu'elles sont fausses, depuis la France, pour lancer une croyance collective. On n'a jamais accepté ce genre de prestation, mais j'imagine que certains le font…On n'est encore qu'aux balbutiements. »* Là est le danger : la manipulation de la deepfake pourrait très prochainement sortir du cadre expérimental pour être utilisée par des entités malveillantes.

Vincent Nozick est maître de conférence à l'Université Paris-Est. Dans ses bureaux, à la décoration minimaliste installés à l'ESIEE, école spécialisée en innovation technologique basée à Noisy-Le-Grand, ce chercheur en informatique est devenu un interlocuteur phare depuis qu'il a mis au point une méthode capable de détecter les deepfakes. Il a vu le phénomène se perfectionner et prendre de l'ampleur : « *Un état peut avoir un intérêt à ce qu'un autre pays s'exprime d'une certaine façon sur un sujet, quitte à créer une deepfake d'un président étranger pour créer une angoisse d'un conflit par exemple, et la diffuser à grande échelle. »* En Belgique, une deepfake humoristique faisait dire à Trump des propos pro-écologistes. « *C'était écrit dessus, en sous-titre, que c'était une deepfake. Mais c'était trop tard, psychologiquement on garde une trace. »* À l'inverse, un représentant politique qui a tenu des propos filmés qu'il n'assume plus, pourrait prétendre avoir été victime d'une deepfake…

Wonder woman et hacker anonyme

Pour comprendre la force de frappe de cette technologie dévastatrice, il faut remonter à 2017 pour voir les premières deepfakes apparaitre sur la toile…et ses premières victimes. Alors qu'elle est au cinéma la Wonder Woman de Patty Jenkins, film aux 800 millions de dollars de recette pour

lequel elle reçoit le prix de la meilleure actrice aux Teen Choice Awards – l'actrice israélienne Gal Gadot s'impose comme l'une des personnalités les plus influentes au monde. Ses tenues légères en super héroïne font saliver les adulescents fans de teen movies et les cinéphiles lubriques. Objet de tous les fantasmes, elle va aussi être victime de la Règle 34, l'une des dix lois les plus célèbres d'internet : « *Si ça existe, il y a du porno à ce sujet* ». Début décembre, une mystérieuse vidéo X fait son apparition sur Reddit – site communautaire américain – avec, en vedette, Gal Gadot, débardeur rouge et petit short noir, prête à s'offrir sur un lit blanc immaculé. L'héroïne avait-elle l'intention de devenir la nouvelle girl next door du X à 33 ans ?

En regardant attentivement le contour du visage, on voit des imperfections trahissant la manipulation.

Elle n'est, en réalité, que la cobaye d'un hacker anonyme qui se cache sous le pseudonyme Deepfakes. Son premier essai, même s'il a quelques défauts de conception – certains mouvements du visage sont flous, désynchronisés – annonce le potentiel d'une arme de destruction massive dont le logiciel est en accès libre sur internet. Par la superposition de visages célèbres sur ceux d'actrices pornos en plein ébat sexuel, la Deepfake matérialise les fantasmes de millions d'internautes, et rend la frontière entre réalité et virtuel encore plus ténue. Scarlett Johansson est également une cible. Dans une interview accordée au Washington Post, elle s'avoue vaincue : « *Internet est un vaste trou de ver qui s'auto-détruit. Le fait est qu'essayer de se protéger d'Internet et de sa dépravation est fondamentalement une cause perdue.* »

La technologie utilisée, le deeplearning, qui existe depuis une dizaine d'années, a en effet besoin d'une base de données solide d'images du visage ciblé sous tous les angles possibles pour construire un montage crédible. « *Avant il s'agissait des falsifications d'images. Personne ne s'était frotté à la vidéo, mais moi j'étais prêt »,* explique Vincent Nozick. « *Pour réaliser une deepfake parfaite, il faut une cible dont on a beaucoup de vidéos et d'images, donc un personnage public. A priori, votre voisin ou votre belle soeur, c'est mal barré. »* Cette intelligence artificielle est capable à elle seule d'imiter nos réseaux de neurones pour échanger un visage avec un autre et traduire ses moindres mouvements. « *Le danger, c'est que tout le monde peut le faire puisque la deepfake est une combinaison de logiciels en open source dont TensorFlow de Google, accessible à tous. »* Pire encore, des sites et des applications comme deepfakesapp proposent de mâcher tout le boulot, légèrement fastidieux. « *Tout vient du machine learning, qui consiste à apprendre manuellement à un ordinateur de reconnaître certains objets, visages, qu'on lui soumet image par image. En réglant quelques paramètres, ça demande juste un peu d'expérience sur la collecte de données mais pas de connaissances particulières, ça prendra la journée, pas plus. »* Avant de rendre sa méthode de détection publique, l'universitaire s'est frotté à une problématique inédite : « *Nos bases de données étaient à 80 % du porno, donc on ne pouvait pas les mettre en ligne, les échanges restaient privées. »* Et tant que la deepfake restait dans le porno, seul l'intégrité morale des cibles étaient atteinte. Si son récent virage en politique pose de sérieux problèmes éthiques aux conséquences bien plus graves, quand est-il dans le privé ?

Face2Face, la Rolls-Royce

Chez Facebook, on a compris depuis longtemps le potentiel du deeplearning. Pour preuve, le Français Yann Le Cun, numéro un de l'intelligence artificielle chez le réseau social qui a changé la

