Fiche d'activité 19



Capture d'écran d'un deepfake copiant une chaîne d'information

Niveau d'enseignement
- Secondaire inférieur
- Secondaire supérieur

Thématique
Fake news

Méthode
Analyse des médias

Compétences visées
- Analyser une vidéo produite par un outil d'intelligence artificielle
- Connaître les stratégies permettant d'identifier une vidéo créée avec un outil d'intelligence artificielle
- Analyse rhétorique de discours : l'argument d'autorité
- Identifier les effets émotionnels d'une vidéo truquée

Quand l'IA s'empare des *News*

L'évolution des outils d'intelligences artificielles génératives permet aujourd'hui la création de vidéos de plus en plus réalistes. C'est ce qu'on appelle les *deepfakes*. Sur les réseaux sociaux, certains comptes s'emparent de cette technologie pour créer des vidéos de désinformation et susciter de vives émotions. Comment les repérer ?

L'objet à analyser

Deepfakes, des procédés si nouveaux ?

Des images de journalistes en pleine interview circulent sur les réseaux sociaux. Ces professionnel·les semblent travailler pour des médias d'information identifiables et reconnus, tel que la RTBF dans ce cas-ci. Certaines de ces vidéos dépassent les centaines de milliers de vues et suscitent des réactions dans les commentaires. Et pourtant, ces vidéos sont fausses : elles sont ce qu'on appelle des *deepfakes*. Les *deepfakes* sont des vidéos créées ou modifiées par intelligence artificielle. L'objectifs de ces dernières est souvent de ressembler au maximum à une vidéo réelle et suciter l'engagement des internautes (commentaires...).

Si le procédé s'est perfectionné grâce aux outils d'IA récents, il ne date pourtant pas d'hier. Avant cela, on pouvait déjà voir circuler des *Cheapfakes* : ce sont des images ou vidéos truquées qui requièrent peu de moyens techniques et de compétences pour être créées : un montage de base, un ralenti, ou encore un changement de légende peut suffire à induire le public en erreur tout en maintenant le sentiment de vraisemblance qu'il a face à une vidéo.

Des arguments rhétoriques

Bien que ces vidéos aient besoin de technologie de pointe pour exister, les mécanismes sur lesquels elles reposent pour faire croire en leur véracité existent depuis des siècles, voire des millénaires. Ceux-ci sont des concepts bien connu de la rhétorique et de l'analyse de discours, des disciplines qui existent déjà depuis l'antiquité.

D'une part, ces images reposent sur ce qu'on appel en rhétorique le ***pathos***. Pour convaincre le public de leur véracité, elles mettent en avant des émotions fortes : tristesse, colère, pitié. En effet, nous sommes plus facilement disposés à réagir rapidement à un contenu si celui-ci suscite chez nous des émotions. Les deux autres procédés existants en rhétorique, au mettre titre que le *pathos*, sont le *logos* (la logique) et l'*éthos* (l'image que renvoi l'orateur).

D'autre part, les vidéos simulant des interviews font également appel à l'ethos en utilisant ce qu'on appelle l'**argument d'autorité**. Ce concept désigne la tendance à accorder davantage de crédibilité aux propos d'une figure perçue comme légitime ou experte, comme un·e scientifique ou un·e professeur·e par exemple. Dans le cas présenté, le fait de croire que la vidéo provient d'un média reconnu et institutionnalisé peut conduire à une baisse de vigilance. Cela laisse à penser que la production est le produit d'une source fiable et nous dissuade de vérifier l'origine du contenu réel. On serait donc tenté de ne pas vérifier le compte sur lequel la vidéo a été publiée par exemple.

Déroulement de l'activité

Étape 1 : observation de la vidéo et analyse formelle

Présentation à la classe d'une vidéo d'actualité qui a circulé sur les réseaux sociaux : qu'est-ce que les élèves en pensent ? Si les élèves ne relèvent pas d'eux·elles-mêmes la question de la véracité des images alors le ou la professeur·e amène le questionnement.

Étape 2 : analyse critique de production vidéo

Lister avec les élèves les éléments de la vidéo qui permettent d'attester de l'utilisation d'un outil d'IA générative.



