# Quels risques les deepfakes représentent-ils pour les entreprises et comment s'en prémunir ?

## Réponse courte

Les deepfakes représentent des risques majeurs pour les entreprises au Luxembourg : **fraude au président** par imitation vocale ou vidéo, usurpation d'identité de dirigeants, manipulation de documents, atteinte à la réputation et désinformation interne. L'AI Act interdit les deepfakes destinés à tromper et impose le **signalement obligatoire** de tout contenu généré ou manipulé par l'IA (article 50).

L'employeur doit mettre en place des mesures de **prévention** : sensibilisation des salariés, procédures de vérification d'identité renforcées, protocoles de validation pour les transactions financières et politique de gestion de crise. Sur le plan juridique, la création et la diffusion de deepfakes malveillants peuvent constituer une **escroquerie**, une usurpation d'identité ou une atteinte à la vie privée, sanctionnées par le Code pénal luxembourgeois.

## Définition

Un **deepfake** est un contenu multimédia (vidéo, audio, image) généré ou manipulé par l'intelligence artificielle pour reproduire de manière réaliste l'apparence, la voix ou les gestes d'une personne. La technologie repose sur des réseaux de neurones capables de synthétiser des contenus quasiment indiscernables de la réalité.

Dans le contexte professionnel, les deepfakes sont utilisés à des fins frauduleuses pour **usurper l'identité** de dirigeants, clients ou partenaires. Le développement rapide de cette technologie, relevant de la réglementation IA, rend la détection de plus en plus difficile et impose aux entreprises une vigilance renforcée.

## Questions fréquentes

**Comment se prémunir contre la fraude au président par deepfake ?**

Il faut instaurer une règle de double validation pour toute instruction inhabituelle reçue par voie électronique, même si elle semble provenir d'un supérieur. Le rappel téléphonique sur un numéro vérifié et un mot de passe interne sont recommandés.

**Faut-il sensibiliser les salariés aux risques de deepfakes ?**

Oui, la formation est prioritaire pour les services financiers, la direction et les assistants de direction. Des exercices de simulation, l'identification des signaux d'alerte et une procédure de signalement complètent le dispositif de prévention.

**L'AI Act interdit-il certains deepfakes ?**

Oui, l'article 5 de l'AI Act interdit les systèmes d'IA utilisant des techniques de manipulation subliminale pour tromper. Les deepfakes destinés à induire en erreur peuvent ainsi tomber dans le champ des pratiques interdites.

**Qu'est-ce qu'un deepfake selon l'AI Act ?**

Un deepfake est un contenu multimédia (vidéo, audio, image) généré ou manipulé par l'IA pour reproduire de manière réaliste l'apparence, la voix ou les gestes d'une personne. L'article 50 de l'AI Act impose le signalement obligatoire de tout contenu généré ou manipulé.

**Quelles sanctions pénales pour la création d'un deepfake malveillant ?**

La création et la diffusion de deepfakes peuvent constituer une escroquerie (Code pénal art. 496), un faux et usage de faux (art. 193 et suivants) ou une atteinte à la vie privée. Le RGPD s'applique aussi en cas de traitement de données biométriques sans consentement.

**Quels risques les deepfakes représentent-ils pour les entreprises ?**

Les deepfakes créent des risques majeurs : fraude au président par imitation vocale ou vidéo, usurpation d'identité de dirigeants, manipulation de documents et atteinte à la réputation. Les cas de fraude peuvent atteindre plusieurs millions d'euros, surtout dans la finance.

## Conditions d’exercice

La prévention des risques liés aux deepfakes repose sur des mesures techniques, organisationnelles et juridiques.

| Critère | Détail |
|---|---|
| **AI Act - Transparence** | Obligation de signaler que le contenu a été généré ou manipulé par l'IA (article 50) |
| **AI Act - Interdictions** | Interdiction des systèmes d'IA utilisant des techniques de manipulation subliminale pour tromper (article 5) |
| **Fraude au président** | Risque d'imitation vocale ou vidéo d'un dirigeant pour ordonner des virements frauduleux |
| **Usurpation d'identité** | Utilisation de deepfakes pour se faire passer pour un salarié, un client ou un partenaire |
| **Atteinte à la réputation** | Création de faux contenus compromettants impliquant des dirigeants ou l'entreprise |
| **Code pénal** | Escroquerie (art. 496), faux et usage de faux (art. 193 et suivants), atteinte à la vie privée |
| **RGPD** | Le traitement de données biométriques pour créer un deepfake requiert le consentement explicite (article 9) |

## Modalités pratiques

La mise en place d'un dispositif de prévention des deepfakes suit un processus structuré.

| Étape | Détail |
|---|---|
| **Sensibilisation** | Formation de tous les salariés aux risques des deepfakes, exemples concrets, signaux d'alerte, procédure de signalement |
| **Procédures de vérification** | Double authentification pour les transactions financières, rappel téléphonique sur un numéro vérifié, mot de passe interne |
| **Outils de détection** | Déploiement d'outils de détection de deepfakes (analyse des métadonnées, détection d'artefacts, watermarking) |
| **Protocole de crise** | Plan de réponse en cas de deepfake détecté : alerte, investigation, communication, dépôt de plainte |
| **Politique interne** | Interdiction de création de deepfakes utilisant l'image de collègues, clients ou partenaires |
| **Veille technologique** | Suivi des évolutions des techniques de deepfake et des outils de détection disponibles |

## Pratiques et recommandations

**Former** prioritairement les services financiers, la direction et les assistants de direction aux techniques de fraude par deepfake, en organisant des exercices de simulation.

**Instaurer** une règle de double validation pour toute instruction inhabituelle reçue par voie électronique, même si elle semble provenir d'un supérieur hiérarchique identifié.

**Déployer** des outils de watermarking sur les contenus officiels de l'entreprise (communications internes, vidéos de formation) pour faciliter la détection de manipulations.

**Prévoir** un plan de communication de crise en cas de diffusion d'un deepfake impliquant l'entreprise ou ses dirigeants, incluant les voies de recours juridiques.

**Collaborer** avec les autorités compétentes (Police, Parquet) et signaler tout incident à la CNPD si des données personnelles sont impliquées.

## Cadre juridique

| Référence | Objet |
|---|---|
| **AI Act (UE 2024/1689) - Article 50** | Obligation de signalement des contenus générés ou manipulés par l'IA |
| **AI Act - Article 5** | Interdiction des systèmes d'IA utilisant des techniques de manipulation |
| **Code pénal - Art. 496** | Escroquerie par manoeuvres frauduleuses |
| **Code pénal - Art. 193 et suivants** | Faux et usage de faux |
| **RGPD - Article 9** | Protection des données biométriques, consentement explicite |
| **RGPD - Article 5** | Principe de licéité et de loyauté du traitement |
| **Art. L.312-1** | Obligation de l'employeur d'assurer la sécurité des salariés |

Les deepfakes constituent une menace croissante pour les entreprises, avec des cas de fraude atteignant plusieurs millions d'euros dans le secteur financier. La combinaison de mesures techniques, organisationnelles et de sensibilisation constitue la meilleure protection, en attendant des outils de détection plus performants.