face du monde, vient tout juste d'être récompensé par le prix Turing, « *l'équivalent du prix Nobel d'informatique,* selon le Journaldugeek, pour avoir *posé les bases du deeplearning au cours des trois décennies passées. »* Contacté par nos soins, l'intéressé n'a pas souhaité répondre à nos sollicitations concernant les moyens mis en oeuvre par Facebook pour lutter contre la propagation de vidéos *deepfakes*. Sujet sensible ? Certaines entreprises auraient de gros intérêts à perfectionner le deeplearning, « *La recherche scientifique est de plus en plus mal financée. Développer un outil performant en deeplearning et le vendre à une société, pour un chercheur c'est le jackpot. »* On peut s'interroger à ce sujet sur le récent silence de l'équipe de chercheurs allemands menée par Matthias Niessner, à Munich. En 2016, ils publiaient sur youtube la Rolls-Royce de la deepfake : un logiciel capable de capturer les mouvements d'un visage et de les calquer sur un autre, mais cette fois en temps réel, avec l'aide d'une simple webcam. Trump, Bush, Poutine, figuraient parmi les victimes de cette méthode appelée Face2Face. « *C'est méga flippant parce que là, il n'y a même plus besoin de collecter des données, c'est génial »,* jubile Vincent Nozick, qui met en garde cependant contre les réelles intentions de ses confrères allemands. « *Ma méthode de détection ne fonctionne pas avec Face2Face, donc elle est déjà dépassée. Là où je m'interroge, c'est que les Allemands n'ont pas mis leur logiciel en accès libre comme d'autres l'ont fait. Soit ils se sont rendus compte de sa dangerosité, soit ils l'ont vendu. Cela peut représenter plus de 100 fois la dotation annuelle de leur laboratoire, c'est inestimable ! »*

De Hollywood à la DARPA

Loin d'être envisagée dans le seul but de nuire, la deepfake est déjà utilisée dans la production audiovisuelle à des fins purement créatives. C'est le cas dans la post-production et l'animation. Et demain ? « *Votre acteur meurt au milieu d'un tournage. Avec la deepfake, en un mois, le problème est réglé, vous pouvez finir votre film avec le visage de l'acteur »* plaisante Vincent Nozick. Dernièrement, une vidéo où la voix de David Beckham est remplacée par celles de survivants de la malaria a permis de récolter trois millions de dollars pour une association humanitaire. Selon Claire Wardle, directrice de la coalition américaine First Draft, organisme de détections des fakes news, l'impact de la deepfake dans la sphère politique n'est, pour l'heure, que pure spéculation. Mais qu'en est-il pour les prochaines échéances électorales ? « *Nous sommes à environ quatre ans du niveau de sophistication qui pourrait causer de sérieux dommages et il existe actuellement une course aux armements pour la production d'outils permettant de détecter efficacement ce type de contenu »* explique-t-elle dans une interview accordée à la fondation Nieman Lab de l'Université d'Harvard. D'ici là, ceux qui n'auront pas investi efficacement et anticiper l'impact de la deepfake, pourraient payer très cher leur retard. Aux États-Unis, la DARPA, l'agence de défense gouvernementale spécialisée dans la recherche et le développement des nouvelles technologies à usage militaire, a déjà investi 60 millions de dollars pour lutter contre les deepfakes.

Son équivalent français, la DGA, n'a rien anticipé de tel. Idem au niveau européen. Le chercheur Vincent Nozick le sait bien : « *Les seuls qui sont au point, c'est les Américains et de très très loin. Ils ont tout l'argent qu'il faut pour être hyper bon, même dans le privé, rien que le budget de la recherche d'Amazon, c'est quatre fois celui du CNRS. C'est juste risible. »* Pierre Ganz, vice-président de l'ODI (l'Observatoire de la déontologie de l'information), conclut : « *Quand on voit que des journalistes ne sont pas capables de détecter un poisson d'avril… Il va falloir traquer, mais avec des formations et des outils. Le risque est plus grand sur le plan démocratique qu'il y a 50 ou 60 ans. »*

4. Faut-il avoir peur de Zao, l'app de deepfakes ?

Source
Korii
Barthélemy Dont
4 septembre 2019

L'application actuellement la plus populaire de Chine soulève bien des inquiétudes.

Après le fulgurant succès de TikTok, une nouvelle app chinoise fait fureur sur internet. Zao est sortie ce vendredi 30 août et est devenue l'application gratuite la plus téléchargée de l'AppStore chinois en quelques jours à peine.

Zao permet de prendre une photo de votre visage puis de l'apposer sur celui d'un personnage choisi parmi une sélection de clips vidéo, tirés essentiellement de films chinois, mais aussi de long-métrages et séries occidentales comme *Titanic* ou *Game of Thrones*. À en juger les vidéos disponibles sur les réseaux sociaux, le résultat est bluffant.

À bien des égards, l'appli est la matérialisation d'un événement prédit depuis quelques mois déjà :

la démocratisation des deepfakes, ces logiciels qui permettent, grâce à l'intelligence artificielle, de calquer un visage sur un corps à partir de simples photos.

Pas d'affolement, mais de la prudence

Les réseaux sociaux se sont rapidement inquiétés des dommages que des deepfakes mis à la portée du grand public pourraient causer – surtout lorsque l'on sait à quelle vitesse des montages mensongers, pourtant techniquement bien inférieurs à ceux de Zao, peuvent se propager sur internet.

Heureusement, Zao ne sera probablement pas responsable d'un incident diplomatique. L'app ne fonctionne pour l'instant qu'avec des clips présélectionnés – ce qui peut d'ailleurs expliquer la qualité des fakes. Rien n'indique que ses algorithmes puissent répliquer leurs prouesses sur d'autres images.

Une autre crainte porte sur la protection des données privées. Les conditions d'utilisation précisent que les photos sont envoyées sur les serveurs de l'entreprise, mais surtout que cette dernière peut ensuite en disposer à sa guise.

CGU : conditions générales d'utilisation

Foudroyée par le géant chinois WeChat pour des raisons de sécurité, Zao a annoncé qu'elle changerait ses CGU, afin que les clichés ne puissent pas être réutilisés dans un autre but sans autorisation.

Compte tenu des méthode de surveillance de masse high-tech du pays, livrer sa tête à une entreprise chinoise reste malgré tout inquiétant. Comme le souligne Wired, ce serait toutefois une méthode bien alambiquée pour le gouvernement chinois, qui n'a pas pour habitude de prendre tant de pincettes pour récupérer des données – en témoigne la récente révélation d'un hack massif d'iPhone.

En définitive, il ne semblerait pas que Zao soit plus nocive que d'autres apps ou sites gratuits auxquels les internautes fournissent leurs données tous les jours – comme Facebook, pour n'en nommer qu'un –, ce qui ne signifie pas qu'il faille lui faire aveuglément confiance.

