Deepfakes (hypertrucage) et arnaques à l'intelligence artificielle : comment les repérer et vous en protéger ?

Les deepfakes (hypertrucages) utilisent l'IA pour imiter visages ou voix, favorisant usurpation, fraude, cyberharcèlement et désinformation. Des indices visuels ou sonores et l'urgence alertent. Protéger ses données, vérifier par un autre canal, alerter la banque, conserver des preuves, signaler via Pharos ou porter plainte sont conseillés.

Écrit le 05/11/2026

L'intelligence artificielle (IA) générative permet aujourd'hui de créer des images, des vidéos ou des voix très réalistes. Si ces technologies offrent de nouvelles opportunités, elles sont aussi détournées par des cybercriminels pour produire des deepfakes (hypertrucages) et mettre en scène de fausses situations ou de faux interlocuteurs. Comment fonctionnent ces arnaques ? Comment les détecter pour s'en protéger ? Que dit la loi ?

Qu'est-ce qu'un *deepfake* et comment est-il utilisé pour vous arnaquer ?

Selon la définition de la CNIL, le mot ***deepfake*** (ou *hypertrucage*) vient de la contraction de *deep learning* (« apprentissage profond ») et *fake* (« faux »). Un hypertrucage désigne un **contenu audio, photo ou vidéo créé ou modifié grâce à des techniques d'intelligence artificielle pour imiter** une voix, un visage ou un mouvement avec un réalisme parfois difficile à distinguer d'un contenu authentique.

Ces techniques permettent par exemple de :

- **remplacer un visage dans une vidéo** (*face swapping*) ;
- **imiter une voix dans un message audio**, en modifiant les paroles et, dans le cas de vidéos, les mouvements des lèvres pour les resynchroniser (*lip synching*) ;
- **créer des vidéos ou des images entièrement ou partiellement fictives**, présentant une personne dans une situation qu'elle n'a jamais vécue.

Une simple photo, vidéo ou note vocale publiée en ligne peut donc être détournée pour fabriquer un hypertrucage sans l'accord de la personne concernée, avec des impacts potentiels sur sa vie privée, sa réputation ou sa sécurité. La reproduction fidèle d'un visage, d'une voix ou d'un comportement peut ensuite servir à usurper

une identité, notamment dans des fraudes et escroqueries visant des proches ou des collaborateurs, ou pour modifier des comptes en ligne.

Dans ce contexte, plusieurs scénarios d'arnaques à l'IA se développent.

Les scénarios d'arnaques à l'IA les plus fréquents

Les faux placements financiers
Des hypertrucages peuvent mettre en scène des personnalités publiques (dirigeants, influenceurs, responsables politiques, journalistes) dans des vidéos falsifiées où elles semblent recommander un investissement « sans risque » ou des gains rapides, par exemple en cryptomonnaies.

Certaines personnes également essayer de vous faire investir de l'argent dans des projets qui n'existent pas.

Le clonage vocal : l'arnaque au faux proche

À partir d'un court extrait de la voix de l'un de vos proches (par exemple une vidéo ou une note vocale publiée sur un réseau social), un escroc peut utiliser l'IA pour imiter cette voix avec un réalisme très convaincant.

La reproduction fidèle de la voix permet alors de se faire passer pour un membre de la famille ou un ami. Le scénario est souvent le même : vous recevez un appel ou un message audio évoquant un accident, une difficulté grave ou un besoin urgent de liquidités, et on vous demande un virement immédiat.

Ce type d'arnaque joue sur l'émotion et l'urgence, ce qui rend plus difficile la prise de recul et la vérification. En cas de demande inhabituelle ou urgente (argent, données personnelles), pensez à vérifier par un autre moyen (appel sur le numéro habituel, message direct) l'identité réelle de la personne avant d'agir.

