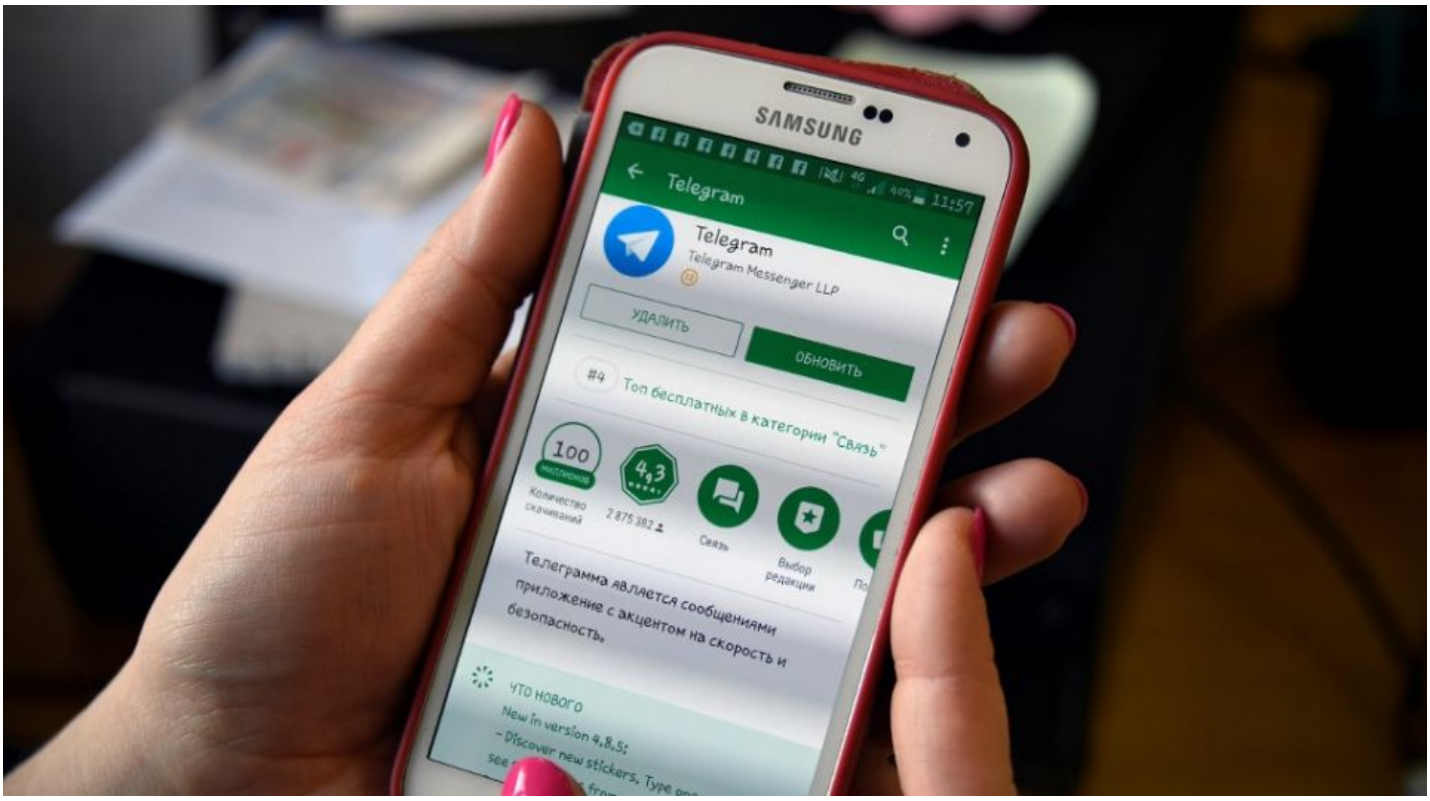
Ces deepfakes de photos de (vraies) femmes nues que des bots génèrent à la chaîne



Des chercheurs en sécurité viennent de découvrir un immense réseau de génération de deepfakes sur Telegram. Au total, plus de 100.000 images de faux nus auraient été générées.

