
DÉSINFORMATION ET FAKE
NEWS GÉNÉRÉES PAR IA
RÉALISÉ PAR
LES SHOES'N

LES SHOES'N

SOMMAIRES

INTRODUCTION

L'intelligence artificielle s'est rapidement imposée dans le quotidien et le monde professionnel, notamment grâce à des outils comme ChatGPT. Elle transforme la production et la consommation des contenus numériques, et près de 60 % des professionnels du numérique l'utilisent désormais chaque jour. Mais cette généralisation pose des enjeux importants : l'IA peut générer en quelques secondes des textes, images ou vidéos très réalistes, rendant plus difficile la distinction entre vrai et faux. Dans un contexte déjà saturé d'informations, elle amplifie les risques de désinformation, ce qui nécessite des réponses techniques, réglementaires et éducatives adaptées.

PARTIE 1 : LES ENJEUX

L’impact de l’IA sur la société

L’impact de l’IA sur la société

EFFICACITÉ

L’IA générative s’est imposée comme une innovation majeure grâce au deep learning :

- Elle peut produire des textes détaillés en quelques secondes.
- Elle génère des images capables de représenter presque n’importe quelle idée.
- Elle crée des vidéos réalistes à partir de simples images.
- Elle synthétise des voix sans enregistrement humain.
- Elle automatise des tâches répétitives et accélère la création visuelle et multimédia

L’impact de l’IA sur la société

EFFICACITÉ

Néamoins elle est également capable d’engendrer des contenus dangereux dans le seul but de tromper

- Elle peut réécrire des textes officiels ou fabriquer de fausses preuves.
- Elle peut créer des contrefaçons d’œuvres ou réinventer l’histoire.
- Elle produit des vidéos réalistes montrant des événements fictifs.
- Elle imite des voix humaines avec une grande précision.

Ces contenus **imitent les codes** des médias, ce qui rend la manipulation **difficile à détecter.**

L’impact de l’IA sur la société

EFFICACITÉ

L’efficacité des Deepfakes repose principalement sur des modèles capables d’imiter le réel avec précision tout en utilisant les algorithmes à leurs avantages.

- Les deepfakes deviennent dangereux lorsqu’ils sont amplifiés par les algorithmes des réseaux sociaux.
- Ces algorithmes ciblent les personnes les plus vulnérables ou influençables.
- Des groupes malveillants peuvent manipuler les tendances ou utiliser des bots pour diffuser un deepfake.

L’impact de l’IA sur la société

Les DeepFakes sont une réalité terrifiante non pas à causes de leurs mensonges mais surtout et avant tout à cause ce fait :

Il faut 10 fois plus d'effort pour réfuter un mensonge que pour le créer.
Loi de Brandolini

L’impact de l’IA sur la société

Ce qui aggrave l’impact des deepfakes sur la société est notamment la perte de confiance de la population :

- Ils remettent en question notre capacité à croire ce que nous voyons ou entendons.
- Cette perte de confiance menace la démocratie et la stabilité sociale.
- Les deepfakes ne détruisent pas seulement la vérité : ils détruisent la confiance elle-même.

L’impact de l’IA sur la société

PRODUCTION CINÉMATOGRAPHIQUE UTILISANT LA NOTION DES DEEPFAKES

L’impact de l’IA sur la société

EXEMPLES CONCRETS

Fraude à Hong Kong (2024)

- Un employé a participé à une visioconférence entièrement générée par IA.
- Les visages et les voix des dirigeants étaient faux.
- Trompé, il a transféré 25,6 millions de dollars aux fraudeurs.

Fraude vocale (2020)

- Une imitation vocale d’un dirigeant a convaincu un manager de transférer 35 millions de dollars.
- Ce cas est devenu emblématique des arnaques financières basées sur les deepfakes.

Risques pour la société

Risques pour la société

ATTEINTE À LA DÉMOCRATIE

L'essor de l'IA générative représente une menace majeure pour les systèmes démocratiques:

- Des acteurs étatiques ou para-étatiques utilisent l'IA pour renforcer leurs campagnes d'influence.
- L'IA augmente massivement la quantité de messages diffusés.
- Elle adapte ces messages aux cultures, aux langues et aux idéologies des publics ciblés
- La désinformation assistée par IA fragilise la confiance envers les institutions.
- Elle peut décourager la participation citoyenne.

Risques pour la société

DESTABILISATION GÉOPOLITIQUE

Ces pratiques ont des conséquences directes sur la confiance accordée aux institutions démocratiques, aux médias et aux processus électoraux :

- Les deepfakes servent à manipuler des récits politiques, influencer l'opinion publique et perturber des processus électoraux.
- Ils sont également utilisés pour amplifier des tensions sociales et créer un climat de division.

