La menace des deepfakes

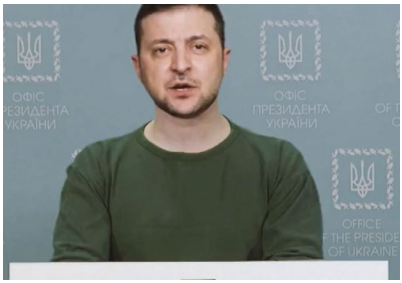
Comprendre, réagir et se protéger

Votre Police et la Prévention Suisse de la Criminalité (PSC) – un organe intercantonal de coordination spécialisé de la Conférence des directrices et directeurs des départements cantonaux de justice et de police (CCDJP), en collaboration avec Institut de Lutte contre la criminalité économique (ILCE) de la Haute école de gestion Arc, Neuchâtel.

Deepfakes – un phénomène en forte progression

Un deepfake, ou hypertrucage, est un contenu image, vidéo ou audio généré ou manipulé à l'aide d'une intelligence artificielle. Cette technique est généralement utilisée pour imiter de manière réaliste la voix ou les caractéristiques physiques d'une personne.

Exemples :



2022 : Vidéo deepfake du président ukrainien Volodymyr Zelensky appelant ses soldats à déposer les armes.

2023 : Image deepfake du pape François avec une doudoune blanche
(Photo: Deepfake généré à l'aide du logiciel IA Midjourney)

Utilisés pour tromper, manipuler ou escroquer, les deepfakes sont difficiles à distinguer de la réalité et peuvent se propager très rapidement sur le web et les réseaux sociaux.

Un deepfake peut notamment servir à :

- diffuser de fausses informations
- manipuler l'opinion publique
- usurper l'identité d'une personne
- commettre des escroqueries
- créer ou diffuser des images pornographiques altérées
- porter atteinte à la réputation d'une personne
- harceler une personne

Des possibilités presque infinies

Les outils dédiés à la création de deepfakes se multiplient et deviennent de plus en plus accessibles. Ils permettent de réaliser avec une certaine facilité des images, des vidéos, ainsi que des enregistrements audios qui déforment ou manipulent la réalité. Ces contenus, diffusés à travers des canaux publics et privés (p. ex. Whatsapp) peuvent prendre des formes et poursuivre différents objectifs.

- Placer une personnalité publique dans une situation inventée ou lui faire prononcer un faux discours.
- Insérer une célébrité dans une publicité ou une scène sans son accord.
- Modifier un événement historique ou une scène d'actualité.
- Créer des visages ou des personnages fictifs pour des profils en ligne.
- Altérer tout détail d'une image ou d'une vidéo.
- Introduire des personnes dans des situations absurdes ou ridicules.
- Générer des scènes fictives à partir d'images ou de vidéos existantes.
- Simuler la voix d'une personne afin de commettre une escroquerie.
- Utiliser le visage d'une connaissance pour créer des contenus pornographiques.
- Tromper un procédé de reconnaissance vocale ou d'identification vidéo.
- Alimenter une campagne de désinformation.
- Etc.

Atteintes à la personnalité et désinformation

L'IA est employée de diverses façons à des fins politiques. Dans certains cas, le but est de peser notamment sur les élections et sur les votations en orientant les opinions sur les réseaux sociaux par le biais de fausses informations ciblées ou d'une communication délibérément unilatérale. À cet effet, les falsifications portent sur des textes, mais aussi sur des photos et sur des vidéos. Dans d'autres cas, le but est de provoquer. Même la Suisse a déjà connu des cas de vidéos falsifiées contenant des déclarations erronées conçues pour dénigrer ou pour calomnier.

Exemple 1

Fraude à l'investissement en ligne

Sur une plateforme en ligne, Paul découvre la mention d'un article paru dans le Blick, où Roger Federer, lors d'une discussion à bâtons rompus, révèle comment il parvient à arrondir sa fortune. Paul, qui est sur le point de prendre sa retraite et sait que sa pension sera très modeste, voit là une occasion idéale de mettre du beurre dans ses épinards. Il clique donc sur le lien afin de lire la totalité de l'interview sur le site du Blick. Il se réjouit de voir que Roger Federer indique un lien menant à la plateforme d'investissement. Les placements y sont possibles à partir de CHF 250 à peine. Trop beau, pas vrai ?

Non ! Car tout est faux. La petite annonce a été mise en ligne par des criminels, le site prétendu du Blick est une copie qui n'a rien à voir avec le journal du même nom, et les conseils financiers donnés par Roger Federer sont de la fiction pure et simple. Sa photo et même sa vidéo ont été falsifiés par l'IA. D'ailleurs, pourquoi Roger Federer parlerait-il en public de ses placements ? Même de telles questions critiques ont déjà été anticipées par des criminels. C'est pourquoi l'on trouve désormais des annonces et des articles évoquant des personnalités connues qui ont prétendument été rouées de coups parce qu'elles avaient révélé les secrets de leurs investissements. Or, vous vous en doutez bien : cela aussi est *fake !*

Exemple 2

Arnaque aux sentiments

Au terme de longues recherches, Anita vient enfin de trouver sur Internet son partenaire de rêve, Rolf. C'est un ingénieur qui travaille actuellement sur une plateforme de forage dans l'Atlantique. Rolf est tombé follement amoureux d'Anita et veut la rencontrer le plus tôt possible, mais les circonstances s'y opposent. Ils dialoguent tous les jours en ligne et échangent régulièrement des photos. Et puis, Rolf tombe malade. Il doit revenir sur le continent pour se faire soigner, mais ses comptes en banques ont été bloqués. Il appelle Anita à l'aide. Dès qu'il sera guéri, ils pourront enfin se voir. Trop beau, pas vrai ?

