

Désinformation amplifiée par l'IA : incidents médiatisés, régulations et technologies pour contrer les manipulations par hypertrucages

Projet de la CLE en
collaboration avec l'Obvia



UNIVERSITÉ
LAVAL



Crédits

Chercheure principale

Nadia Naffi

Chercheure co-responsable de l'Axe Éducation et Capacitation à l'Obvia, titulaire de la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale et Professeure agrégée en technologie éducative au Département d'études sur l'enseignement et l'apprentissage, Faculté des sciences de l'éducation à l'Université Laval

Équipe de recherche et de rédaction – volet médiatique

Mohamad-Hassan Bassal

Professionnel de recherche à la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale de l'Université Laval

Yves Joseph Kuete Tapamo

Auxiliaire de recherche à la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale de l'Université Laval

Équipe de recherche et de rédaction – volet juridique

Julie Goulet

Professionnelle de recherche à la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale de l'Université Laval

Emmy Lapointe

Auxiliaire de recherche à la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale de l'Université Laval

Équipe de recherche et de rédaction – volet technologique

Meryam Chaieb

Auxiliaire de recherche à la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale de l'Université Laval

Sérina Côté

Professionnelle de recherche à la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale de l'Université Laval

Édition du rapport

Viviane Vallerand

Coordonnatrice et auxiliaire de recherche de l'Axe Éducation et Capacitation à l'Obvia

Produit avec le soutien financier des Fonds de recherche du Québec

Fonds
de recherche

Québec



Conseil de recherches en
sciences humaines du Canada

Social Sciences and Humanities
Research Council of Canada

Canada

Table des matières

Introduction	4
Lexique	6
1. Recension des ressources hautement partagées sur les réseaux sociaux abordant l'hypertrucage	11
1.1 Des définitions de l'IA, de l'hypertrucage et de la désinformation du point de vue juridique	11
1.2 Méthodologie	13
1.3 Résultats	14
2. Contrer les dérapages de l'IA du point de vue juridique	21
2.1 Méthodologie	21
2.2 Résultats	23
3. Contrer les dérapages de l'IA et de l'hypertrucage du point de vue technologique	27
3.1 Méthodologie	27
3.2 Résultats	28
3.3 Liste non-exhaustive des chercheurs et chercheuses travaillant sur le phénomène de l'hypertrucage	34
Discussion	39
Limites	47
Conclusion	50
Références	52
Annexes	58

Introduction

Au cours des dernières années, les avancées fulgurantes de l'intelligence artificielle (IA) et des technologies d'hypertrucage ont ouvert de nouvelles perspectives fascinantes, mais elles ont également engendré des risques considérables. Ces technologies permettent de créer des contenus audiovisuels hyperréalistes, souvent indiscernables de la réalité, qui peuvent être utilisés à des fins malveillantes telles que la manipulation de l'opinion publique, la diffusion de fausses informations, et l'atteinte à la réputation.

Plusieurs organisations anticipent un possible dérapage de l'IA et de l'hypertrucage avec des effets délétères au niveau de nos démocraties mêmes. D'une part, le groupe Horizons, un Centre de recherche du Canada affilié au ministère de l'Emploi et du Développement social du Canada, prévoit cette tendance. Dans son rapport « Perturbations à l'horizon » publié en 2024, Horizons identifie les 10 perturbations les plus probables et celles ayant le plus d'impact au Canada (Gouvernement du Canada, 2024b). En tête de la liste des perturbations les plus probables figure l'incapacité des gens à distinguer le vrai du faux. De plus, Horizons classe « l'IA qui se déchaîne » comme la dixième perturbation la plus probable. D'autre part, le World Economic Forum (Forum économique mondial) (2024), dans son Global Risks Report 2024, partage cette inquiétude, considérant la désinformation et la mésinformation comme le risque le plus important à court terme à l'échelle mondiale. La majorité des personnes expertes interviewées anticipent une polarisation accrue des idéologies dans les décennies à venir et exhortent les autorités à coopérer à l'échelle internationale pour trouver des solutions à ce problème pressant.