Risques pour la société

LIMITES DE L'IA

Même si les deepfakes peuvent se propager rapidement, leurs effet n'est pas garanti :

- Les plateformes utilisent des outils de détection, de signalement et de filtrage pour limiter leur diffusion.
- Son efficacité dépend autant du contexte et de l'écosystème de diffusion que de la vidéo elle-même.
- Les modèles ont du mal à reproduire les micro-expressions et les émotions fines.
- Les vidéos longues présentent des variations de qualité qui révèlent leur caractère artificiel.

PARTIE 2 : LES SOLUTIONS

Détection des fake news générées par IA

Détection des fake news générées par IA

TECHNIQUES DE VÉRIFICATION PERSONNELLE

Face à la sophistication croissante des outils de génération, le premier rempart contre la désinformation reste l'utilisateur. Cependant, l'intuition ne suffit plus ; il faut adopter des réflexes méthodiques de “fact-checking citoyen”. Cette vérification s'articule autour de trois piliers :

- **L’Examen Visuel Approfondi**

L'IA commet des erreurs de logique physique que notre cerveau ignore au premier regard.
- Anomalies anatomiques : Examiner les mains, le nombre de dents et la jonction des membres.
- Incohérences physiques : Scruter les reflets asymétriques dans les pupilles et les distorsions des arrière-plans (lignes courbes, objets fusionnés).

Détection des fake news générées par IA

Détection des fake news générées par IA

TECHNIQUES DE VÉRIFICATION PERSONNELLE

- **L'Enquête Contextuelle**

Une image ne doit jamais être analysée de manière isolée.

- Recherche inversée : Utiliser Google Lens ou TinEye pour vérifier si l'image est détournée d'un événement passé.

- Traçabilité : Identifier si le contenu provient d'une banque d'images IA ou d'un forum de création numérique.

Détection des fake news générées par IA

Détection des fake news générées par IA

TECHNIQUES DE VÉRIFICATION PERSONNELLE

- **La Validation Institutionnelle**

L'urgence du partage est l'alliée de la désinformation.

- Le test de la source unique : Si une information spectaculaire n'est relayée par aucun média de référence (AFP, agences de presse), la probabilité de manipulation est élevée.

- Vérification de l'intention : Se demander pourquoi ce contenu a été créé et s'il cherche à provoquer une réaction émotionnelle forte.

Détection des fake news générées par IA

OUTILS DE FACT-CHECKING « AUTOMATISÉS »

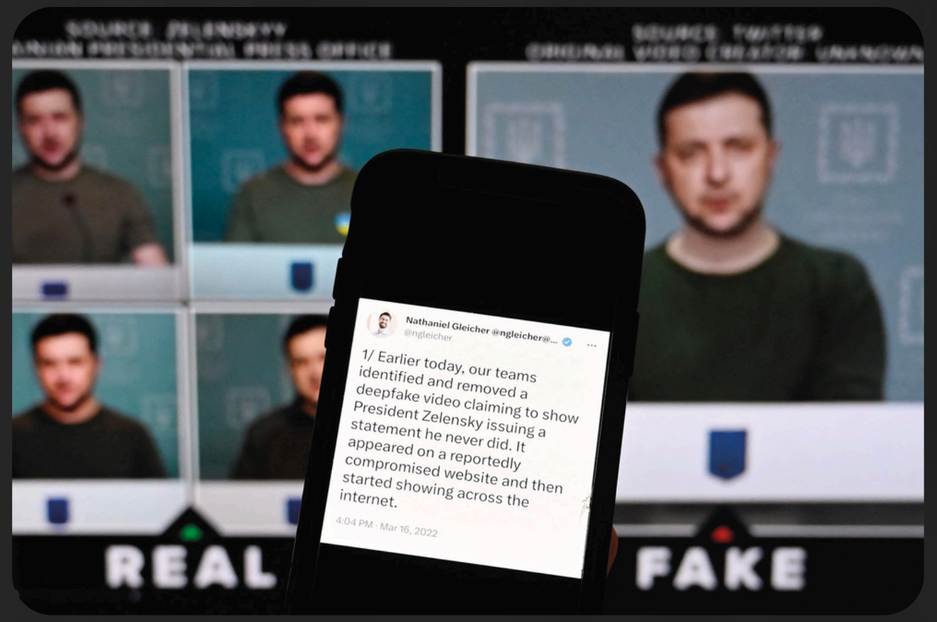
Définition :

Le fact-checking consiste à vérifier la véracité d'un texte, d'une image ou d'une vidéo avant de la considérer comme vraie. Issu du journalisme, il s'est imposé avec l'explosion des fake news et des deepfakes. Il repose sur une méthode rigoureuse : identifier la source, analyser le contenu, vérifier les incohérences et croiser avec des archives ou des bases de données. Aujourd'hui, il combine méthodes journalistiques, esprit critique et outils numériques comme la recherche inversée ou l'analyse des métadonnées.

