Hypertrucage (deepfake) : comment se protéger et signaler les contenus illicites ?

03 février 2026

De plus en plus réalistes, les vidéos, images et sons truqués peuvent porter atteinte à la vie privée et à la réputation. Les créer ou les partager n'est pas sans risques et peut engager la responsabilité de leurs auteurs. Quels enjeux, quels risques et quels bons réflexes à adopter ?

L'hypertrucage : qu'est-ce que c'est ?

Le mot *deepfake* (ou hypertrucage) vient de la contraction de deux mots anglais : *deep learning*, « apprentissage profond » et *fake*, « faux ».

Un hypertrucage est ainsi un contenu audio, photo ou vidéo créé ou modifié grâce à des techniques d'intelligence artificielle. Elles permettent d'**imiter une voix, un visage ou un mouvement** avec un réalisme de plus en plus difficile à distinguer d'un contenu authentique. Dans certains cas, ces hypertrucages peuvent se répandre très vite en ligne, notamment sur les réseaux sociaux.

Ils peuvent notamment être utilisés pour :

- remplacer un visage dans une vidéo (*face swapping*) ;
- imiter une voix dans un message audio : les paroles prononcées sont modifiées et, dans le cas de contenus vidéos les mouvements des lèvres sont modifiés et resynchronisés (*lip synching*) ;
- créer des vidéos ou images entièrement ou partiellement fictives.

Quels sont les risques pour vous et quels sont les bons réflexes à adopter ?

Désormais, de nombreuses applications et logiciels permettent de réaliser des hypertrucages. Ceux-ci sont de plus en plus élaborés et leur utilisation largement accessible au grand public. La diffusion d'hypertrucages a donc considérablement augmenté, tout comme les nombreux risques liés.

Les utilisations de l'**intelligence artificielle** à des fins malveillantes sont punies par la loi.

Atteinte à la vie privée et usurpation d'identité

Une photo, une vidéo ou un enregistrement vocal publié en ligne peut techniquement être détourné pour créer un hypertrucage, sans l'accord de la personne. L'image ou la voix peuvent alors être utilisées pour la mettre dans des situations fausses, gênantes ou préjudiciables, portant atteinte à sa vie privée et à sa réputation.

Ils peuvent ainsi être utilisés dans le cadre de phénomènes de « sextorsion » (extorsion ayant pour support le chantage lié à la diffusion de contenus à caractère sexuel).

Pour rappel, le montage réalisé avec l'image d'une personne sans son consentement peut être puni d'**un an emprisonnement** et de **15 000 € d'amende** (article 226-8 du code pénal). Si le montage est à **caractère sexuel** et/ou **généré par un service de communication en ligne** (par exemple un réseau social), **les peines peuvent être plus lourdes**.

Cyberharcèlement

Les hypertrucages peuvent servir à humilier, menacer ou faire chanter une personne. Des images ou vidéos truquées, parfois à caractère choquant, peuvent être diffusées pour nuire à une victime, en particulier les jeunes, avec des conséquences importantes sur leur bien-être.

Pour rappel, le harcèlement, en ligne ou non, est puni de **2 ans d'emprisonnement** et **30 000 € d'amende** (article 222-33-2-2 et suivants du code pénal). Les peines encourues sont plus lourdes si la victime est un conjoint, un ex-conjoint ou une personne vulnérable (par exemple un mineur).

> Pour en savoir plus : cyberviolences et cyberharcèlement : que faire ?

Fraudes et escroqueries

La reproduction fidèle d'un visage, d'une voix ou d'un comportement peut servir à usurper l'identité d'une personne : ouverture ou modification de comptes en ligne, fraude auprès de services, escroquerie visant des proches ou des collaborateurs, etc.

L'escroquerie, qui consiste à tromper une personne en vue de se voir remettre de l'argent, un bien ou un service est punie de **5 ans d'emprisonnement** et de **375 000 € d'amende** (article 313-1 du code pénal).

> Pour en savoir plus : *spam*, hameçonnage, arnaques : signaler pour agir

Désinformation et manipulation

Les hypertrucages peuvent être créés pour tromper le public. Des vidéos ou des audios falsifiés de personnalités peuvent être diffusés pour manipuler l'opinion, nuire à une réputation ou propager de fausses informations.

La fabrication ou la diffusion de désinformation est, dans certains cas, **punie par la loi** (article 27 de la loi du 29 juillet 1881, article 322-14 du code pénal, article L97 du code électoral).

Génération de contenus haineux ou pédopornographiques

Les hypertrucages peuvent permettre de générer des contenus illégaux comme des contenus haineux (racistes ou xénophobes) ou pédopornographiques.

Certains services gratuits ou payants utilisant l'IA prétendent « déshabiller » des personnes réelles, notamment à partir de clichés de personnes mineurs : **ne les utilisez pas, ne partagez pas les contenus.**

Par ailleurs, certaines pratiques comme, par exemple la publication de photos et de vidéos de mineurs sur les réseaux sociaux (ou **sharenting**) peuvent participer à nourrir les hypertrucages.

