Village de la Justice www.village-justice.com

Le deepfake ou l'hypertrucage - Connaitre la règlementation et s'en prémunir. Par Sophie Renaudin, Avocate.

Parution : mercredi 19 février 2025

Adresse de l'article original :
https://www.village-justice.com/articles/deepfake-hypertrucage-connaitre-reglementation-premunir,52460.html

Le mot « *deepfake* » se traduit littéralement part « *hypertrucage* » en français et vient de « *deep learning* » (apprentissage profond) et « *fake* » (faux).

Il s'agit d'une technique qui permet de réaliser, grâce à l'intelligence artificielle (IA), des montages de vidéos, d'images ou de son.

Nous avons pu voir, par exemple un certain nombre de vidéos, comme une vidéo du Président Macron reprenant un titre de la chanteuse Angèle, Madame Le Pen en train de rapper, Barack Obama qualifiant Trump de « *sombre merde* », etc.

Au sommaire de cet article...

Le Règlement européen (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) no 300/2008, (UE) no 167/2013, (UE) no 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle) définit l'hypertrucage comme étant

« *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* ».

Nous constatons d'emblée que la définition de l'hypertrucage ne fait référence qu'aux images, aux audios et aux vidéos, laissant ainsi de côté les écrits.

La pratique de l'hypertrucage n'est pas forcément interdite par les textes. En effet, elle peut être un moyen au service de la liberté d'expression ou de la création satirique, parodique ou artistique.

Il apparaît que notre règlementation n'en est qu'aux prémices en la matière, néanmoins, l'arsenal juridique est en train de se construire petit à petit pour pouvoir encadrer ces pratiques **(I)** et sanctionner les dérives **(II)**.

L'enjeu est également, à l'heure des Fake news, de pouvoir distinguer du contenu généré par IA, d'un contenu réel **(III)**.

I. Tour d'horizon des règles juridiques applicables.

La loi n° 2023-451 du 9 juin 2023 visant à encadrer l'influence commerciale et à lutter contre les dérives des influenceurs sur les réseaux sociaux.

Et notamment l'article 5 de cette Loi qui impose une obligation spécifique d'information pour les influenceurs en matière d'hypertrucage dans deux hypothèses :

Soit l'influenceur diffuse un contenu commercial comprenant des images visant à affiner ou à épaissir la silhouette ou à modifier l'apparence du visage, dans ce cas il doit accompagner sa diffusion de la mention : « *images retouchées* » ;
Soit ce contenu a été créé par l'IA et dans ce cas, l'influenceur doit accompagner la diffusion de son contenu de la mention « *images virtuelles* ».

Le Règlement sur l'intelligence artificielle de 2024.

À titre liminaire, rappelons qu'un Règlement européen est directement applicable dans notre droit interne sans que l'adoption d'une loi de transposition ne soit nécessaire. De sorte que le Règlement européen sur l'intelligence artificielle est applicable en France depuis son entrée en application.

Concernant le chapitre spécifique à l'hypertrucage, ce dernier entrera en application le 2 août 2025 [1]. Celui-ci imposera dès son entrée en application une obligation de transparence pour les fournisseurs et déployeurs d'IA [2].

Concernant les fournisseurs d'IA, ils devront marquer techniquement que le contenu a été produit par l'IA dans un format lisible par machine et l'identifier comme ayant été généré ou manipulé par une IA [3].

Si l'hypertrucage est interactif, ils devront informer les personnes concernées du fait qu'elles interagissent avec une IA « ***sauf si cela ressort clairement du point de vue d'une personne physique normalement informée et raisonnablement attentive et avisée, compte tenu des circonstances et du contexte d'utilisation*** » [4].
Autre exception à cette obligation de transparence, les systèmes d'IA dont la loi autorise l'utilisation à des fins de prévention ou de détection des infractions pénales, d'enquêtes ou de poursuites en la matière, sous réserve de

garanties appropriées pour les droits et libertés des tiers, sauf si ces systèmes sont mis à la disposition du public pour permettre le signalement d'une infraction pénale.

Concernant le déployeur (utilisateur), l'article 50.4 dispose que « *les déployeurs d'un système d'IA qui génère ou manipule des images ou des contenus audio ou vidéo constituant un hypertrucage indiquent que les contenus ont été générés ou manipulés par une IA* ». Il y a donc une véritable obligation de transparence.

Il existe deux exceptions à cette obligation d'information :

L'utilisation est autorisée par la loi à des fins de prévention ou de détection des infractions pénales, d'enquêtes ou de poursuites en la matière ;
Le contenu fait partie d'une œuvre ou d'un programme manifestement artistique, créatif, satirique, fictif ou analogue. Dans ce cas, les obligations de transparence énoncées se limitent à la divulgation de l'existence de tels contenus générés ou manipulés d'une manière appropriée qui n'entrave pas l'affichage ou la jouissance de l'œuvre.