Face à ces défis, il est urgent de développer des stratégies efficaces pour contrer les dérapages de l'IA et protéger nos sociétés et l'éducation peut jouer un rôle crucial à cet égard. C'est dans un tel contexte que ce rapport s'insère. Il a pour objectif d'analyser les besoins sur le terrain et d'évaluer les mesures actuelles en matière de régulation juridique et technologique. Il met en lumière les lacunes existantes dans ces domaines afin de mieux préciser le rôle que l'éducation peut jouer en soutenant à la fois les actions juridiques et technologiques pour contrer les dérives de l'IA, tout en palliant les limites relevées dans ces deux domaines. Pour mener à bien cette analyse, une équipe composée d'auxiliaires et de professionnels de recherche en informatique, en design, et en droit, incluant une doctorante en intelligence artificielle et une avocate, a été mobilisée. Leur expertise a permis de garantir une approche rigoureuse et multidisciplinaire, essentielle pour aborder ces enjeux complexes.

La recension conjointe des aspects médiatiques, juridiques et technologiques dans ce rapport vise à documenter les défis posés par l'hypertrucage. Cette triple approche permet de dresser un état des lieux d'incidents médiatisés, de textes juridiques et de technologies existantes. D'une part, la recension des ressources médiatiques hautement partagées sur les réseaux sociaux offre un aperçu de différents types d'incidents d'hypertrucage. D'autre part, le recensement de textes juridiques et de technologies disponibles permet de documenter des efforts actuels pour encadrer ces technologies et les détecter. En conjuguant ces trois volets complémentaires, le rapport fournit un portrait de la situation actuelle concernant des incidents d'hypertrucage, leur encadrement juridique et des outils technologiques développés pour les contrer.

Ce rapport constitue une première étape pour identifier les pistes d'action et de réflexion sur les dérapages de l'IA et de l'hypertrucage.

Ce rapport s'articule ainsi autour de trois principaux volets :

1

Volet médiatique

Cette section présente une recension des ressources médiatiques hautement partagées sur les réseaux sociaux documentant des incidents d'hypertrucage.

2

Volet juridique

Cette partie présente une recension des textes juridiques, adoptés ou en cours d'analyse, qui visent à encadrer diverses dérives que peut engendrer l'utilisation de l'hypertrucage.

3

Volet technologique

Cette partie présente une recension des techniques et outils existants pour créer et détecter les hypertrucages, en les classant selon leur origine : entreprises et développeurs, gouvernements, organismes à but non lucratif, et recherche académique.

Ce rapport constitue une première étape pour identifier les pistes d'action et de réflexion sur les dérapages de l'IA et de l'hypertrucage. Sans prétendre à une couverture exhaustive de toutes les initiatives existantes, il vise à fournir une base solide pour des discussions futures et des actions concrètes. Cette approche se veut humble, reflétant la complexité et l'évolution rapide des technologies en question, et souligne la nécessité d'une adaptation continue ainsi que d'une collaboration interdisciplinaire.

Lexique

Apprentissage automatique (*Machine learning*) :

« Mode d'**apprentissage** par lequel un **agent** évalue et améliore ses performances et son efficacité sans que son programme soit modifié, en acquérant de nouvelles connaissances et aptitudes à partir de données et/ou en réorganisant celles qu'il possède déjà. » (« Apprentissage profond »,2020).

Abroger :

Annuler ou révoquer officiellement une loi, un règlement ou une disposition légale. Lorsqu'une loi ou un règlement est abrogé, il cesse d'avoir force de loi et n'est plus applicable. L'abrogation peut être totale, supprimant entièrement la loi ou le règlement, ou partielle, ne supprimant qu'une partie spécifique de celle-ci. Au Canada, l'abrogation d'une loi est généralement effectuée par l'adoption d'une nouvelle législation ou d'un acte législatif explicitement destiné à annuler les dispositions en vigueur.

Arrêt :

Un arrêt est une décision juridique rendue par une juridiction supérieure, telle qu'une cour d'appel ou une cour suprême. Contrairement aux décisions rendues par les tribunaux de première instance, un arrêt résulte d'un appel où une partie conteste la décision initiale.