Source
Numerama
Perrine Signoret
31 juillet 2019

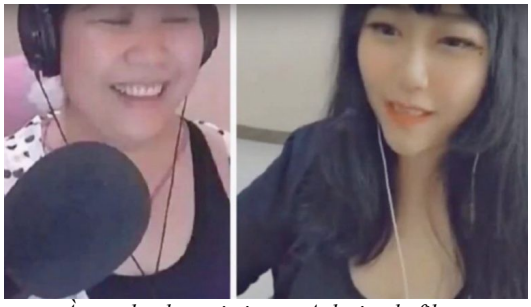
5. Les fans d'une influenceuse ont découvert son vrai visage à cause d'un bug de filtre

Une influenceuse chinoise s'est fait passer pour une femme beaucoup plus jeune grâce à un filtre qu'elle appliquait en continu. Un bug a révélé son vrai visage.

C'est ce qu'on appelle découvrir le vrai visage d'une personne… au sens propre. Une influenceuse chinoise subit des railleries depuis que le filtre qu'elle utilisait dans toutes ses vidéos s'est temporairement désactivé. Elle prétendait être une jeune femme et ses abonnés ont découvert qu'elle était en fait plus âgée et qu'elle ne ressemblait en rien à l'image qu'elle s'était donnée, a rapporté le site HITC ce 30 juillet.

L'influenceuse se fait appeler Your highness Qiao Biluo (son altesse Qiao Biluo). Elle compte 100'000 abonnés sur le réseau social Douyu. Elle est régulièrement complimentée sur son physique qui correspond parfaitement aux critères de beauté stéréotypés de la société en Chine. Elle a un visage petit et fin, une peau lisse, un petit nez légèrement retroussé, des cheveux longs… ou du moins, c'est ce qu'elle faisait croire.

Celle que le China's Global Times décrivait comme une « mignonne déesse » était en fait une pure invention numérique. Pour ressembler à ceci, Qiao Biluo utilisait en permanence un filtre dans ses vidéos. Lors d'un direct le 25 juillet, le filtre a été temporairement désactivé, à cause d'un bug technique. Les abonnés ont alors découvert le vrai visage de l'influenceuse… qui ne ressemble en rien à son « avatar » et qui est plus âgée.



À gauche, le vrai visage. A droite, le filtre.

Un phénomène très répandu en Chine

L'utilisation de filtres sur les réseaux sociaux est très courante. En Chine particulièrement, de nombreuses personnes (majoritairement des femmes) en appliquent lors de leurs directs vidéos.

Ces filtres qui rajeunissent sont souvent accusés de renforcer la pression que les femmes peuvent ressentir sur leur physique. Des chirurgiens racontent que de plus en plus de femmes demandent à se faire opérer pour correspondre à leur image avec un filtre Snapchat ou Instagram. Ce phénomène porte d'ailleurs un nom, la « dysmorphie Snapchat ».

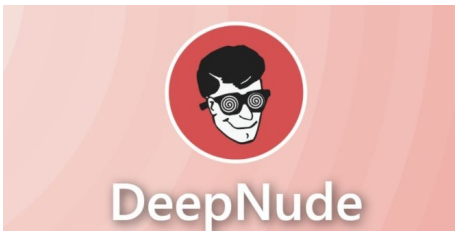
On peut aussi considérer que Qiao Biluo avait décidé d'utiliser le système à son avantage, en profitant des outils technologiques pour se créer une apparence stéréotypée mais mieux acceptée par la société de consommation, et en tirer un profit. Cynique mais habile ? Il semblerait d'ailleurs qu'elle ait continué sa discussion en direct alors que le filtre était tombé, ce qui laisse penser qu'elle ne portait pas grande importance à cette « révélation » — si tant est qu'elle s'en soit rendu compte.

L'histoire a fait beaucoup réagir : selon HITC, il y aurait eu plus de 50'000 hashtags différents au sujet de l'incident. Certains sont moqueurs envers Qiao Biluo, d'autres envers ses abonnés jugés naïfs. Qiao Biluo elle, a a gagné 650'000 abonnés depuis son direct… On ignore si elle continuera à faire ses vidéos avec ou sans filtre.

Source
lebigdata.fr
Bastien L
28 juin 2019

6. DeepNude : l'application qui déshabille les femmes en photo grâce à l'IA

DeepNude est une application qui permet de déshabiller n'importe quelle femme sur une photo en un clic et trente secondes. Elle repose sur l'intelligence artificielle, et a rapidement rencontré un succès fou. Cependant, par souci d'éthique, son créateur a préféré la supprimer au bout de quelques jours…



Parmi les différentes applications d'intelligence artificielle, DeepNude est sans doute celle qui risque d'intéresser le plus de monde. Cette appli permet tout bonnement de retirer les vêtements de n'importe quelle femme sur une photo en un seul clic…

Pour ce faire, les réseaux de neurones se basent sur la photo de la personne habillée pour en créer une nouvelle sur laquelle elle est entièrement nue. Ses vêtements sont automatiquement remplacés par des seins et un vagin. Le processus ne prend que 30 secondes.

Le logiciel est basé sur pix2pix, un algorithme open-source créé par l'Université de Californie en 2017. Cet algorithme utilise les réseaux génératifs antagonistes, qui peuvent être entraînés à générer de fausses images de type Deep Fakes à partir de données.

Dans le cas de DeepNude, l'IA a été entraînée sur plus de 10'000 photos de femmes nues. C'est pourquoi le tour de magie ne fonctionne que sur les images de femmes. Le développeur explique que les photos de femmes sont plus faciles à trouver sur le web, mais compte également développer une version fonctionnant sur les hommes.

Lancée le 23 juin 2019 sur Windows et Linux, l'application fonctionnait comme n'importe quel logiciel pour PC. Elle était proposée en version gratuite ou payante. La version gratuite laissait une watermark sur les images générées. La version payante, proposée pour 50 dollars, se contentait d'une inscription « FAKE » en haut à gauche de la photo.