Détection des fake news générées par IA

OUTILS DE FACT-CHECKING « AUTOMATISÉS »

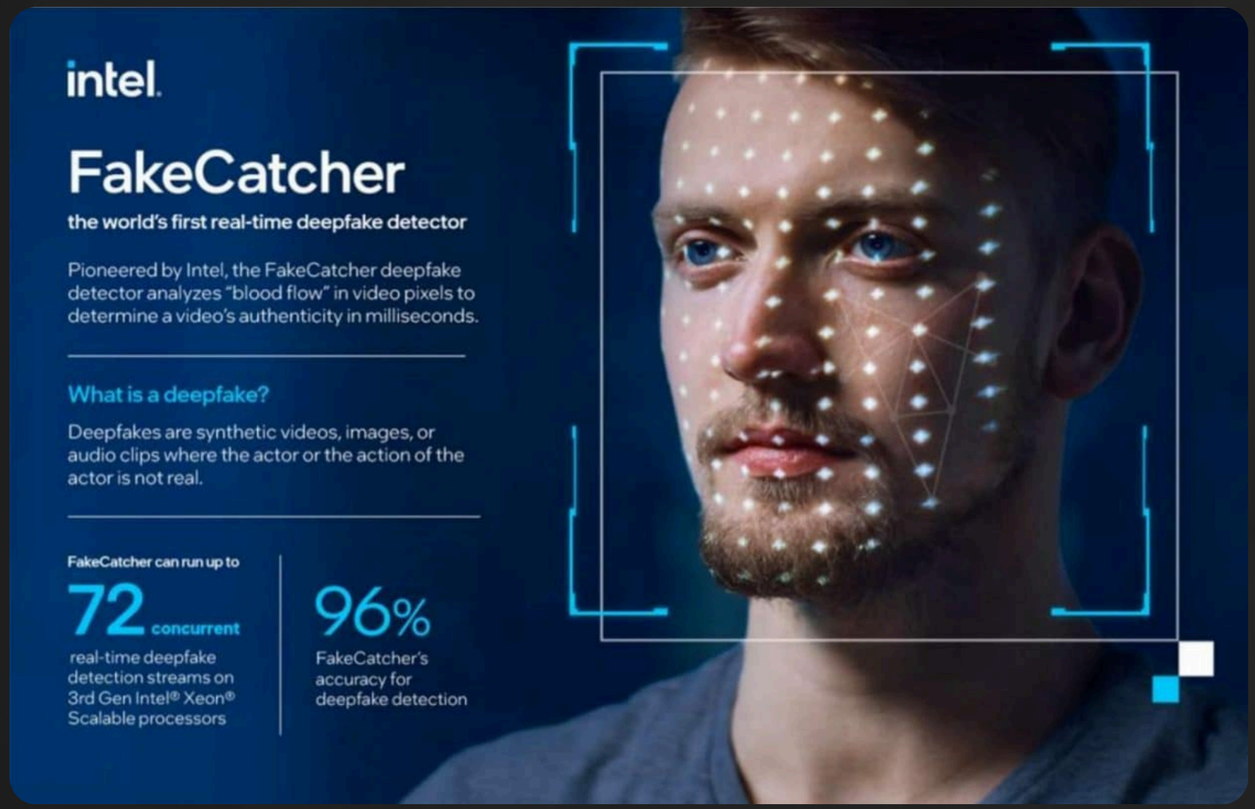
- Deepfakes → imitent parfaitement l'apparence humaine
- Risques : manipulation émotionnelle, confusion, perte de confiance
- Nécessité d'outils spécialisés comme FakeCatcher (Intel)
- Détection via micro-variations de couleur du visage (PPG)



Faux Zelensky appelant à déposer les armes, 2022

Détection des fake news générées par IA

OUTILS DE FACT-CHECKING « AUTOMATISÉS »



FakeCatcher, une I.A capable de détecter en temps réel des vidéos truquées. Image : Intel.

Moyens mis en oeuvre pour contrer le phénomène

Moyens mis en oeuvre pour contrer le phénomène

RÉGULATION DES PLATEFORMES

Les plateformes numériques ont désormais un rôle central dans la lutte contre les dérives de l'IA, car la responsabilité ne peut plus reposer uniquement sur les individus.