Analyse juridique :

La partie du jugement où le juge examine les arguments des parties, applique les lois et les précédents, et explique comment ils s'appliquent aux faits de l'affaire.

Cadre réglementaire flexible :

Ensemble de règles et de règlements qui peuvent être adaptés ou modifiés en fonction des évolutions technologiques et des besoins du marché, tout en garantissant la protection des droits et la sécurité.

Commissaire à l'intelligence artificielle et aux données :

Fonctionnaire ou responsable chargé de superviser et de réguler l'utilisation de l'intelligence artificielle et des données, assurant leur utilisation éthique et légale.

Commission canadienne de la sécurité numérique :

Organisme chargé de réguler et de superviser les aspects de la sécurité numérique, y compris la protection contre les cybermenaces et la désinformation.

Contrôle d'application substantiel :

Processus par lequel les autorités veillent à ce que les lois et règlements soient effectivement respectés et appliqués, avec des mesures correctives en cas de non-conformité.

Cybersécurité :

« Ensemble des mesures de sécurité informatique appliquées aux données qui passent par le réseau Internet. » (OQLF,2020)

Décision finale :

La conclusion de la cour sur l'affaire, déterminant l'issue du litige et les conséquences pour les parties impliquées.

Désinformation :

Diffusion intentionnelle de fausses informations dans le but de tromper, influencer ou manipuler l'opinion publique ou les comportements.

Disposition :

La conclusion du jugement indiquant la décision de la cour sur chaque question en litige et précisant les ordonnances ou mesures que les parties doivent suivre.

Dispositions administratives :

Les sections qui traitent de la mise en œuvre et de l'administration de la loi, y compris les rôles et responsabilités des organismes ou autorités compétentes.

Dispositions générales :

Les règles et principes de base qui régissent le projet de loi. Cela inclut souvent des déclarations de politique générale et des objectifs.

Dispositions spécifiques :

Les articles et sections détaillées du projet de loi qui énoncent les règles, obligations et interdictions spécifiques. Chaque article traite d'un aspect particulier de la loi.

Élections Canada :

Risque de dommage ou de perte pouvant résulter d'une action, d'une décision ou d'une omission, affectant les individus ou les organisations

Encadrement légal :

Ensemble des lois et règlements qui régissent un domaine spécifique qui assure que les pratiques respectent les normes juridiques et éthiques.

En-tête :

La partie supérieure d'une décision juridique qui contient des informations essentielles telles que le nom de la cour, le numéro de dossier, les noms des parties, et la date du jugement.

Fraude :

Acte délibéré de tromperie visant à obtenir un avantage illégal ou à causer un préjudice à une autre personne ou entité.

Hypertrucage (Deepfake) :

Technologie utilisant l'intelligence artificielle pour créer des vidéos ou des enregistrements audio falsifiés qui paraissent réels, souvent utilisés pour la désinformation ou la fraude (Obvia, 2024).

Images intimes non consensuelles :

Photos ou vidéos d'une personne dans des situations intimes partagées sans son consentement.

Intelligence artificielle (IA) :

Technologie qui permet aux machines d'exécuter des tâches nécessitant normalement l'intelligence humaine, telles que l'apprentissage, la reconnaissance de la parole et la prise de décision.

Jeux de données (Data set) :

« Groupe de données cohérent et structuré, portant sur un sujet déterminé, accessible pour la consultation ou le téléchargement en un ou plusieurs fichiers » (OQLF, 2020)

Justice décisionnelle :

Processus et principes guidant la prise de décision équitable et impartiale dans le système judiciaire, assurant que les décisions sont basées sur les faits et le droit.

Jugement :

Dans son sens le plus large, le terme « jugement » désigne toute décision rendue par une juridiction de l'ordre judiciaire, incluant, par exemple, les arrêts de la Cour suprême du Canada. Cependant, en français, « jugement » est également utilisé dans un sens plus restreint, en contraste avec d'autres termes tels que « arrêt ».