La « fraude au président » renforcée par l'IA

Dans le milieu professionnel, les hypertrucages peuvent être utilisés pour imiter le visage et la voix d'un dirigeant lors d'une visioconférence ou d'un appel. Un collaborateur reçoit une instruction urgente de virement qui semble émaner de son supérieur hiérarchique, avec une apparence visuelle et sonore très crédible.

Ce type de fraude, déjà connu sous le nom de « fraude au président », est désormais renforcé par l'usage de deepfakes : l'usurpation d'identité ne repose plus seulement sur un mail ou un coup de fil, mais sur une représentation audiovisuelle truquée qui vient conforter le caractère « authentique » de la demande. Là encore, le schéma est toujours le même : sentiment d'urgence, demande inhabituelle concernant un transfert de fonds, pression pour agir sans délai.

Quels sont les risques des *deepfakes* ?

Au-delà des arnaques financières, les hypertrucages peuvent avoir d'autres conséquences graves pour les personnes :

- **Atteinte à la vie privée et usurpation d'identité** : une image ou une voix peuvent être utilisées pour placer une personne dans des situations fausses, gênantes ou préjudiciables, et alimenter des phénomènes de sextorsion (chantage à la diffusion de contenus à caractère sexuel).
- **Cyberharcèlement** : des images ou vidéos truquées, parfois choquantes, peuvent être diffusées pour humilier, menacer ou faire chanter une personne, en particulier les jeunes, avec des impacts importants sur leur bien-être.
- **Désinformation et manipulation** : des vidéos ou audios falsifiés de personnalités peuvent être diffusés pour manipuler l'opinion, nuire à une réputation ou propager de fausses informations, notamment en période électorale.
- **Création de contenus haineux ou pédopornographiques** : certains outils d'IA peuvent servir à générer des contenus illégaux (contenus racistes, xénophobes, pédopornographiques), à partir de photos de personnes réelles, y compris de mineurs.

Ces usages sont susceptibles de constituer des infractions pénales (harcèlement, escroquerie, diffusion de contenus haineux ou pédopornographiques, etc.), avec des peines pouvant aller jusqu'à plusieurs années d'emprisonnement et des amendes importantes.

Comment repérer un contenu généré par l'IA ? Les signaux d'alerte

Les hypertrucages sont de plus en plus sophistiqués, mais la technologie n'est pas infaillible. Certains indices peuvent vous mettre la puce à l'oreille :

Des défauts visuels : image pixélisée ou légèrement floue, clignements d'yeux ou mouvements du regard peu naturels, bouche déformée ou mal synchronisée avec les paroles, lumière et ombres anormales sur le visage.

Une synchronisation labiale imparfaite : les mouvements des lèvres ne coïncident pas parfaitement avec les mots prononcés, ou la voix semble artificielle, métallique, dépourvue de nuances et d'émotions naturelles.

Des gestes ou postures « figés » : mouvements saccadés, transitions peu fluides, éléments du corps maladroitement générés (mains, dents, contours du visage).

Des contenus trop spectaculaires ou choquants : une vidéo ou un message particulièrement surprenant, choquant ou sensationnaliste peut être un hypertrucage. Il est recommandé de **vérifier la source** avant de croire ou de partager, et de consulter des sites de vérification des faits (*fact-checking*).

Le sentiment d'urgence : dans les arnaques, la demande est presque toujours urgente : virement immédiat, communication rapide de données sensibles, décision financière sans délai. Ce contexte doit vous alerter.

Le réflexe clé : **ne pas se fier uniquement à ce que vous voyez ou entendez dans une vidéo ou un audio isolé**, surtout s'il vous demande d'agir vite ou de réaliser un paiement.

Pour ne pas tomber dans le piège de la désinformation et des arnaques à l'IA, plusieurs bons réflexes peuvent être adopté.