- Transparence et Étiquetage Obligatoire

Sous l'impulsion de réglementations comme l'IA Act, les plateformes imposent de nouveaux standards :

- Marquage (Watermarking) : Identification automatique des contenus générés par IA.
- Signalétique claire : Affichage d'étiquettes "Généré par IA" pour informer l'utilisateur en un coup d'œil.

Moyens mis en oeuvre pour contrer le phénomène

RÉGULATION DES PLATEFORMES

- Détection Algorithmique Proactive

Les plateformes déploient leurs propres outils pour stopper la désinformation à la source :
- Scan automatique : Repérage des deepfakes avant leur diffusion massive.
- Analyse de comportement : Identification des campagnes de manipulation orchestrées par des robots.

- L'Équilibre : Sécurité vs Liberté

Une régulation complexe qui nécessite une supervision humaine constante :
- Distinguer l'intention : Faire la différence entre une création artistique (humour, art) et une manipulation malveillante.
- Protection de l'espace public : Modérer les contenus sans basculer dans la censure injustifiée.

Moyens mis en oeuvre pour contrer le phénomène

ÉDUCATION AUX MÉDIAS

Explication :

L'éducation aux médias et à l'information (EMI) vise à aider les citoyens à comprendre comment les contenus sont produits, diffusés et transformés. Elle développe l'esprit critique face à un environnement numérique saturé d'informations, où fausses nouvelles, biais cognitifs et manipulations circulent rapidement. Le Ministère de la Culture la considère comme un pilier de la citoyenneté numérique, indispensable pour analyser les messages médiatiques et résister aux dérives informationnelles.

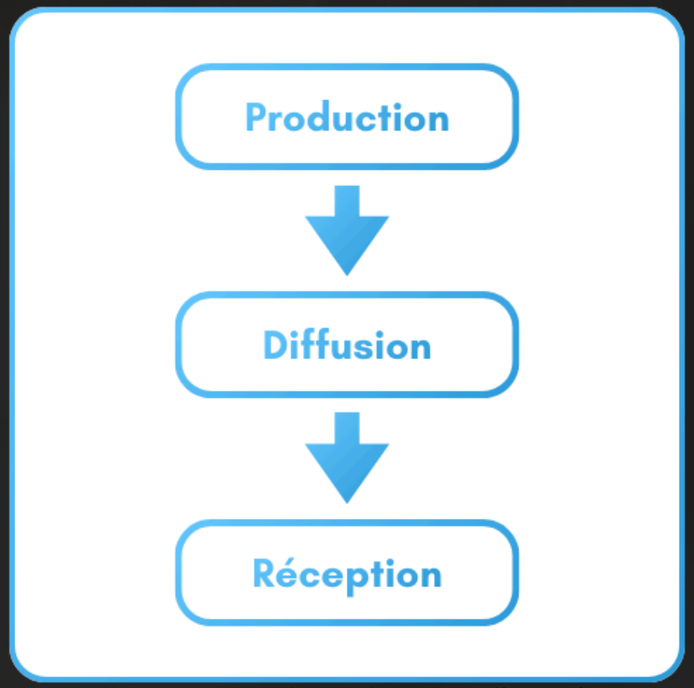


Schéma du voyage d'une information

Moyens mis en oeuvre pour contrer le phénomène

ÉDUCATION AUX MÉDIAS

Pourquoi l'EMI est indispensable aujourd'hui ?

- Explosion des réseaux sociaux → flux d'infos non vérifiées
- Biais cognitifs, bulles de filtres, viralité
- Perte de confiance envers les médias
- Nécessité de comprendre comment une info circule

Illustration pour expliquer ce qu'est la bulle de filtres

Moyens mis en oeuvre pour contrer le phénomène

ÉDUCATION AUX MÉDIAS

Deepfakes : un défi inédit pour l'EMI

- Deepfakes → brouillent la frontière entre vrai et faux
- Crise du savoir (UNESCO) : perte de confiance dans les preuves visuelles
- L'EMI apprend à vérifier, contextualiser, douter intelligemment