Juridiquement contraignant :

On dit d'un texte de droit qu'il est « juridiquement contraignant » lorsque le respect de celui-ci a un caractère obligatoire.

Législation :

Ensemble des lois et des règlements adoptés par une autorité législative, comme le Parlement du Canada ou une assemblée législative provinciale, pour régir et organiser divers aspects de la vie sociale, économique et politique. La législation inclut les actes législatifs (lois) et les textes réglementaires (règlements) qui définissent les droits, les obligations et les responsabilités des citoyens et des institutions.

Loi :

La « loi » est une disposition normative posant une règle juridique juridiquement contraignante et donc, d'application obligatoire à ceux qui y sont assujettis. Elle tend à définir des droits et/ou des devoirs. Pour devenir une loi, un projet de loi doit être approuvé par l'autorité compétente. La dernière étape confirmant qu'un projet de loi est devenu loi s'illustre souvent par son approbation officielle (sanction) par le chef d'État ou la personne autorisée pour ce faire (ex. : gouverneur général du Canada, lieutenant-gouverneur du Québec).

Loi sur la protection de la vie privée des consommateurs :

Législation canadienne visant à protéger les renseignements personnels des consommateurs contre l'utilisation abusive et à garantir leur confidentialité.

Métadonnée :

« Donnée qui renseigne sur la nature de certaines autres données dans le but d'en faciliter la compréhension et la gestion » (OQLF, 2020)

Mésinformation (Misinformation) :

un contenu faux ou trompeur partagé sans intention nuisible, même si ses effets peuvent l'être, par ex. lorsque des personnes partagent de bonne foi de fausses informations avec leurs proches.

Modèle :

« Construction mathématique générant une déduction ou une prédiction à partir de données d'entrée. Le modèle est estimé à partir de données annotées lors de la phase d'apprentissage (ou d'entraînement) du système d'IA. » (Commission nationale de l'informatique et des libertés, s. d.)

Motifs du jugement :

Les raisons détaillées pour lesquelles la cour a pris sa décision, y compris l'interprétation des lois et des précédents juridiques.

Parties impliquées :

Les personnes ou entités directement concernées par le litige et qui participent à la procédure judiciaire (par exemple, le demandeur et le défendeur).

Pornographie juvénile :

Toute représentation d'une personne de moins de 18 ans se livrant à une activité sexuelle, tout matériel montrant la région anale ou un organe sexuel d'une personne de moins de 18 ans, tout matériel écrit dont la caractéristique principale est d'encourager ou de représenter une activité sexuelle avec une personne de moins de 18 ans et tout enregistrement sonore décrivant, présentant ou simulant une activité sexuelle avec une personne de moins de 18 ans sont inclus.

Préambule :

Une section introductive qui expose les motifs et les objectifs du projet de loi. Elle fournit le contexte et les raisons pour lesquelles la législation est proposée.

Préjudice potentiel :

Risque de dommage ou de perte pouvant résulter d'une action, d'une décision ou d'une omission, affectant les individus ou les organisations

Préjudicieux (préjudiciable) :

Se dit de quelque chose qui cause ou est susceptible de causer un dommage, un tort ou un préjudice à une personne, un groupe ou une organisation.

Projet de loi :

Il s'agit d'un texte législatif n'étant pas encore juridiquement contraignant, qui peut être présenté par un ministre ou un député par exemple, en vue de son adoption souhaitée par l'instance officielle ayant autorité pour ce faire (ex. : Assemblée nationale, Chambre des Communes et Sénat, Congrès des États-Unis, Parlement européen et Conseil de l'Union européenne, etc.). À ce stade, le texte peut encore faire l'objet d'amendements. Un projet de loi doit franchir certaines étapes avant de devenir loi. Tous les projets de loi ne deviennent pas forcément une loi, car il est possible qu'il ne soit pas adopté.

Questions en litige :

Les points de droit ou de fait sur lesquels la cour doit se prononcer pour résoudre l'affaire.