Clairement, compte tenu de l'ampleur du phénomène et de l'ascension fulgurante de l'IA, des lois dédiées devront nécessairement voir le jour pour encadrer davantage ces pratiques.

II. Les sanctions du recours à l'hypertrucage.

Plusieurs leviers légaux existent déjà et peuvent s'appliquer à l'hypertrucage.

L'infraction de montage de l'article 226-8 du Code pénal [5].

> « *Est puni d'un an d'emprisonnement et de 15 000 euros d'amende le fait de porter à la connaissance du public ou d'un tiers, par quelque voie que ce soit, le montage réalisé avec les paroles ou l'image d'une personne sans son consentement, s'il n'apparaît pas à l'évidence qu'il s'agit d'un montage ou s'il n'en est pas expressément fait mention.* ***Est assimilé à l'infraction mentionnée au présent alinéa et puni des mêmes peines le fait de porter à la connaissance du public ou d'un tiers, par quelque voie que ce soit, un contenu visuel ou sonore généré par un traitement algorithmique et représentant l'image ou les paroles d'une personne, sans son consentement, s'il n'apparaît pas à l'évidence qu'il s'agit d'un contenu généré algorithmiquement ou s'il n'en est pas expressément fait mention*** ».

Si l'infraction a été réalisée en utilisant un service de communication au public en ligne, les peines sont portées à deux ans d'emprisonnement et 45.000€ d'amende.

Pour caractériser ce délit, il faut que le montage généré par IA l'ait été sans le consentement de la personne. Par ailleurs, il ne doit pas apparaitre à l'évidence qu'il s'agit d'un contenu généré par l'IA ou qu'il n'en soit pas fait expressément mention.

Ainsi, un contenu généré sans le consentement de la personne mais dont il est fait mention qu'il a été créé via l'IA ou que cela paraisse évident, ne tombe pas sous le coup de cette infraction.

L'infraction de montage à caractère sexuel de l'article 226-8-1 du Code pénal [6].

> « *Est puni de deux ans d'emprisonnement et de 60 000 euros d'amende le fait de porter à la connaissance du public ou d'un tiers, par quelque voie que ce soit, un montage à caractère sexuel réalisé avec les paroles ou l'image d'une personne, sans son consentement. Est assimilé à l'infraction mentionnée au présent alinéa et puni des mêmes peines le fait de porter à la connaissance du public ou d'un tiers, par quelque voie que ce soit, un contenu visuel ou sonore à caractère sexuel généré par un traitement algorithmique et reproduisant l'image ou les paroles d'une personne, sans son consentement* ».

Les peines sont portées à trois ans d'emprisonnement et à 75 000 euros d'amende lorsque la publication du montage ou du contenu généré par un traitement algorithmique a été réalisée en utilisant un service de communication au public en ligne.

Pour ce délit, la mention du caractère virtuel du contenu ne suffit pas à exonérer son auteur de sa responsabilité pénale.

Ainsi, le délit sera constitué dès lors que le contenu est à caractère sexuel, généré par l'IA et qu'il ait été produit sans le consentement de la personne.

Étant précisé que la tentative de ces deux délits est réprimée de la même manière. Par ailleurs, les personnes morales peuvent également être condamnées et sanctionnées.

L'atteinte au droit à l'image.

L'article 9 du Code civil dispose que chacun a droit à sa vie privée.
Sur la base de cet article, il est possible d'engager la responsabilité civile d'une personne qui utiliserait l'image d'une personne sans son consentement.
L'utilisation de l'image d'une personne, même modifiée par l'IA, pourrait tomber sous le coup de cet article. Pour le moment, la jurisprudence ne s'est pas positionnée sur la question.

L'atteinte aux données personnelles des personnes physiques.

L'hypertrucage peut avoir recours à la voix ou encore l'image des personnes. Selon la règlementation en vigueur, en leur qualité d'attributs de la personnalité, ces données sont des données à caractère personnel [7].

L'utilisation de telles données pourrait être sanctionnée sur la base du RGPD dans la mesure où :

La personne qui a créé le contenu est un responsable de traitement au sens du RGPD ;
L'utilisation n'est pas faite à des fins strictement personnelles ou privées ;
Le traitement est illicite (ne repose sur aucune base légale listée à l'article 6 du RGPD [8]).

Nous noterons que, contrairement au Règlement sur l'IA, le RGPD s'appliquera même si le contenu relève de la liberté d'expression ou artistique. En effet, si le RGPD prévoit que certains articles ne soient pas applicables en cas de tels contenus, le traitement de données devra néanmoins obligatoirement reposer sur une base légale pour pouvoir être licite. Cela pourrait donc être incompatible avec l'hypertrucage.

