

COMPRENDRE LES INTENTIONS DERRIÈRE LA DIFFUSION DE DEEPFAKES

Prérequis

- Les élèves ont déjà travaillé sur les notions d'information, de source et de désordres informationnels.

Tous les supports intégrés à cette fiche sont disponibles sur le site du CLEMI.

Niveaux de classe : 2^{de} professionnelle et 1^{re} CAP.

Durée : 2 heures.

Objectif

- Apprendre à reconnaître les vidéos manipulées par des outils d'IA, en s'interrogeant notamment sur l'intentionnalité de l'émetteur.

Entrées programmes

EMC

Réfléchir sur les problèmes soulevés par l'intelligence artificielle (IA).

Français

S'informer, informer : les circuits de l'information : développer des réflexes critiques et de vérification des supports vidéo.

PSE

Appliquer une démarche d'analyse à une situation donnée.

ÉTAPE 1. DÉFINIR CE QU'EST UN DEEPFAKE (OU HYPERTRUCAGE)

Dans un premier temps, en classe entière, l'enseignant interroge les élèves sur leur connaissance des *deepfakes* (ou hypertrucages) : connaissent-ils ce terme ? Ont-ils déjà été confrontés à ce type de contenus ? Sur quelles plateformes ou réseaux sociaux ? Quels étaient les objectifs de ces vidéos selon eux ? À partir des réponses des élèves, l'enseignant relève les mots-clés pour faire émerger une définition de ces vidéos générées ou manipulées par l'intelligence artificielle. L'enseignant peut proposer cette définition en complément : « [deepfake] fait à la fois référence à l'usage de l'IA (*deep*, pour *deep learning*, "apprentissage profond") et à la manipulation (*fake*, qui signifie "faux"). » [Morgane Tual, « Deepfake, un terme imparfait pour une réalité appelée à durer », *Le Monde*, 6 mars 2024, consulté le 7 octobre 2024]. Un lien peut être fait entre *deepfake* et infox.

ÉTAPE 2. RECONNAÎTRE LES DIFFÉRENTES INTENTIONNALITÉS

Dans un deuxième temps, l'enseignant présente un corpus de vidéos utilisant des *deepfakes*. Les élèves, répartis en groupes, analysent les différentes vidéos et déterminent leurs objectifs en répondant aux questions sous forme de tableau (voir fiche élève) : quelles sont les caractéristiques techniques de la vidéo ? (indices/observations) ; y a-t-il une volonté de tromper ou de nuire ? ; quelles sont les intentions de l'émetteur ? (intentionnalités/objectifs)

Vidéo 1 : Le Monde - Comment nous avons recréé le « vrai » Appel du 18 Juin (extrait de 10:15 à 13:59)

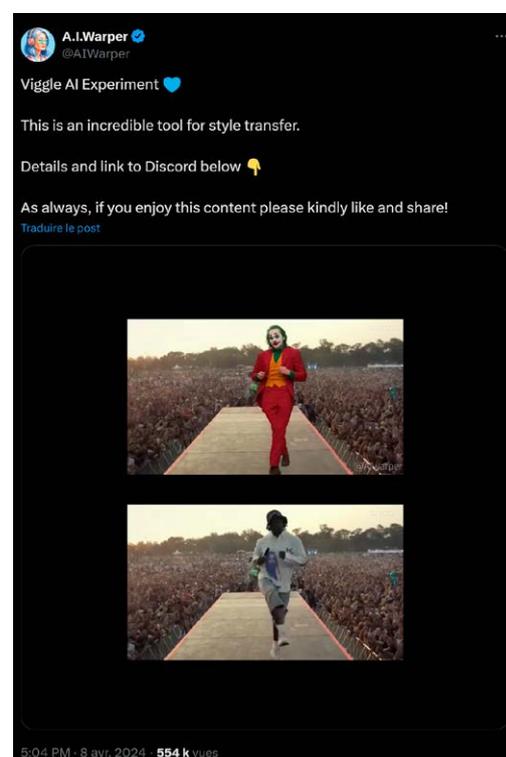


Le Monde expérimente l'utilisation de l'IA pour reconstituer, à l'aide d'historiens, de chercheurs et de documentations historiques, le

discours de l'Appel du 18 juin 1940 de Charles de Gaulle par un *deepfake* audio, interprété par un comédien. La vidéo a été réalisée dans l'objectif d'illustrer un événement historique.