Réseaux de neurones convolutifs :

« Un réseau neuronal convolutif (ConvNet ou CNN) est un réseau neuronal artificiel (ANN) qui utilise des algorithmes d'apprentissage profond pour analyser des images, classer des éléments visuels et effectuer des tâches de vision par ordinateur. » (Pathak, 2024)

Renseignements personnels :

Informations relatives à un individu identifiable incluant son nom, adresse, numéro de téléphone, informations médicales, et autres données sensibles.

Responsabilité légale :

Obligation de répondre de ses actions devant la loi, incluant la conformité aux règlements et la réparation des préjudices causés.

Résumé des faits :

Une description détaillée des événements et circonstances pertinentes à l'affaire, tels que présentés par les deux parties.

Sanctionner :

Dans le contexte canadien, sanctionner signifie approuver ou adopter officiellement un texte législatif. Par exemple, un projet de loi est sanctionné lorsqu'il reçoit l'approbation finale et devient une loi en vigueur.

Traitement du langage naturel :

« Le TLN est un sous-domaine de l'IA qui se consacre au développement d'algorithmes et à la construction de modèles capables d'utiliser le langage de la même manière que les humains. Il est couramment utilisé dans les assistants virtuels comme « Siri » et « Alexa » ou dans les recherches et traductions Google. Le TLN permet d'analyser et d'extraire des renseignements des sources non structurées, d'automatiser la réponse aux questions et de mener une analyse des sentiments et à la synthèse du texte. Les langues naturelles (communication) étant le principal moyen de collecte et d'échange de connaissances en santé publique et en médecine, le TLN est la clé pour libérer le potentiel de l'IA dans les sciences biomédicales. » (Santé Publique du Canada, 2020)

Usurpation d'identité :

Utilisation frauduleuse des informations personnelles d'une autre personne sans son consentement, souvent pour commettre un délit ou obtenir un gain financier.

Victimisation sexuelle :

Exposition d'une personne à des actes sexuels non consensuels, souvent impliquant coercition, manipulation ou violence.

Vision par ordinateur :

« La vision par ordinateur est un domaine de l'intelligence artificielle (IA) qui permet aux ordinateurs et aux systèmes de dériver des informations significatives à partir d'images numériques, de vidéos et d'autres entrées visuelles, et de prendre des mesures ou de faire des recommandations sur la base de ces informations. Si l'intelligence artificielle permet aux ordinateurs de penser, la vision par ordinateur leur permet de voir, d'observer et de comprendre. » (Naskar, 2024)



1.

Recension des ressources hautement partagées sur les réseaux sociaux abordant l'hypertrucage

L'hypertrucage, ou deepfake, est une technologie de manipulation numérique qui utilise l'intelligence artificielle pour créer des contenus audio et vidéo extrêmement réalistes mais entièrement falsifiés. Cette technologie, bien que fascinante, pose des défis considérables en matière de sécurité, de vie privée et d'intégrité de l'information. Le présent rapport vise à dresser un état des lieux des utilisations de l'hypertrucage à travers le monde, en identifiant les objectifs généralement recherchés par ces utilisations et les domaines les plus touchés. L'objectif est de mieux comprendre comment soutenir une éducation à une utilisation responsable de l'hypertrucage et de sensibiliser le public à ses effets potentiellement délétères.

La gravité de cette menace est soulignée par diverses études et organisations. Caldwell et al. (Cité dans Gouvernement du Canada, 2024a) qualifient l'hypertrucage de « plus importante menace posée par l'usage de l'IA à des fins criminelles et terroristes ». Europol (2022), de son côté, met en garde contre les multiples utilisations malveillantes de ces contenus manipulés, allant du harcèlement en ligne à la manipulation de l'opinion publique, en passant par l'extorsion, la fraude, et même l'exploitation sexuelle des enfants.

Ce premier volet du rapport présentera une recension des ressources hautement partagées sur les réseaux sociaux abordant l'hypertrucage dans le monde au cours des dernières années. Une fois l'hypertrucage bien défini (section 1.1), la méthodologie (section 1.2) de ce volet sera décrite. La section des résultats présente une analyse détaillée des données recueillies, avec une sous-section dédiée aux types de ressources trouvées, aux domaines touchés par l'hypertrucage, aux objectifs recherchés par ces utilisations, une autre sur des ressources hautement partagées sur les réseaux sociaux portant sur les impacts de l'hypertrucage sur les personnes âgées et une dernière sur les femmes.