La CNIL s'est d'ailleurs positionnée sur le sujet en rappelant que les systèmes d'IA comprenant des données personnelles devaient disposer d'une base légale (consentement de la personne par exemple) [9].

Les sanctions du RGPD sont les suivantes :

20 000 000€ d'amende ou
4% du chiffre d'affaires mondial consolidé.

Le Règlement sur l'IA.

Le Règlement précise que le non-respect des règles en matière de transparence pourra être sanctionné de 15 000 000 € pour une personne physique ou 3% du chiffre d'affaires mondial consolidé si la personne est une personne morale.

La contrefaçon.

Le Code de la propriété intellectuelle pourrait s'appliquer en cas d'utilisation par l'IA d'un contenu protégé par le droit d'auteur, voire par un titre de propriété industrielle comme une marque ou des dessins et modèles et sanctionner de telles reproductions à l'identique ou par similarité. Sauf à ce que le contenu tombe sous le coup des exceptions prévues par le Code (Exemple : parodie, pastiche, caricature, etc.). Une étude au cas par cas devra être menée pour déterminer si le contenu tombe sous le coup du délit de contrefaçon.

L'escroquerie, abus de confiance, etc.

Si l'hypertrucage est utilisé comme moyen pour recevoir des fonds, valeurs ou autre, comme c'est le cas dans certaines escroqueries comme par exemple les arnaques au Président, la personne pourrait se faire sanctionner au titre des délits d'escroquerie, abus de confiance, etc.

III. Les outils permettant de détecter qu'un contenu est généré par l'IA.

Il est possible de détecter qu'un contenu a été généré par IA grâce à certains signaux :

Positionnement inhabituel du visage ;
Positionnement inhabituel des mains ou des pieds ;
Expressions maladroites ;
Mouvements du visage ou du corps non naturels ;
Couleurs incohérentes sur l'ensemble de la vidéo ;
Son mal adapté ou incohérent ;
Clignement des yeux non naturel ;
Fautes d'orthographe et des phrases grammaticalement incorrectes ;
Déroulement des phrases qui ne semble pas naturel ;
Adresses électroniques suspectes ;
Messages hors sujet ou non pertinents.

Et dans tous les cas, il faut toujours vérifier la source du *contenu* !

Avec le développement de l'IA, cette dernière deviendra de plus en plus performante et il deviendra compliqué pour les personnes physiques de savoir distinguer les hypertrucages du vrai.

Des sociétés du secteur ont également développé certains outils pour détecter les hypertrucages :

Microsoft a créé un outil de détection fonctionnant avec une intelligence artificielle pour analyser les vidéos et les photos. Il indique avec précision le taux de probabilité de manipulation.
Opération Minerva : cet outil utilise le catalogage des sites *deepfakes* connus et des empreintes numériques. Les vidéos sont comparées à ce catalogue pour détecter les éventuelles modifications.
Outil de détection Sensity : la plateforme propose un logiciel qui détecte les médias falsifiés. Son fonctionnement est le même que celui des antivirus. En cas de contenus falsifiés, elle envoie une alerte par courrier électronique.
Mozilla lance l'extension Deep Fake Detector qui permet de détecter les contenus générés par IA.

Quoiqu'il en soit, de nombreux vides juridiques entourent encore l'IA et l'hypertrucage. Nos représentants doivent d'ores et déjà s'emparer de cette problématique pour proposer un encadrement clair de ces pratiques. Compte tenu de la vitesse avec laquelle l'IA évolue, il est urgent d'intervenir !

Sophie Renaudin Avocate au Barreau de Saint-Denis de la Réunion [->avocat@srenaudin-avocat.com]
https://www.srenaudin-avocat.com/

[1] Article 113 du Règlement.

[2] Article 3.4 du Règlement : un déployeur est « *une personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant sous sa propre autorité un système d'IA sauf lorsque ce système est utilisé dans le cadre d'une activité personnelle à caractère non professionnel* ». En d'autres termes, ceux qui utilisent l'IA.

[3] Article 50.2.

[4] Article 50.1.

[5] Article modifié par la Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique.

[6] Article créé par la Loi n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique.

[7] Article 4 RGPD : « *la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement ; lorsque les finalités et les moyens de ce traitement sont déterminés par le droit de l'Union ou le droit d'un État membre, le responsable du traitement peut être désigné ou les critères spécifiques applicables à sa désignation peuvent être prévus par le droit de l'Union ou par le droit d'un État membre* ».

[8] Consentement, contrat ou mesures précontractuelles, obligation légale, sauvegarde des intérêts vitaux, exécution d'une mission de service public, intérêt légitime.

[9] Sur ce sujet, voir les recommandations de la CNIL https://www.cnil.fr/fr/developpement-des-systemes-dia-les-recommandations-de-la-cnil-pour-respecter-le-rgpd

L'auteur déclare ne pas avoir utilisé l'IA générative pour la rédaction de cet article.