Pour rappel :

- le fait de produire, de détenir, d'enregistrer, de transmettre ou de consulter habituellement des contenus pédopornographiques est puni de **5 ans d'emprisonnement** et de **75 000 € d'amende**, y compris si ces contenus ont été produits par des outils d'intelligence artificielle (article 227-23 du code pénal).
- les contenus haineux visant des personnes en raison de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une nation, une prétendue race ou une religion déterminée est punie de l'**amende** prévue pour les contraventions de la 5^{e} classe (jusqu'à 1 500 € ou 3 000 € en cas de récidive) (article R. 625-7 du code pénal).

Comment se protéger et signaler les contenus illicites ?

Protéger son image et ses données

Une fois en ligne, vos données peuvent être copiées et détournées. Veillez donc à :

- Éviter de diffuser de manière incontrôlée (c'est-à-dire accessible à tous) des images de vous ou de vos proches notamment sur les réseaux sociaux : privilégiez le partage par messagerie privée instantanée sécurisée ou des partages éphémères avec une audience réduite.
- Éviter que vos photos de profils puissent être utilisées dans des hypertrucages, par exemple, en prenant une photo de loin, en la pixellisant, en la floutant (par exemple avec l'application de la CNIL FantomApp), en utilisant des filtres ou en masquant une partie de votre visage.
- Faites le ménage régulièrement dans vos photos et vidéos, notamment les plus anciennes.

Apprendre à reconnaître un hypertrucage

- Vous pouvez apprendre à reconnaître un hypertrucage même si cela n'est pas toujours facile. Certains indices peuvent aider : l'image peut paraître pixélisée ou floue, les yeux peuvent bouger de manière non naturelle, la bouche peut sembler déformée ou mal synchronisée, la lumière et les ombres sur le visage peuvent sembler anormales.
- Méfier vous aussi des contenus surprenants : une vidéo ou un message choquant ou trop spectaculaire peut être un hypertrucage. Prenez le temps de vérifier la source avant de croire ou de partager. Consulter également les sites spécialisés dans la vérification de faits (*fact checking*).

Réagir à la publication d'un hypertrucage

1 Vérifier avant d'agir

Si vous soupçonnez qu'une image, une vidéo ou un **fichier** audio est un hypertrucage, ne la partagez pas afin de ne pas aggraver le préjudice subi par la personne.

De même, en cas de demande inhabituelle (argent, informations personnelles, consignes urgentes, etc.), même semblant venir d'un proche ou d'un collègue, assurez-vous par un autre moyen (appel, message direct) de son authenticité.

2 Agir rapidement en cas de problème

- Utilisez d'abord les outils de signalement des plateformes pour contenus trompeurs ou abusifs.
- Si vous estimez que l'hypertrucage a été utilisé dans le but de **commettre un délit** ou de **porter atteinte à votre réputation**, conservez des preuves (captures d'écran des images truquées, liens) et, si nécessaire, déposez une plainte auprès de la police nationale ou d'une brigade de gendarmerie.
- Pour signaler des **contenus illicites et graves sur Internet** (harcèlement, menaces, images sexuelles de mineurs, escroqueries) : PHAROS
- En cas d'**arnaque** ou de **cyberharcèlement**, plusieurs guichets sont disponibles pour vous aider (signalement, dépôt de plainte, aide aux victimes, etc.). Vous trouverez des informations à ce sujet, notamment dans l'application de la CNIL FantomApp.

- Si vos **données personnelles** (y compris votre visage, votre voix, votre nom et prénom) **sont utilisées sans votre accord**, vous pouvez également adresser une plainte à la CNIL.

Attention : une plainte auprès de la CNIL ne remplace pas une plainte auprès de la police ou gendarmerie. La CNIL ne peut pas obtenir pour vous des dommages et intérêts ou sanctionner des auteurs de crimes ou délits.

Que fait la CNIL ?

La CNIL agit pour **protéger votre vie privée et vos données personnelles** face aux risques liés aux hypertrucages. Ses missions principales sont :

- **Informer et conseiller** le grand public sur les hypertrucages et leurs risques.
- **Encadrer les technologies** et promouvoir des pratiques responsables, par exemple à travers :
  - Le projet GenFakes du Pôle d'Expertise de la Régulation Numérique (PEReN) en partenariat avec le laboratoire d'innovation numérique de la CNIL (LINC) qui explore la génération et la détection de deepfakes.
  - La publication de contenus sur le tatouage numérique ou watermarking et d'autres techniques de détection de contenus artificiels sur le site du LINC.
  - Le suivi des travaux de recherche liés, tels que ceux menés dans le cadre du PEPR (Programme et équipement prioritaire de recherche) COMPROMIS sur la sécurité des données multimédia.

- **Participer aux travaux européens**, comme le code de bonnes pratiques pour les contenus générés par IA, pour renforcer la lutte contre les contenus trompeurs et malveillants.
- **Veiller au respect des règles sur la protection des données personnelles par les entreprises, qu'elle peut contrôler et sanctionner.**

L'image d'illustration de cet article a été générée par un système d'intelligence artificielle afin de montrer le réalisme des visages créés.

Pour approfondir

- Partage de photos et vidéos de votre enfant sur les réseaux sociaux : quels sont les risques ?
- Tous les contenus de la CNIL sur l'intelligence artificielle (IA)
- Panorama et perspectives pour les solutions de détection de contenus artificiels [1/2]
- Partenariat avec le PEReN : les risques de l'intelligence artificielle dans le cadre des hyper-trucages
- Mineurs : floutez votre photo avec FantomApp (en ligne ou sur l'application)