Vidéo 2 : Le Joker en concert.

Le compte X AI Warper diffuse une vidéo du personnage du Joker en concert, en reprenant une vidéo du rappeur Lil Yachty. La vidéo a été réalisée à des fins de divertissement.



Vidéo 3 : Le célèbre musicien Niska a lancé son application « AZTEC STONES »

Un faux extrait d'un JT du 20 heures de TF1 présente la nouvelle application de jeux d'argent du rappeur Niska et incite à la télécharger et à jouer de l'argent pour s'enrichir. La vidéo est une arnaque.



Vidéo 4 : [France TV – Le visage de cette jeune iranienne n'existe pas.](#)

La réalisatrice du documentaire « Nous, jeunesse(s) d'Iran », Solène Chalvon-Fioriti, explique avoir eu recours à l'IA pour modifier le visage des témoins afin de garantir leur anonymat et les protéger.



Vidéo 5 : [RMC - Deepfake sur Kamala Harris : l'IA peut-elle changer le cours de l'élection présidentielle américaine ?](#)

Une victime d'un accident de la route accuse Kamala Harris, alors candidate à l'élection présidentielle américaine, de l'avoir renversée puis d'avoir pris la fuite. Ce deepfake a été diffusé dans l'intention de nuire à Kamala Harris et d'influencer l'opinion à son encontre.



En classe entière, une restitution est faite à partir des réponses de chaque groupe. L'enseignant fait une synthèse au tableau en listant les différentes intentions guidant les auteurs

de deepfakes : restitution historique, divertissement, publicité et arnaque, protection des sources journalistiques (garantir l'anonymat), propagande politique. Ces cas pratiques doivent permettre de comprendre que toutes les vidéos manipulées ne sont pas nécessairement malintentionnées ou nuisibles ; cela dépend de l'intention de l'auteur et du contexte dans lequel la vidéo est créée et diffusée. Les élèves doivent ainsi comprendre la pluralité des intentions et leur degré de dangerosité ou de nuisance. L'enseignant pourra présenter des méthodes de vérification et leurs limites : par exemple, les outils de reconnaissance des deepfakes ou la consultation de sites de vérification de faits.

Pour conclure cette deuxième partie, l'enseignant diffuse la vidéo de « Quotidien » (« Les deepfakes de Taylor Swift scandalisent les Swifties »). Cet extrait concernant les deepfakes dont a été victime Taylor Swift permet d'amorcer un échange avec les élèves sur la volonté de nuire en portant atteinte à la réputation d'une personne. Il permet d'ouvrir la discussion sur d'autres intentionnalités comme la pornodivulgateur et le cyberharcèlement.

PROLONGEMENT : QUESTIONNER L'ÉTHIQUE DE L'UTILISATION DE L'IA

En s'appuyant sur les vidéos visionnées précédemment, les élèves sont amenés à s'interroger sur les différentes questions éthiques soulevées par l'utilisation des deepfakes. L'enseignant pourra si besoin orienter la discussion autour de la déontologie des journalistes ayant recours aux deepfakes, à la notion de droits d'auteur, de droit à l'image et de propriété intellectuelle, à la question des sources ou encore à l'atteinte à la vie privée.

Flora Rodriguez, formatrice CLEMI, et Audrey Démonière-Rouvel, professeure documentaliste (académie de Créteil)

Ressources complémentaires

- « [Savoir analyser les désordres informationnels](#) », Xavier Gillet, formateur CLEMI, et Elsie Russier, responsable du pôle Labo formation du CLEMI, *Dossier pédagogique de la SPME 2023*.
- « [Évaluer l'information](#) », Elsie Russier, responsable du pôle Labo formation du CLEMI, *Les Essentiels éducation aux médias et à l'information 2024-2025*.