DeepNude a été supprimée par son créateur, terrifié par son succès viral

Cependant, en quelques jours seulement, DeepNude a été à la fois victime de son succès et de la colère de nombreuses femmes. L'application a généré un tel buzz que les serveurs ont planté. En parallèle, un grand nombre de femmes ont été offensées par la façon dont elle objectifie la femme.

En réaction, le créateur de DeepNude, qui garde l'anonymat, a préféré supprimer son application. Il avoue ne pas avoir pensé qu'elle deviendrait virale, et craint désormais que sa technologie soit utilisée à mauvais escient. En effet, avant ce logiciel, les DeepFakes n'avaient jamais été aussi faciles à créer.

Le développeur a fait le choix de privilégier la morale plutôt que son propre profit. Par conséquent, aucune nouvelle version ne sera lancée et plus personne n'est autorisé à utiliser l'application. Selon son créateur, « le monde n'est pas prêt pour DeepNude »…

On ne peut que saluer cette sage décision, même si certains seront sans doute un peu déçus. Néanmoins, ce logiciel présage d'un inquiétant futur dans lequel n'importe qui pourra diffuser de fausses photos réalistes d'une personne pour la discréditer.

Même si le créateur de DeepNude a fait le choix de supprimer son outil, il est fort probable que d'autres développeurs n'aient pas autant de scrupules à l'avenir. Il suffit d'observer son succès phénoménal et immédiat pour percevoir le potentiel lucratif d'une telle application…

Source
Siècle digital
Arthur Vera
4 juillet 2019

6.1. DeepNude : l'application renaît un peu partout sur le web

La semaine dernière, on apprenait la disparition de l'application DeepNude, en ligne depuis seulement le 23 juin. Cette application permettait d'enlever les vêtements d'une femme en photo grâce à une IA. Cependant, une fois un produit en ligne, on a beaucoup de mal à le faire disparaître. C'est exactement ce qu'il est en train de se passer avec DeepNude.

Si l'application officielle a bien été supprimée par son créateur, de nombreuses copies se développent de plus en plus. Le média américain The Verge a mené une investigation jusqu'à trouver des liens de téléchargement de DeepNude à plusieurs endroits. Que ce soit sur Telegram, YouTube, ou encore GitHub, l'application est encore accessible très facilement. Motherboard, par Vice, aurait même révélé que le code de l'application serait vendu sur un serveur Discord à partir de 20 dollars. Ces vendeurs ont déclaré que le code a été modifié pour rendre le logiciel plus stable, tout en supprimant la fonctionnalité filigrane qui devait empêcher l'utilisation de ces fausses images à des fins malveillantes.

Sur Discord, les vendeurs se vantent d'avoir en leur possession « la version complète et propre de DeepNude V2 ». Cependant, comme tout logiciel libre ayant été modifié, il faut se montrer extrêmement prudent. Il est fort probable que certaines versions aient été modifiées dans le but d'inclure des logiciels malveillants. De ce fait, il est recommandé de se montrer responsable, en évitant de télécharger ces fichiers.

DeepNude est condamnée à continuer d'exister, le tout en causant de nombreux dommages chez plusieurs utilisateurs. Si les commentaires sont souvent négatifs à propos de la qualité de ces images, certaines restent très réussies et ne sont pas faciles à déceler en un coup d'œil.

Source
Futura
Fabrice Auclert
17 juin 2019

7. Espionnage : de faux comptes LinkedIn créés par l'IA

Un espion aurait utilisé une image générée par une intelligence artificielle pour créer un faux compte sur le réseau LinkedIn. Ce profil lui aurait permis d'entrer en contact avec plusieurs personnalités politiques américaines.

L'agence de presse américaine *Associated Press* est tombée sur une affaire des plus curieuses en découvrant non seulement un faux profil sur le réseau social professionnel LinkedIn, mais également que le profil utilise la photo d'une femme qui n'existe pas ! Selon l'AP, ce personnage ferait partie d'une vaste armée de profils fantômes qui rôdent sur le réseau.

Le profil, qui utilise le nom de Katie Jones, ne semblait pourtant pas sortir de l'ordinaire à première vue. Cette femme virtuelle avait établi 52 connexions avec d'autres utilisateurs du site. Contrairement à Facebook, qui privilégie les connexions personnelles, comme les amis et les

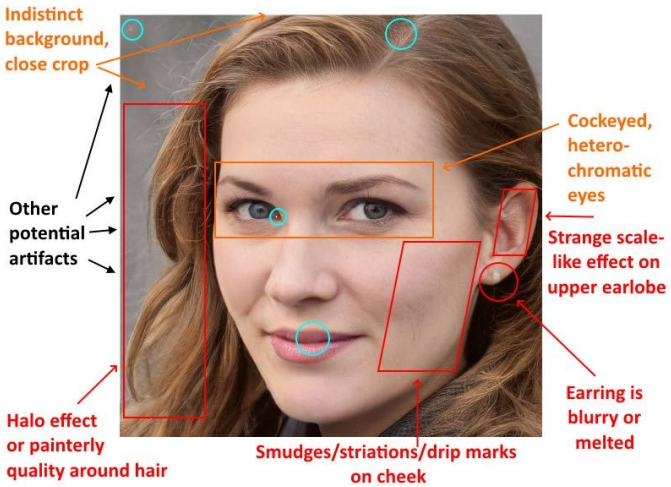
membres de la famille, les connexions sur LinkedIn sont avant tout professionnelles et constituent un terrain propice pour les espions.

Un espion avec des contacts bien placés

Malgré un nombre de contacts limité, Katie Jones était particulièrement bien connectée. Elle se vantait de travailler au Centre d'études stratégiques et internationales, un groupe de réflexion basé à Washington, et comptait parmi ses contacts des personnes influentes. Ses connexions incluaient des organismes comme la *Brookings Institution* ou *Heritage Foundation*, mais également un sous-secrétaire d'État adjoint, un haut conseiller d'un sénateur, ainsi que l'économiste Paul Winfree, candidat pour un siège à la Réserve fédérale.