Comme pour les images fixes créées par IA génératives, il existe actuellement plusieurs indices qui permettent de conclure qu'une vidéo a été générées par Intelligence Artificielle :

- Focus sur l'arrière-plan : généralement moins précis que la scène qui se déroule à l'avant-plan, l'arrière-plan est généralement là où la majorité des défauts sont repérables.
  - Présences d'écritures : ces dernières sont floues, illisibles et changeantes au fil de la vidéo.
  - Fond flou et peu détaillé.
  - Visage et corps mal représentés.

- À l'avant plan : les IA ont, à l'heure d'aujourd'hui, encore des difficultés à reproduire certains éléments.
  - Yeux et regards des personnes : les yeux font des mouvements aléatoires et ne regardent pas dans une direction définie ou, à l'inverse, les yeux fixes un point sans suivre les mouvements naturels du visage.
  - Micro du ou de la journaliste : dans certains cas, le micro ne sera pas mis devant la bouche de la personne qui parle.
  - Présence de Watermark du programme générateur d'image.
  - Logo du média d'information erroné.

Voici les éléments figurant dans la vidéo analysée :

- Logo de la RTBF différent du logo réel et illisible par endroit.
- Présence de la mention « AI generated ».
- À l'arrière-plan :
  - Roues de tracteurs fusionnées entre elles.
  - Écriture illisible sur le gilet jaune.

Étape 3 : au-delà de la technique

Expliquer les limites de l'analyse de l'image :

Bien qu'il existe certains éléments formels qui nous indiquent qu'une vidéo est générée par IA, ces indices visuels deviennent de plus en plus subtils et difficiles à repérer. On peut en effet s'attendre à ce que ces perfectionnements et ses avancées techniques continuent et rendent donc les moyens de détection de plus en plus difficile, voire impossible. C'est pourquoi, il est important de munir les élèves d'autres outils permettant d'analyser les images en allant au-delà de la question du vrai ou du faux. On peut donc se fier à des moyens plus traditionnels de lutte contre la désinformation et d'analyse de discours.

En dehors de l'analyse de l'image, lister avec les élèves les questions à se poser qui sont relatives au contexte de diffusion :

Voici quelques exemples :

- Où a été postée la vidéo ? Sur quelle plateforme et sur quel compte ? Est-ce un compte officiel ?
- Que disent les commentaires ?
- Pourquoi serait-on amené à croire ce qui est dit dans la vidéo ? Explication de l'argument d'autorité et du rôle des émotions.

Pour aller plus loin

On pourrait aussi faire un exercice de comparaison. Pour souligner les différences existantes avec une source d'information fiable, il est possible d'effectuer un exercice de comparaison entre une vidéo générée par IA et une vidéo créée par le média d'information sur la même plateforme. Par exemple, la vidéo en IA qui utilise le logo de la RTBF et une vidéo issue de la page TikTok de la RTBF.

- Quels sont les différences ?
- Est-ce que les thèmes abordés sont similaires ou non ? De quelle manière sont-ils abordés ?

Ressources et prolongements

- ***La rhétorique du discours***
  Quelles sont les techniques de rhétoriques fallacieuses ? Comment les repérer ? Action Médias Jeunes ASBL propose une méthode d'animation complète. Lien : https://acmj.be/outilultime/les-arguments-fallacieux/
- ***Les différents types de deepfakes***
  Tous les deepfakes ne sont pas créés avec les mêmes intentions. Quelles sont-elles ? Lien : https://www.clemi.fr/ressources/ressources-pedagogiques/comprendre-les-intentions-derriere-la-diffusion-de-deepfakes

Supports pour l'activité : source

Article de la rubrique de *fact-chacking* de la RTBF, FAKI, *Cette fausse vidéo d'un agriculteur au micro de la RTBF a été générée par intelligence artificielle* publié en décembre 2025 par Ambroise Carton.

Lien : https://www.rtbf.be/article/cette-fausse-video-d-un-agriculteur-au-micro-de-la-rtbf-a-ete-generee-par-intelligence-artificielle-11650738



Cofinancé par l'Union européenne. Les points de vue et avis exprimés n'engagent toutefois que leur(s) auteur(s) et ne reflètent pas nécessairement ceux de l'Union européenne ou de l'Agence exécutive européenne pour la santé et le numérique (HaDEA). Ni l'Union européenne ni l'autorité octroyant la subvention ne peuvent en être tenues pour responsables.