Avec Vincent Claveau

Atlantico.fr : Plus de 100.000 femmes ont vu récemment leurs photos téléchargées par un robot sur l'application Telegram afin que des faux nus soient générés. Pourquoi il y a de plus en plus ce type de pratiques ?

Vincent Claveau : Ce fait divers renoue avec un des premiers usages "popularisant" les deepfake, à savoir le face-swapping sur des vidéo pornographiques (visages de célébrités sur des actrices porno).

Il y a également eu des app (dont DeepNude cité) que tout un chacun pouvait télécharger (avant d'être bannies des app stores) générant une image d'une personne déshabillée à partir d'une photo de la personne habillée.

L'émergence de ces pratiques relèvent de différents facteurs :

- la technologie pour faire des deepfake s'est démocratisée (une personne maîtrisant quelques concepts de base et sachant un peu programmer peut le faire),
- les capacités de calcul pour les générer deviennent plus accessibles (un ordinateur de "gamer" avec une bonne carte GPU suffit)
- la grande disponibilité des données en ligne pour construire ces modèles d'une part, et pour les appliquer d'autre part (dans le cas présent, appliqué à des photos des victimes)

Que faire pour lutter contre ce phénomène ?

Les fake (deepfake ou outils de désinformation/manipulation) sont d'une manière générale une menace pour la société. Il y a des usages différents : désinformation de masse, attaques personnelles/à la réputation, attaques financières (faux ordres de virement) ...

Et d'un point de vue informatique, il y a des technologies différentes pour construire ces fakes.

La réponse ne peut donc être unique.

Technologiquement, pour chacun des types de fake, il y a des outils, plus ou moins aboutis, pour détecter qu'une image ou vidéo a été manipulée ou générée artificiellement et la recherche est active dans ce domaine.

Les plateformes (réseaux sociaux, etc.) ont un rôle absolument essentiel mais délicat. Ce sont les plus à même d'éradiquer la prolifération et le partage de ce type de contenu, mais dans le même temps, il y a un risque à confier à une entreprise privée, rarement de droit français, la régulation de la liberté de parole et d'échange. Cela pose des problèmes autant éthiques que techniques. Mis à part quelques annonces, la situation évolue très peu sur ce front-là.

Il y a nécessité d'éduquer et d'informer sur ces outils de manipulation de contenu et des mésusages qui en sont faits. D'une part pour dissuader en expliquant les dommages qu'ils provoquent, mais aussi, pour tous, pour expliquer que les images ou vidéos qui circulent sur le web ne doivent jamais être considérées comme vraies, sauf à en connaitre la source.

Sur le cas spécifique des "deepnudes", il faut comprendre que l'algo génère une image que lui, en fonction de ses données d'entrainement, considère plausible. Mais évidemment, c'est juste un "dessin très réaliste", pas le vrai corps de la personne (ce qui n'ôte rien au caractère traumatisant de se voir victime d'un photomontage humiliant).

La dernière réponse est évidemment juridique. Ce n'est pas mon domaine, mais les diffamations, injures, atteintes à l'honneur, etc. sont sanctionnables et la diffusion sur internet est, il me semble, vue comme circonstance aggravante.

Comment protéger les populations les plus fragiles ?

On peut tous être victimes, soit individuellement, soit collectivement (en tant que société), de ce type de manipulation.

Évidemment, le sexisme exacerbé de certaines communautés en ligne n'arrange rien à cela, et les femmes en font les frais.

Les conseils de bon sens sont de ne pas laisser d'informations personnelles en ligne si on ne veut pas risquer de les voir utilisées à de mauvaises fins. Plus facile à dire qu'à faire, notamment si on y est obligé pour des raisons professionnelles par exemple (CV en ligne, page web, annuaire d'entreprise...).

Pour le reste, voir les réponses précédentes sur la lutte.

Comment savoir si nous avons été, à notre insu, victime de deep fake ?

C'est très difficile, voire impossible en général. Si le fake est accessible sur internet, il est parfois possible de le trouver en utilisant des moteurs de recherche par image similaire, mais dans le cas de visages, les moteurs grand public n'y arriveront sans doute pas.