Cependant, son activité sur le réseau s'est révélée suspecte. Keir Giles, un spécialiste londonien sur la Russie, était méfiant en recevant une invitation de cette femme, après avoir déjà été victime d'une affaire d'espionnage. Se présentant comme chercheuse sur la Russie et l'Eurasie au Centre d'études stratégiques et internationales, il aurait dû déjà en entendre parler, mais Katie Jones lui était inconnue. Un porte-parole du centre a également confirmé que personne portant ce nom n'y travaillait.

Un visage généré par l'intelligence artificielle



Plusieurs experts, dont Mario Klingemann, un artiste allemand qui expérimente depuis des années avec des portraits générés automatiquement, ont commenté la photo et confirmé qu'il s'agit bien d'une fausse. Plusieurs éléments révèlent l'origine de la photo, comme le fond flou, des traits sur sa joue, ou encore une boucle d'oreille qui semble à moitié fondue. Il s'agirait d'un deepfake, une génération automatique basée sur le deep learning. Le procédé utilise deux réseaux neuronaux antagonistes (GAN), l'un pour générer des visages après avoir été entraîné sur des images existantes, et le second pour estimer le réalisme de l'image ainsi produite, et la valider ou non.

Plusieurs experts ont indiqué que son profil était assez typique dans les opérations d'espionnage. De nombreux utilisateurs acceptent toutes les invitations, y compris des personnes qu'ils ne connaissent pas. Si une simple connexion ne présente pas de danger en soi, tous les contacts bien placés de Katie Jones ont beaucoup augmenté sa crédibilité. L'espion, qui se cache derrière ce faux profil, peut ensuite prendre contact avec d'autres personnes beaucoup plus facilement. Le compte a depuis été supprimé, et LinkedIn a indiqué prendre régulièrement des mesures contre les faux comptes, avec plusieurs milliers supprimés au cours du premier trimestre de 2019.

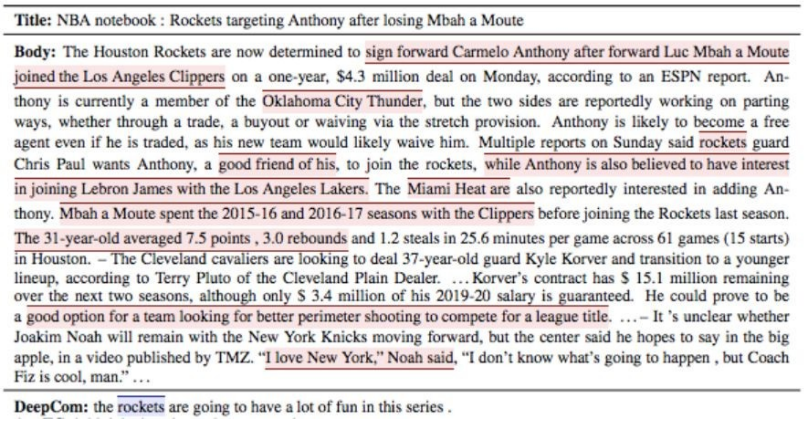
Source
Siècle digital
Vincent Glad
3 octobre 2019

8. Personne ne le réclamait, ils l'ont quand même inventé : le bot qui crée de faux commentaires de presse

Il ne manquait plus que cela. Alors que l'Internet est déjà inondé de fake news, de trolls ou de bots, une équipe chinoise de Microsoft annonce avoir créé une IA permettant de générer de faux commentaires sur les sites de presse. Dans un article publié sur arXiv, les chercheurs arguent de

l'utilité de leur recherche : de faux commentaires peuvent permettre de lancer la conversation autour d'un article quand aucun humain n'a commenté. Ils peuvent aussi, assurent-ils, améliorer l'expérience de lecture sous un article peu commenté.

Le bot, nommé DeepCom, a ingurgité une base de données de millions de commentaires d'articles postés en chinois et une autre base de données de commentaires en anglais, issus de Yahoo! News. L'IA s'attache d'abord à définir le sujet principal de l'article, celui qui doit générer normalement le plus de commentaires. Par exemple, le nom de l'actrice ou acteur principal si l'article porte sur un film. Ensuite, DeepCom lâche son com'.



Title: NBA notebook : Rockets targeting Anthony after losing Mbah a Moute

Body: The Houston Rockets are now determined to sign forward Carmelo Anthony after forward Luc Mbah a Moute joined the Los Angeles Clippers on a one-year, $4.3 million deal on Monday, according to an ESPN report. Anthony is currently a member of the Oklahoma City Thunder, but the two sides are reportedly working on parting ways, whether through a trade, a buyout or waiving via the stretch provision. Anthony is likely to become a free agent even if he is traded, as his new team would likely waive him. Multiple reports on Sunday said rockets guard Chris Paul wants Anthony, a good friend of his, to join the rockets, while Anthony is also believed to have interest in joining Lebron James with the Los Angeles Lakers. The Miami Heat are also reportedly interested in adding Anthony. Mbah a Moute spent the 2015-16 and 2016-17 seasons with the Clippers before joining the Rockets last season. The 31-year-old averaged 7.5 points , 3.0 rebounds and 1.2 steals in 25.6 minutes per game across 61 games (15 starts) in Houston. – The Cleveland cavaliers are looking to deal 37-year-old guard Kyle Korver and transition to a younger lineup, according to Terry Pluto of the Cleveland Plain Dealer. … Korver's contract has $ 15.1 million remaining over the next two seasons, although only $ 3.4 million of his 2019-20 salary is guaranteed. He could prove to be a good option for a team looking for better perimeter shooting to compete for a league title. …– It 's unclear whether Joakim Noah will remain with the New York Knicks moving forward, but the center said he hopes to say in the big apple, in a video published by TMZ. "I love New York," Noah said, "I don't know what's going to happen , but Coach Fiz is cool, man." …

DeepCom: the rockets are going to have a lot of fun in this series .

Exemple de faux commentaire généré par DeepCom. En rouge, ce que le bot a jugé important et digne d'être commenté dans l'article.

