Deepfakes 2.0: La suite est encore plus terrifiante

Shelly Palmer - Août 2019 https://www.shellypalmer.com/2019/08/deepfakes-2-0-sequel-even-scarier/amp/

[Deepfake, hypertrucage, ou permutation intelligente de visages, est une technique de synthèse d'images basée sur l'intelligence artificielle. Elle sert principalement à superposer des images et des vidéos existantes sur d'autres images et/ou vidéos NdT]

À la fin de l'année dernière, j'ai écrit un article au sujet des hypertrucages et de la façon dont, pendant près de 200 000 ans, nous nous sommes fiés à nos yeux et à nos oreilles pour séparer la vérité des mensonges et les faits de la fiction. Même si nous mettons de côté la montée des fausses nouvelles [fake news Ndt], la technologie est sur le point de rendre cela impossible, nous ne pourrons plus savoir si ce que nous voyons et entendons est réel ou faux.

Cela fait moins d'un an que j'ai écrit cet article, et la technologie a évolué dans un sens qui fait passer les contrefaçons de l'an dernier pour des fakes primitives. Voici un tour d'horizon des avancées et des utilisations réalisés au cours de la dernière année.

Déjà illégal: DeepNudes [DeepNudes est une application d'hypertrucage NdT]

La pornodivulgation [divulgation d'un média à caractère sexuel, comme une vidéo ou une photo, produit avec ou sans le consentement d'une personne, afin de lui nuire, NdT] a été interdite dans de nombreuses juridictions à travers le monde. En juillet, la Virginie est devenue l'un des premiers États à interdire le partage de pornographie générée par ordinateur, ou deepfakes. Pour ce faire, elle a modifié une loi existante qui criminalise la pornodivulgation, en précisant que cette catégorie comprend désormais le matériel " faussement créé ". (Image ci-dessous).

Nudité fake

Terrifiant : DNC Deepfake

Le Parti démocrate a hypertruqué son propre président pour mettre en lumière les préoccupations de 2020.

Début août, au DEF CON - l'un des plus grands congrès mondiaux pour hackers - les participants ont appris que le président du DNC [Comité national démocrate, NdT], n'a pas pu assister à la présentation du DNC.

Au lieu de cela,Tom Perez a "Skypé" et chatté. Sauf qu'il ne l'a pas fait. Vous savez vous, à quoi ressemble la voix de Perez ? Les personnes présentes non plus.

Le DNC s'est hypertruqué tout seul

Bob Lord, le chef de la sécurité du DNC, a dirigé l'exercice (avec l'aide d'experts dans le domaine de l'IA) pour démontrer l'art du possible. C'est ce qui nous attend jusqu'aux élections de 2020. La vidéo parle d'elle-même. S'il vous plaît, regardez là. [IA: intelligence artificielle NdT]

Amusant: "Réparer" Le Roi Lion

Le Roi Lion "réparé"

Il est important de se rappeler que les deepfakes ne sont pas tous malveillants. Les puristes de Disney n'ayant pas aimé l'infographie (époustouflante) de The Lion King cet été, ils ont "réparé" la bande-annonce en fusionnant les visages des personnages de la version originale (animée) et les animaux de la nouvelle version. Les images qui en résultent sont si bien faites qu'il vous faut une seconde pour réaliser ce que vous regardez.

Quelque part entre les deux : FaceApp

Vieilli par FaceApp

Au cas où vous auriez vécu sur une île déserte cet été, ou simplement arrêté Internet, FaceApp est une application sur Android et iOS qui utilise l'IA pour "générer des transformations très réalistes de visages à partir de photos". Fondamentalement, FaceApp met le pouvoir des deepfakes à votre portée, ce qui fait qu'il est facile de faire paraître une personne sur une photo plus âgée, plus jeune, ou de la faire changer de sexe.

Sortons de cette petite controverse - Fast Company a découvert que chaque photo téléchargée sur l'application a été envoyée sur les serveurs cloud de FaceApp (ce que vous avez accepté lorsque vous avez coché la case des termes et conditions de l'application, mais clairement sans les lire) - la technologie ici est remarquable. FaceApp propose ses effets photo depuis début 2017, mais les résultats se sont nettement améliorés depuis son lancement.