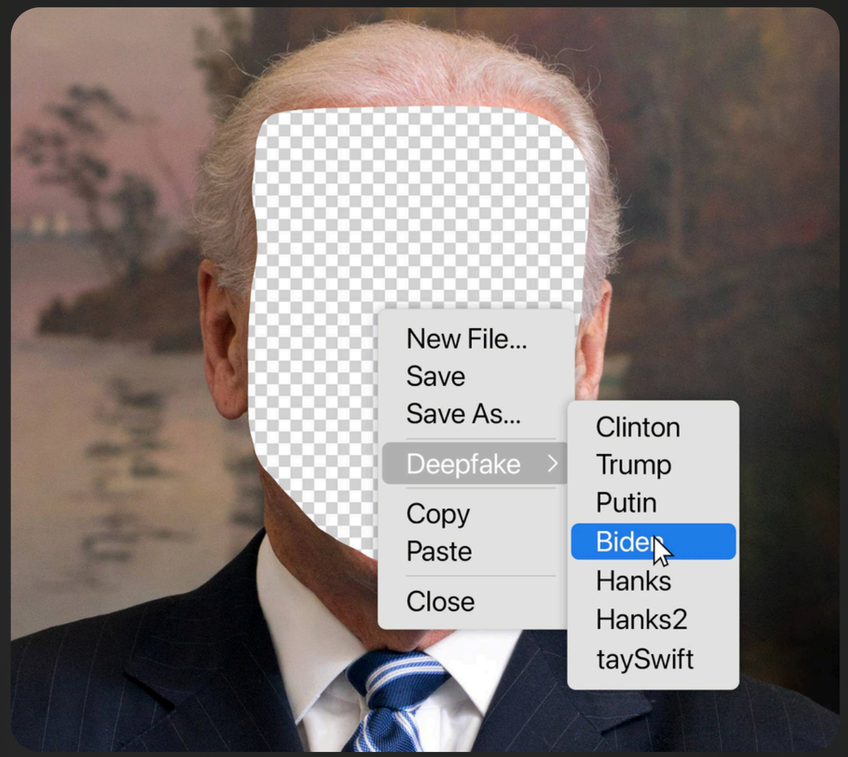


Illustration par John DiJulio, Université de Virginia

PARTIE 3 : AVANTAGES DE L'IA CONTRE LA DÉSINFORMATION

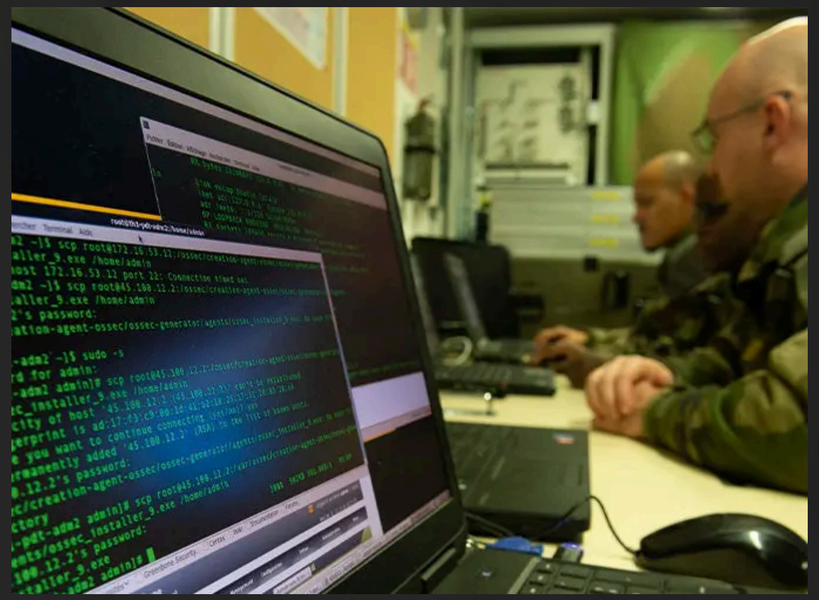
L’IA comme outil de lutte

L’IA comme outil de lutte

Voies d’utilisation:

- Traitement à grande échelle
- Outils de vérification avancés
- Complément à l’humain
- Réduction des biais cognitifs

→ Combinaison entre puissance de calcul, analyse linguistique et neutralité



L’intégration stratégique de l’IA

L’intégration stratégique de l’IA

Actions nécessaires pour que cela soit mis en place

- Intégration institutionnelle
- Profiter de son rôle stratégique
- Surveillance et alerte précoce
- Éducation aux médias
- Coopération internationale

CONCLUSION

Pour conclure, l'IA nous oblige à élever notre niveau d'exigence. Nous avons vu que la réponse n'est pas seulement technique ou juridique, elle est surtout humaine. En combinant notre vigilance individuelle à la responsabilité des entreprises, nous pouvons transformer ce défi en opportunité. L'objectif n'est pas de craindre la technologie, mais de restaurer la confiance dans l'information pour protéger notre démocratie.