Comme l'illustre cet exemple d'un article sur une équipe de NBA, DeepCom est encore loin d'être pertinent et de pouvoir vraiment lancer la conversation. Qu'en sera-t-il quand ce genre d'outil sera plus performant ? La publication des chercheurs chinois a suscité **l'ironie sur les réseaux sociaux**. Arvind Narayanan, informaticien et professeur associé à l'Université de Princeton, salue avec sarcasme l'irruption de cette « *nouvelle technique de machine learning dont l'intérêt principal semble être le trolling et la désinformation* ».

Les auteurs de l'étude sont conscients des questions éthiques que pose leur outil : « Il y a un risque que des internautes ou des organisations utilisent ces techniques à grande échelle pour publier de faux commentaires, à des fins de manipulation ou de persuasion politique. » Malgré cet avertissement timide, « l'article n'aborde pas les usages inattendus potentiels de l'outil », dénonce Arvind Narayanan. « Malheureusement, EMNLP [la prestigieuse conférence dans laquelle sera présentée cette étude] ne semble pas demander aux auteurs de questionner les implications éthiques de leur travail. La sécurité informatique a un long historique de travaux utilisés par la suite à mauvais escient ».

Un DeepCom plus robuste, capable de produire des commentaires crédibles, pourrait devenir l'outil rêvé pour l'astrosurfing, cette pratique consistant à créer artificiellement un mouvement de foule sur Internet. En Chine, on parle de « water army » pour qualifier ces usines à trolls payés pour commenter de manière insincère sur les réseaux, une technique actuellement utilisée pour tenter de discréditer les manifestants pro-démocratie à Hong-Kong. Dans le cadre d'une expérience récente, une filiale de Google avait montré qu'il était possible d'acheter les services de troll russes spécialisés dans la désinformation pour seulement 250 dollars. Une intelligence artificielle pourrait encore faire baisser la facture.

Source
Siècle digital
Vincent Glad
8 octobre 2019

9. L'apocalypse du deepfake n'est pas pour tout de suite (sauf dans le porno)

D'après un premier audit, la menace de fake news n'est pas encore avérée mais les fausses vidéos pornographiques prolifèrent.

La rumeur de l'arrivée imminente des deepfakes bruisse dans la presse depuis plusieurs mois

mais la menace peine toujours à se matérialiser. Où sont ces fameux deepfakes ? Sont-ils si faciles à réaliser ? À partir de quand doit-on vraiment paniquer ?

Deeptrace, une start-up d'Amsterdam, a réalisé le premier audit sur la question pour mesurer l'ampleur actuelle du phénomène et savoir quels sont les outils disponibles en libre accès sur Internet pour réaliser de fausses vidéos.

Premier enseignement : la masse de deepfakes s'accroît sensiblement. En sept mois, le nombre de vidéos contrefaites a presque doublé, pour atteindre le nombre (très précis) de 14'678, selon la start-up amstellodamoise. Alors que le risque de fausses vidéos utilisées à des fins électorales suscite un fort intérêt médiatique, cela reste à l'état actuel une menace fantôme.

En réalité, 96% des deepfakes sont des vidéos… porno. Les quatre sites les plus fréquentés se dédiant à cette activité ont reçu pas moins de 134 millions de vues grâce à de fausses vidéos de célébrités. Cette tendance qualifiée d' »inquiétante » par le rapport cible exclusivement les femmes. Au contraire, les deepfakes non-pornographiques s'attaquent majoritairement à des hommes (dans 61 % des cas).

En dehors des sites porno, les deepfakes sont pour l'instant surtout des parodies de célébrité (81 % des contenus non-pornographiques). Plusieurs YouTubeurs jouent avec les premiers outils qu'ils ont à leur disposition. L'idée n'est pas tant de manipuler le public que de tester ces nouveaux joujoux.

Keanu Reeves dans une fameuse scène de Forrest Gump.

Le terme deepfake est apparu fin 2017 sur le forum Reddit, avec la création d'une section u/deepfakes. Reddit l'a rapidement fermé mais la communauté s'est répandue en d'autres lieux. Le rapport de Deeptrace dénombre 20 forums ou sites consacrés à cette activité (dont 4chan et 8chan), avec une petite centaine de milliers de participants. Avec un minimum de bagage technique, il est possible de réaliser un deepfake (en pratique un « faceswap ») grâce à un code disponible sur GitHub.

La plupart des vidéos jusqu'ici présentées dans les médias, beaucoup plus réalistes —et donc beaucoup plus inquiétantes — ne proviennent pas de cette activité amateure des bas-fonds du web mais plutôt de travaux de laboratoires académiques. Ces méthodes de deepfake plus abouties ne sont pas disponibles pour le commun des mortels sur le net. « On ne publie jamais un bout de notre code. Ce serait trop dangereux », explique Siwei Lyu de l'Université de Buffalo, cité par IEEE Spectrum.

La politique devient parano avec les deepfakes

Si les deepfakes n'ont pas encore causé de réel trouble dans la politique, leur simple potentialité change déjà la donne. S'il est possible de fabriquer une fausse vidéo, en conséquence toutes les vidéos deviennent suspectes. Le rapport de la start-up amstellodamoise évoque un cas saisissant au Gabon. Fin 2018, alors que des rumeurs laissent entendre que le Président Ali Bongo, ayant disparu des écrans depuis plusieurs mois, est très malade, la présidence publie une vidéo de Bongo où il présente ses vœux à la nation.

Loin de calmer la rumeur, la vidéo va au contraire relancer les spéculations, des opposants politiques dénonçant un deepfake. L'affaire est loin d'être anecdotique: quelques jours plus tard, la vidéo supposément fake sera utilisée pour justifier une tentative de coup d'État (avorté). Des analyses ultérieures approfondies n'ont trouvé aucune trace de manipulation sur cette vidéo. Menace réelle ou supposée, les deepfakes à des fins politiques sont déjà au centre des préoccupations.

Source
Siècle Digital
Valentin Cimino
30 novembre 2019

10. Diffuser une deepfake en Chine constitue désormais un crime

Comme la Californie, la Chine décide de sévir contre les internautes qui voudraient faire passer de fausses informations grâce à cette technique d'IA. Le gouvernement chinois a fait savoir qu'à compter du 1er janvier 2020, la *Cyberspace Administration of China* (CAC) appliquera une nouvelle loi qui consiste à punir un créateur de deepfakes comme un criminel.

