

Deepfake : enjeux techniques, juridiques et éthiques

Article rédigé par Isabelle Laratte, avocate au barreau de Paris

Les deepfakes représentent un défi majeur pour nos sociétés. Bien que des solutions techniques et juridiques existent, elles n'empêchent pas le phénomène de prospérer.

Depuis 2023, le « Deepfake », l'enfant terrible de l'intelligence artificielle, s'invite sur tous les réseaux, plateaux, débats publics, et législations. De Hollywood au Dark Web, cette technologie crée événements et scandales, n'épargnant aucun secteur. Miroir déformant, reflet trompeur de notre réalité, source de manipulation, de chantage et d'extorsion, elle nous impacte tous, enfants, citoyens, personnalités publiques, célébrités, sociétés, jusqu'aux démocraties, modifiant même notre rapport à l'image et au réel.

Définition et périmètre

Le deepfake est un néologisme né de la contraction de deep learning (apprentissage profond) et fake (faux). En droit français et européen, on parlera également d'hypertrucage défini par la loi européenne sur l'intelligence artificielle adoptée le 21 mai 2024 comme « une image ou un contenu audio ou vidéo généré ou manipulé par l'IA, présentant une ressemblance avec des personnes, des objets, des lieux, des entités ou événements existants et pouvant être perçu à tort par une personne comme authentiques ou véridiques » (art. 3).

Utilisant des techniques d'apprentissage automatique, le deepfake permet de produire des contenus faux mais réalistes incluant la création ou la modification de contenus vocaux (conversion du texte en parole), la création ou la modification de musique, la génération de visages (facial reenactement) ou d'avatar, le remplacement de visages (face swap) ou des manipulations subtiles du visage, de la voix ou du corps ainsi que la simulation et la modification de modèles 3D.

Sans être illégal par nature, le recours au deepfake pose des problèmes complexes.

Les problématiques : exemples d'utilisations positives et négatives

Dans l'industrie du cinéma, de la télévision et du jeu vidéo, les trucages ont toujours été utilisés en post production dans le cadre des effets spéciaux. L'IA est devenue une technique supplémentaire. En 2020, pendant la période Covid, la série de France 3 « Plus belle la vie » a eu recours au deepfake pour pallier l'absence d'une de ses comédiennes, éloignée du plateau en raison de la pandémie. Le documentaire Welcome to Chechnya (2020) de David France, qui témoigne de la persécution des personnes LGBTQ+ en Tchétchénie a utilisé le deepfake pour protéger l'anonymat de ses sujets tout en conservant leur présence à l'écran. La technologie de l'hypertrucage a permis à l'acteur Val Kilmer, qui a perdu sa voix de "parler".

L'IA améliore les techniques de « de-aging » ou « rajeunissement » utilisées depuis les années 2000 par les studios (Benjamin Button, the Irishman, Captain Marvel). Le dernier Indiana Jones, « Indiana Jones ou le cadran de la destinée » présente un Harrison Ford rajeuni de 40 ans dans une scène de lutte dans un train. L'artiste FKA Twigs a eu recours à cette technologie pour créer son avatar virtuel, capable d'interagir avec son réseau d'abonnés.

Cette accélération du champ des possibles permise par l'IA a suscité une réaction et une inquiétude très forte des acteurs, musiciens et scénaristes américains qui ont refusé l'idée, encore dystopiques, du clonage d'interprétation ou du double numérique, dans le cadre desquels des acteurs seraient payés une somme forfaitaire pour numériser leur image, utilisable ensuite à perpétuité sans autorisation complémentaire. Ces préoccupations ont en grande partie motivé les grandes manifestations d'Hollywood en 2023, à l'occasion

desquelles la Writer Guild Of America, qui regroupe les scénaristes d'Hollywood, unie pour la première fois depuis 1960 avec le syndicat des acteurs d'Hollywood, sont entrés en conflit contre l'Alliance des producteurs de films et de télévision (AMPTP) pour s'opposer avec succès à l'empiètement croissant de l'intelligence artificielle sur leur profession.

Avec l'accessibilité accrue de l'intelligence artificielle, un cap a été nettement franchi. La technologie s'est considérablement démocratisée. Les logiciels permettant de réaliser des trucages se multiplient et sont faciles d'accès, comme FakeApp, Deep faceLab sur les applications d'IA générative comme Midjourney, Stable Diffusion, Meta et OpenAI avec Dalle E 2. Cette évolution facilite et accélère la création d'images et de vidéos de plus en plus réalistes avec moins de données.