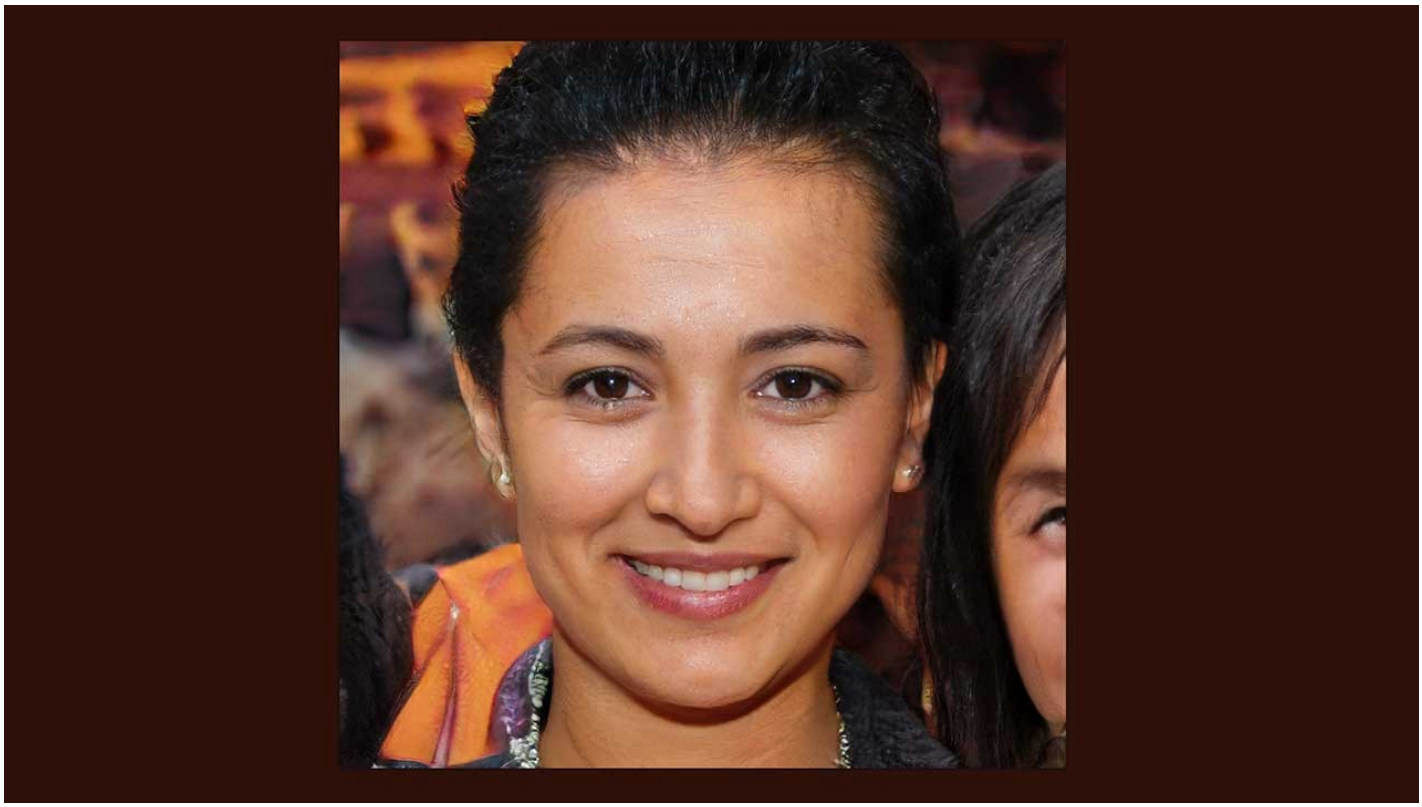
Bien que FaceApp ait des fonctionnalités limitées, c'est l'un des créateurs d'hypertrucage parmi les plus faciles et les plus populaires à la disposition du grand public. Sa vitesse et ses résultats impressionnants le rendent amusant... et dangereux.

Qu'est-ce que tout cela veut dire ?

La technologie derrière les deepfakes était déjà facile à trouver et à utiliser l'automne dernier, mais elle l'est encore plus aujourd'hui. Imaginez où on en sera l'été prochain, au moment des élections de 2020, dans quatre mois.

Elizabeth Warren a-t-elle vraiment dit ça ? La vidéo de Bernie Sanders dans cette publicité offensive était-elle réelle, ou était-ce un hypertrucage? Comme les publicités politiques trafiquent déjà les dires des candidats et manipulent la vérité pour obtenir l'extrait sonore parfait, pouvez-vous croire quoi que ce soit de ce que vous entendez alors que tout cela peut être fabriqué sur n'importe quel ordinateur portable que vous pouvez trouver chez Best Buy ?