Un créateur de deepfake est considéré comme un criminel

En Chine, dans un mois, il sera obligatoire de préciser qu'une vidéo a été créée grâce à l'intelligence artificielle et qu'elle rapporte de fausses informations, pour qu'elle soit publiée de manière légale. Si ces mentions n'apparaissent pas, le créateur de la deepfake sera considéré comme

un criminel aux yeux des autorités chinoises et donc traité comme tel. Le *South China Morning Post* rapporte les propos de la CAC :

> *"Avec la démocratisation des nouvelles technologies et de l'intelligence artificielle dans l'industrie de la vidéo et de l'audio en ligne, nous avons identifié plusieurs risques. Des contenus comme les deepfakes peuvent perturber la véracité des informations diffusées ainsi que l'ordre social dans le pays. L'intelligence artificielle utilisée dans cet objectif précis pourrait avoir un impact négatif sur la sécurité nationale de la Chine. Pour cette raison, nous ne tolérerons aucun écart"*

Un phénomène qui inquiète les gouvernements du monde

Les deepfakes sont conçues grâce à une technique d'intelligence artificielle qui consiste à superposer des images et des vidéos existantes sur d'autres images et/ou vidéos pour faire dire ce qu'on veut à n'importe qui. Au début du mois d'octobre, la Californie légiférait également sur la question de l'interdiction de ces vidéos trompeuses.

Le gouverneur de Californie a ratifié, le 3 octobre, 2 lois contre les deepfakes politiques et pornographiques. Dans l'État américain, il est désormais illégal de publier une vidéo manipulée pour discréditer un candidat Californie, dans les 60 jours qui précèdent une élection. La seconde loi vise à encadrer l'usage de cette technique dans la pornographie.

Les géants du web s'en mêlent aussi

Facebook et Twitter tentent de trouver des moyens efficaces de lutter contre ce fléau. À ce propos, Facebook a récemment annoncé ses équipes allaient travailler avec Microsoft et le MIT pour mieux détecter les deepfakes et pouvoir les éliminer. Dans le cadre de cette alliance, Facebook promet de créer un ensemble de données qui pourrait être utilisé pour créer de meilleurs outils de détection. 10 millions de dollars (9 millions d'euros) sont investis dans ce projet de défense. Objectif : lutter contre les deepfakes en prévision de l'arrivée des élections présidentielles de 2020.

Tous les experts s'y attendent : les deepfakes risquent de poser un problème majeur lors des élections américaines à venir. Ce processus de permutation intelligente des visages, est en réalité une technique de synthèse d'images basée sur l'intelligence artificielle. Souvenez-vous, en avril 2018, une vidéo de Barack Obama avait été créée grâce à l'intelligence artificielle. Il convient de préciser que pour l'instant, les deepfakes sont majoritairement utilisés pour créer des contenus pornographiques. Seules 4 % des vidéos publiées le sont dans un cadre politique, et bien souvent humoristique.

Source
sciencepost.fr
Brice Louvet
30 avril 2021

11. Les dangers sous-estimés de la géographie deepfake

L'imagerie satellitaire Deepfake constitue une menace de plus en plus sérieuse pour les sécurités nationales d'après des chercheurs. Mais de quoi parle-t-on exactement ? Et quels peuvent être les dangers ?

Le deepfake est une technique de synthèse d'images s'appuyant sur l'intelligence artificielle (IA) pour superposer des fichiers audio et vidéo déjà existants. Vous obtenez alors de nouveaux contenus entièrement faux. Jusqu'à présent, les inquiétudes concernant les deepfake étaient centrées sur les vidéos manipulées par machine de célébrités et autres dirigeants mondiaux disant ou faisant quelque chose qu'ils n'ont en réalité jamais dites ou faites.

Face à ces menaces, certaines grandes entreprises de technologie comme Amazon, Facebook et Microsoft ont lancé conjointement un défi de détection de ces faux.

Toutefois, si les faux discours de politiciens et autres scènes de pornographie impliquant des célébrités se répandant sur les réseaux sociaux ont reçu une large attention du public au cours de ces dernières années, il convient de ne pas sous-estimer une autre menace : celle des **images trafiquées de la Terre elle-même**.

La géographie deepfake

La convergence croissante de l'IA et des systèmes d'information géographique (SIG) ont permis de faire des progrès spectaculaires dans le domaine de l'intelligence artificielle géospatiale. Toutefois, au cours de ces dernières années, des chercheurs ont également été témoins de

conséquences inattendues et problématiques de cette convergence. Citons des problèmes de signaux GPS fabriqués ou encore de fausses photos d'environnements géographiques.

Pour l'heure, ces "faux" n'ont pas encore proliféré, mais certains scientifiques sont de plus en plus préoccupés par la propagation de ces données générées par l'IA. Et pour cause, de telles informations pourraient induire en erreur de diverses manières. Les humains "mentent" en effet avec leurs cartes depuis à peu près aussi longtemps que les cartes existent. Cependant, les conséquences pourraient ne pas être les mêmes aujourd'hui.

Ces fausses informations pourraient notamment être utilisées pour **discréditer des histoires basées sur des images satellitaires réelles**. Pour The Verge, James Vincent prend ainsi l'exemple des camps de détention Ouïghours, en Chine, qui ont gagné en crédibilité grâce aux preuves satellitaires. "*À mesure que la géographie deepfake se généralise*, *le gouvernement chinois pourrait affirmer que ces images sont également fausses*".

Une menace pour la sécurité nationale

Ce type de technique pourrait également être un **problème de sécurité nationale** pour certains pays devant composer avec des adversaires géopolitiques s'appuyant sur des faux pour les tromper.

Todd Myers, responsable de l'automatisation pour la Direction de la technologie CIO à la National Geospatial-Intelligence Agence, a mis en garde l'armée américaine contre cette perspective dès 2019 suite au progrès de la Chine dans ce domaine. Grâce à une technique émergente appelée "réseaux antagonistes génératifs", le pays peut en effet inciter les ordinateurs à **"voir" des objets qui n'existent pas dans des paysages ou dans des images satellites**.