Les deepfake envahissent tous les secteurs rendant difficile la distinction entre réalité et fiction. En politique, on pense immédiatement au deep fake canular d'Emmanuel Macron ramassant des poubelles, de Barak Obama et Angela Merkel bâtissant des châteaux, ou bien encore du Pape François en doudoune blanche créé d'ailleurs avec Midjourney. En musique, un titre anonyme de Ghoswriter utilisant les voix de Drake et The Weekend générées par l'IA est devenu viral sur les réseaux sociaux avant d'être supprimé à la demande d'Universal Music Group. Comme celui à l'été 2023 du beatmaker nancéen nommé Lnkhey remixant, sans autorisation aucune, au moyen du logiciel RVC, permettant de reproduire n'importe quelle voix à l'identique, une chanson des rappeurs Heuss l'Enfoiré et Gazo avec la voix de la chanteuse Angèle. En janvier 2024, la plateforme X a verrouillé la recherche « Taylor Swift » après la diffusion massive de nus de la star générés par l'IA.

Ces images abondamment partagées sur les réseaux sociaux peuvent générer des millions de vues et créer un préjudice considérable à ceux qu'elles visent.

Le 5 juin 2023, le FBI tirait la sonnette d'alarme relativement à la recrudescence des pratiques d'extorsions et sextorsions, au deepfake suite à la réception d'un nombre accru de rapports de victimes, y compris des enfants mineurs, dont les photos ou les vidéos généralement capturées à partir du compte d'un individu sur les médias sociaux, sur internet ont été modifiées en contenu explicite ou sexuel. Ces photos ou vidéos sont ensuite diffusées publiquement sur les médias sociaux ou sur des sites pornographiques, dans le but de harceler les victimes ou de pratiquer la sextorsion. Or, une fois qu'elles ont circulé, les victimes peuvent être confrontées à des difficultés considérables pour empêcher le partage continu du contenu manipulé ou son retrait de l'internet.

Le secteur des affaires n'est pas épargné comme le montre l'escroquerie impliquant Patrick Hillman, COO de la société de cryptomonnaie Binance. Des pirates qui avaient récupéré des images de son visage sont parvenus à créer un avatar virtuel pour mettre en place une escroquerie à la cotation en bourse lors de plusieurs réunions Zoom avec des représentants de projets cryptos. Le but ? Soutirer de l'argent à des usagers potentiels de portefeuilles numériques.

En avril 2022, Europol a rapporté des utilisations criminelles de la technologie incluant la désinformation, la falsification de preuves, la fraude documentaire et la production de pornographie non consentu. Le 28 juin 2022, le FBI a également signalé une hausse des plaintes concernant les deepfakes utilisés pour obtenir des postes à distance avec accès à des informations confidentielles.

En septembre 2023, un rapport du New York Times alertait le public sur le fait que des escrocs utilisaient à grande échelle des imitations de voix pour inciter les banques à transférer l'argent de leurs clients. En février 2024, une multinationale de Hong Kong a subi une escroquerie de 25,6 millions de dollars impliquant des deepfakes. La police de Hong Kong a révélé que l'escroquerie consistait en une vidéoconférence à plusieurs personnes dont tous les participants, à l'exception de la victime, étaient des créations de deepfake. Et les exemples sont nombreux.

Les institutions publiques sont également ciblées comme en 2023 lorsqu'un deepfake créé par trois collégiens a attribué des propos racistes au dirigeant d'un collègue. La sécurité nationale et la démocratie sont également menacées par cette technologie qui est une source de désinformation et d'influence politique et électorale. En 2022 un deepfake grossier diffusé sur la chaîne d'informations Ukraine 24 montrait Volodymyr Zelensky appelant ses soldats à la reddition. En janvier 2024, un Joe Biden beaucoup plus convaincant cette fois appelait les électeurs du New Hampshire avant les primaires et demandait de rentrer chez eux. En février 2024, une fausse séquence de France 24 prétendait qu'Emmanuel Macron avait été victime d'une tentative d'assassinat.

Avec la digitalisation croissante de nos vies, la crédibilité des images est en déclin tandis que leur potentiel de nuisance augmente. Les efforts technologiques et législatifs sont désormais essentiels pour maîtriser cette technologie et préserver l'équilibre bénéfique risque.

Source : https://www.app.asso.fr/propriete-intellectuelle/deepfake-enjeux-techniques-juridiques-et-ethiques.html#_ftn1