Non ! Car tout est faux. En réalité, Rolf est un criminel qui ne ressemble en rien à sa photo, parle une autre langue qu'Anita, n'habite pas à l'adresse indiquée et ne s'intéresse qu'à l'argent d'Anita. Des assistants IA lui permettent de traduire ses messages dans la langue de son choix. Il a détourné une photo sur Internet et a utilisé l'IA pour la mettre en scène selon ses besoins. Et comme on s'en doute bien, il ne viendra jamais voir Anita.

Réagir face aux deepfakes

Les deepfakes et autres contenus manipulés peuvent paraître très réalistes, même pour des spécialistes. Il est donc essentiel d'être vigilant et d'adopter les bons réflexes avant de croire ou de partager une information. La prudence est de mise si vous utilisez des outils prétendant identifier de manière certaine les deepfakes. Il est dès lors essentiel de développer de bonnes pratiques afin de distinguer les contenus manipulés dans les informations consultées quotidiennement.

Réfléchir – vérifier – signaler

1. Adoptez un esprit critique

- Méfiez-vous des contenus qui jouent sur les émotions fortes
- Ne réagissez pas de manière hâtive face à ce type de publication
- Demandez-vous à qui a diffusé ce contenu et pour quelles raisons

Les deepfakes visent souvent à susciter la peur, la colère ou la surprise, afin de provoquer des réactions irrationnelles.

2. Vérifiez les sources

- Identifiez l'auteur du contenu
- Evaluez la fiabilité de la source
- Contrôlez si l'information apparaît dans plusieurs médias reconnus

Une information vraie est généralement relayée par plusieurs sources fiables.

3. Signalez le deepfake

- Ne le partagez pas
- Annoncez-le sur la plateforme où il est diffusé
- Informez votre entourage pour limiter sa diffusion

En signalant et en évitant de partager un deepfake, vous contribuez à combattre ce phénomène.

Que faire si vous êtes victime ?

Les deepfakes peuvent viser directement des personnes et entraîner des conséquences graves telles que des atteintes à la vie privée ou des dommages économiques. C'est le cas en particulier avec l'usurpation d'identité, le harcèlement, la diffusion de contenus pornographique ou l'escroquerie. Il est donc essentiel de limiter la diffusion du contenu et ses impacts.

- **Ne diffusez pas le deepfake,** même pour alerter votre entourage
- **Conservez toutes les preuves :**
  - captures d'écran
  - liens vers les contenus
  - messages échangés
- **Signalez le deepfake** sur la plateforme afin d'en demander sa suppression

Plusieurs infractions pénales peuvent entrer en ligne de compte, notamment :

- Escroquerie (art. 146 CP)
- Diffamation (art. 173 CP) et calomnie (art. 174 CP)
- Usurpation d'identité (art. 179^{decies} CP)
- Harcèlement (art. 181b CP)
- Pornographie (art. 197 CP), transmission indue d'un contenu non public à caractère sexuel (art. 197a CP) et désagréments d'ordre sexuel (art. 198 CP)

Des actions civiles sont également possibles en vertu de la protection de la personnalité (art. 28, 28a et 28b CC).

Selon la gravité de l'atteinte, vous pouvez déposer plainte auprès de la police.

Informations complémentaires et conseils

En cas de doute ou de suspicion de deepfakes, vous pouvez consulter le site de l'Office fédéral de la cybersécurité (www.ncsc.admin.ch) et de cybercrimepolice.ch, qui propose des informations actualisées, des conseils de préventions et des recommandations pratiques. Vous y trouverez également des indications pour déposer plainte.

Des informations complémentaires et des ressources en matière de cybersécurité sont également disponible sur le site de la Prévention Suisse de la Criminalité (www.skppsc.ch).

Face aux deepfakes

→ adoptez un esprit critique
→ vérifiez les sources
→ signalez le deepfake

Chaque vérification contribue à limiter la diffusion de fausses informations. Ne diffusez pas de deepfakes !

Suivez-nous sur les réseaux :

Facebook : @Prévention Suisse de la Criminalité **Instagram :** @skppsc_suisse
LinkedIn : @Prévention Suisse de la Criminalité **YouTube:** @SKPPSCSCP

Plus d'informations : www.skppsc.ch

SKPPSC

Prévention Suisse de la Criminalité
Maison des cantons
Speichergasse 6
3001 Berne

www.skppsc.ch

Juin 2026