Et ça empire



Une image créée par thispersondoesnotexist.com

Alors que les informations, émission après émission, ont fait état d'allégations de manipulation possible des médias sociaux avant les élections de 2016, quelqu'un est-il prêt pour la suite des événements ? Imaginons rapidement comment un système assisté par IA aggravera le problème.

Allez sur le site ThisPersonDoesNotExist.com. Vous voyez ce visage ? Cette photo réaliste ? Eh bien ce n'est pas une vraie personne. Le site " présente une synthèse d'images humaines entièrement automatisée en générant à l'infini des images qui ressemblent à des portraits de visages humains".

Rigolo ! N'est-ce pas ? Seulement jusqu'à ce que vous commenciez à réfléchir aux conséquences possibles. Combien de travail faudrait-il pour construire une histoire et une vie entière pour cette personne ? Il suffit d'utiliser l'intelligence artificielle pour générer le nom et l'histoire de la personne (peut-être en faire une nécrologie pour qu'elle puisse trouver ses parents vivants sur les médias sociaux, à côté d'une mémoire de stockage d'images " réelles ") ?

Créez une adresse e-mail, un compte Facebook et un profil Twitter. Utilisez FaceApp pour vieillir la photo de 30 ans, 60 ans. Photoshop pour la ressemblance de la personne avec ses (vraies) photos de famille. Faites la même chose pour d'autres photos.

Utilisez un logiciel assisté par IA pour générer des tweets et des messages Facebook. Utilisez la méthode procédurale de génération pour faire en sorte que les messages de la personne semblent réels, et non planifiés dans le temps. Faites retweeter les choses par la personne. Répondez aux trucs. Comme des trucs. Interagissez avec le monde.

C'est comme la célèbre arnaque des "Yahoo Boys", mais automatisée et amplifiée au énième degré, alimentée par la technologie et l'apprentissage automatique. En informatique, la génération procédurale (ou le *modèle procédural*) est la création de contenu numérique (niveau de jeu, modèle 3D, dessins 2D, animation, son, musique, histoire, dialogues) à une grande échelle (en grande quantité), de manière automatisée répondant à un ensemble de règles définies par des algorithmes1,2. Le *modèle procédural* s'appuie sur les informations d'un algorithme pour créer. NdT] [les yahoo boys sont des gangs virtuels, se faisant passer pour

des militaires américains qui échangent leurs techniques et exhibent leurs trophées dans forums et réseaux sociaux. NdT]

Construisez un réseau adverse génératif (GAN) pour créer des scanners rétiniens et des empreintes digitales (de la même manière que les faux visages sont générés sur thispersondoesnotexist.com). Piquez quelques numéros de sécurité sociale de l'internet obscur, et... Pouf ! Nous avons créé une personne virtuelle qui est pratiquement impossible à distinguer d'une personne réelle.

Il faudrait pas mal de recherches pour comprendre que cette fabrication n'est pas une personne vivante qui respire.[En intelligence artificielle, les réseaux adverses génératifs sont une classe d'algorithmes d'apprentissage non-supervisé. Ces algorithmes ont été introduits par Goodfellow et al. 2014. Ils permettent de générer des images avec un fort degré de réalisme. NdT]

Utilisez le serveur Chrome pour créer quelques milliers de comptes par jour (je ne vais pas vous dire comment masquer votre connexion réseau, mais vous pourriez facilement faire ça).

Trompez l'Intelligence Artificielle en la qualifiant de "rouge" ou "bleue". Ou, si vous préférez d' "anarchiste" ou "organisateur syndical", ou présentez la personne que vous avez créée comme sympathique à toute cause qui sert votre objectif.

Apprenez à l'IA à suivre les "vrais" comptes (ou les faux comptes que vous avez créés). Les tweets et retweets générés par l'intelligence artificielle ou les messages et partages Facebook vont paraître (et être) réels. Ils ne viendront pas d'êtres humains, c'est tout.

Activez votre réseau d'un demi-million de comptes d'influenceurs . Et attendez jusqu'au 3 novembre 2020 pour voir ce que vous avez réussi à faire.

Vous n'aimerez peut-être pas ce jeu, mais ce n'est ni vous ni moi qui allons y jouer. Ce jeu sera joué par des États-nations qui ont des ressources illimitées. Bouclez votre ceinture de sécurité. Le voyage va être coriace.

Note de l'auteur: Ceci n'est pas un billet sponsorisé. Je suis l'auteur de cet article et il exprime mes propres opinions. Ni moi, ni mon entreprise non plus ne recevons quelque compensation que ce soit pour cet article.