Protéger votre image et vos données

Une fois en ligne, vos données peuvent être copiées et détournées. Il est recommandé de :

- éviter de diffuser de façon publique et incontrôlée des images de vous ou de vos proches sur les réseaux sociaux ;
- privilégier le partage par messagerie privée sécurisée ou des partages éphémères avec une audience limitée ;
- veiller à vos photos de profil : utiliser si besoin des filtres, un cadrage plus éloigné, un floutage (par exemple avec l'application de la CNIL FantomApp) ou une pixellisation pour limiter les risques de réutilisation dans des hypertrucages ;
- faire régulièrement le ménage dans vos photos et vidéos, notamment les plus anciennes.

Limiter aussi la mise en ligne de **notes vocales, enregistrements ou vidéos en accès public** réduit la matière exploitable par des escrocs pour cloner une voix ou un comportement.

Vérifier avant d'agir

En cas de demande inhabituelle (argent, informations personnelles, consignes urgentes), même si elle semble provenir d'un proche, d'un collègue ou d'un supérieur :

- prenez le temps de **vérifier par un autre canal** : appelez la personne sur son numéro habituel, écrivez-lui via une messagerie que vous utilisez régulièrement avec elle ;
- ne vous fiez pas uniquement à l'apparence de la voix ou du visage, surtout si la demande vous met sous pression ou sort du cadre habituel ;
- si un « proche » vous appelle en détresse avec un numéro inconnu, raccrochez et rappelez sur son numéro habituel ;
- vous pouvez instaurer **un mot de sécurité familial** : un code connu de vous seuls, à demander en cas de doute lors d'un appel urgent.

Vous êtes victime ou témoin ? Les démarches et recours

Si vous repérez un *deepfake* visant à escroquer ou à désinformer, ou si vous en êtes victime, plusieurs démarches sont possibles.

Réagir immédiatement en cas de paiement ou de tentative d'escroquerie

- **Alertez votre banque** : si vous avez réalisé un virement ou un paiement, contactez votre banque sans délai pour tenter de bloquer les fonds ou d'en limiter les conséquences.
- **Conservez des preuves** : captures d'écran, liens vers les contenus, messages reçus, relevés bancaires. Elles pourront être utiles en cas de plainte ou de procédure.

Signaler le contenu illicite en ligne

- Utilisez les outils de **signalement internes des plateformes** (réseaux sociaux, services de partage de vidéos) pour dénoncer le contenu comme trompeur, abusif ou illicite.
- Pour les contenus graves (harcèlement, menaces, images sexuelles de mineurs, escroqueries, etc.), vous pouvez utiliser la plateforme nationale Pharos de signalement des contenus illicites sur internet.

Déposer plainte

En cas d'escroquerie ou de cyberharcèlement, vous pouvez :

- déposer plainte auprès d'un commissariat de police ou d'une brigade de gendarmerie ;
- utiliser, pour les escroqueries sur internet, les dispositifs de plainte en ligne ;
- en cas de doute sur la démarche à suivre, contacter des services d'information et d'aide aux victimes.

Lorsque vos données personnelles ont été utilisées sans votre accord dans un hypertrucage, vous pouvez également **saisir** l'autorité de protection des données compétente pour qu'elle examine la situation et, le cas échéant, intervienne auprès des acteurs concernés. Cette démarche ne remplace pas une plainte pénale mais peut compléter votre action.

Ressources complémentaires

- Hypertrucage (deepfake) : comment se protéger et signaler les contenus illicites ?
- CNIL
- Les arnaques utilisant l'intelligence artificielle (IA)
- Se prémunir contre l'escroquerie aux faux ordres de virement ou « arnaque au président »
- L'arnaque au président ou escroquerie aux faux ordres de virement (FOVI)
- Cybermalveillance.gouv.fr
- Pharos
- Vous pensez être victime d'un acte de cybermalveillance ?
- Tout savoir sur l'application FantomApp
- AMF (Autorité des marchés financiers) : pour vérifier la liste noire des plateformes d'investissement frauduleuses.
- L'Autorité de protection des données & vous