L'analyste avait à l'époque imaginé un scénario dans lequel un logiciel de planification militaire était trompé par de fausses données révélant un pont à un emplacement incorrect. "*D'un point de vue tactique, vous pourriez alors entraîner vos forces à suivre une certaine route vers ledit pont, alors qu'en réalité il n'y a aucun pont*", expliquait alors Myers. "*Et sur place, une grande surprise vous attend*".

Sensibiliser et contrer le problème

Pour Bo Zhao, de l'Université de Washington, la première étape pour s'attaquer à ce problème est de reconnaître la menace. Dans un récent article, le chercheur détaille comment il a pu créer avec son équipe leurs propres images satellites générées par l'IA. Ainsi qu'il le détaille dans The Verge, le but était alors de "*démystifier l'idée que les images satellites sont d'une fiabilité absolue*" et de "*sensibiliser le public à l'influence potentielle de la géographie deepfake*". Selon lui, son article est en effet probablement le premier à aborder le sujet de ces "faux" dans ce domaine.

Dans le cadre de leur étude, Zhao et ses collègues ont également créé un **logiciel de détection capable de repérer les contrefaçons satellitaires** en fonction de caractéristiques telles que la texture, le contraste et la couleur. Ils soulignent en revanche qu'un tel outil aurait besoin de mises à jour constantes pour suivre les améliorations du deepfake.

Source
20 minutes.fr
Laure Beaudonnet
23 mars 2021

12. Guerre en Ukraine : qu'est-ce que le deepfake de Zelensky laisse craindre pour la suite du conflit ?

Une fausse vidéo de Volodymyr Zelensky le montrait appelant les Ukrainiens à rendre les armes.

C'est une première en temps de guerre. Un deepfake du président ukrainien Volodymyr Zelensky où il exhorte sa population à « rendre les armes » s'est retrouvé la semaine dernière sur le site de la chaîne d'information Ukraine 24 qui a été piraté, sur Facebook, Youtube, Telegram et sur le réseau social russe VKontakte. Pour rappel, un deepfake est une vidéo ou un enregistrement audio qui peut faire dire tout et n'importe quoi à n'importe qui grâce au deep learning. À l'aide de réseaux de neurones, toute l'image est re-générée à partir de matériaux qui n'existaient pas et c'est ainsi que la semaine dernière, le monde a vu naître cette fausse vidéo plutôt réaliste où Volodymyr Zelensky annonce sa reddition à l'invasion russe.

Jusqu'ici, il s'agissait de preuve de concept comme la vidéo de Barack Obama dans laquelle il traitait Donald Trump de « deep shit » ou, plus récemment, celle de Tom Cruise où on le voyait faire un tour de magie. Sans compter l'univers du porno, un certain nombre de deepfakes ont fait irruption dans la sphère politique américaine ces deux dernières années. On pense notamment à la vidéo d'un discours de la présidente de la chambre de représentants Nancy Pelosi accusée d'être ivre par des

proches de Donald Trump. Le résultat était toutefois assez grossier. Si le deepfake de Zelensky a rapidement été supprimé, il ne présage rien de bon pour la suite du conflit (et le futur des deepfakes).

Jeter un discrédit sur Zelensky

« On a affaire à une opération et qui dit opération, dit planification, dit équipe en charge de certaines tâches. Cela fait partie de la palette d'action de la Russie », pointe Julien Nocetti, chercheur à Géopolitique de la datasphère (Géode) pour qui « il paraît plausible que le gouvernement russe en soit à l'origine compte tenu des objectifs de guerre contre la personne de Zelensky ». Mais quel intérêt de diffuser une telle vidéo du président ukrainien ? Il s'agit « avant tout de semer le doute et de jeter un discrédit sur Zelensky qui, jusqu'à présent, est très populaire et très respecté en Ukraine et en Occident », poursuit-il. Le montrer incitant la population rendre les armes a pour vocation de casser les liens entre le président et son peuple. « C'est précisément ce que recherche la Russie dans ses objectifs de guerre, casser ce lien qui s'est vraiment renforcé à la faveur du conflit », note Julien Nocetti.

Si le deepfake est loin d'être parfait, la voix n'est pas la même et l'image peut attirer l'attention, il monte toutefois en gamme comparé à celui de Nancy Pelosi. « C'est quand même largement suffisant pour convaincre un auditoire assez large », observe le spécialiste. L'idée n'est pas d'interpeller l'œil aguerri d'une poignée de geeks mais de s'adresser aux masses qui n'auront pas conscience qu'il s'agit d'un fake. Les deepfakes font partie de l'arsenal russe parce que leurs potentialités se sont accrues et ils permettent de diffuser plus massivement des contenus peu ou prou réalistes. Parfait pour manipuler les esprits.

La menace nucléaire

Et ce n'est qu'un avant-goût. Il n'est pas exclu qu'un deepfake de Joe Biden ou d'Emmanuel Macron puisse créer une escalade dans le conflit. « Toute la doctrine et toute la pratique de la politique étrangère russe consiste à se situer à égalité avec les États-Unis, note Julien Nocetti. Des deepfakes qui suggéreraient des actions de Joe Biden paraissent réalistes ». Imaginons une fausse vidéo dans laquelle le président français ou son homologue américain agiterait la menace nucléaire, les conséquences pourraient être inquiétantes. « Ce serait peut-être aussi un signe de faiblesse de la part de la Russie, nuance-t-il. Cela montre une forme de désespoir du Kremlin sur le théâtre ukrainien ; que la fuite en avant est la seule piste qui permettrait à la Russie de tirer son épingle du jeu. Ce serait une escalade extrêmement dangereuse parce que ce facteur nucléaire est évoqué à tout bout de champs depuis le début du conflit et on ne maîtrise pas les conséquences ».

Au stade de l'expérimentation il y a peu de temps, les deepfakes se transforment peu à peu en armes de communication. On est loin d'avoir vu tout ce que cette technologie nous réserve et ce n'est pas rassurant.