Le présent rapport vise à dresser un état des lieux des utilisations de l'hypertrucage à travers le monde, en identifiant les objectifs généralement recherchés par ces utilisations et les domaines les plus touchés.

1.1 Des définitions de l'IA, de l'hypertrucage et de la désinformation du point de vue juridique

L'IA connaît de multiples définitions. Ainsi, l'Office québécois de la langue française, par exemple, mentionne qu'il s'agit d'un « domaine d'étude ayant pour objet la reproduction artificielle des facultés cognitives de l'intelligence humaine dans le but de créer des systèmes ou des machines capables d'exécuter des fonctions relevant normalement de celle-ci » (« Intelligence Artificielle », 2017). Au niveau canadien, le projet de loi C-27 parle plutôt de « système d'IA », qu'il décrit comme suit : « système technologique qui, de manière autonome ou partiellement autonome, traite des données liées à l'activité humaine par l'utilisation d'algorithmes génétiques, de réseaux neuronaux, d'apprentissage automatique ou d'autres techniques pour générer du contenu, faire des prédictions ou des recommandations ou prendre des décisions » (Parlement du Canada, 2022, partie 3, art. 2). Le Parlement européen traite pour sa part de l'IA comme étant « la possibilité, pour une machine, de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité » (Parlement européen, 2020).

Plus récemment, en 2024, dans le Règlement du Parlement européen et du Conseil sur l'IA, un « système d'IA » est défini comme « un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels » (Parlement européen et Conseil, 2024, art. 3, 1). Sensiblement au même effet, aux Philippines, une loi appelée « *Act establishing a regulatory framework for a robust, reliable, and trustworthy development, application, and use of artificial intelligence systems, creating the Philippine Council on artificial intelligence, delineating the roles of various government agencies, defining and penalizing certain prohibited acts* » opte pour une définition très courte de l'IA : « simulation de l'intelligence humaine dans les machines qui sont programmées pour penser comme les humains et imiter leurs actions » (Republic of the Philippines, 2023 – traduction libre).

À contrario, du côté des États-Unis, le « *National Artificial Intelligence Initiative Act of 2020* », toujours à l'étape de projet de loi, propose pour sa part une déclinaison se voulant plus précise et détaillée des actions pouvant être effectuées par l'intermédiaire de l'IA et énonce que ce terme « désigne un système basé sur une machine qui peut, pour un ensemble donné d'objectifs définis par l'homme, faire des prédictions, des recommandations ou prendre des décisions influençant des environnements réels ou virtuels, [étant entendu que] les systèmes d'IA utilisent des entrées mécaniques et humaines pour : (A) percevoir les environnements réels et virtuels; (B) résumer ces perceptions dans des modèles grâce à une analyse de manière automatisée; et (C) utiliser l'inférence de modèle pour formuler des options d'information ou d'action » (Congress.gov, 2020, sec. 3 – traduction libre). Cette définition, par ailleurs, a semblé servir de source d'inspiration à celle qui fut retenue afin de traiter de la notion d'IA dans le texte « *Act Concerning Artificial Intelligence – Senate Bill 2* », actuellement analysé au Connecticut et qui entrerait en vigueur au 1er février 2026 s'il devait être adopté (Connecticut Senate, 2024 et Rice, 2024).

D'ailleurs, pour ceux qui se sentent concernés par l'encadrement juridique de l'IA dans sa globalité et au sens large, cette loi, laquelle est décrite par Madame Tatiana Rice, Deputy Director for U.S. Legislation, comme « le premier texte législatif aux États-Unis régissant le développement et le déploiement de l'IA dans le secteur privé à une échelle comparable à celle de l'UE sur l'IA », est sans doute digne d'intérêt (Rice, 2024 – traduction libre). Force est de constater que l'effervescence entourant le développement de l'IA semble interpeller les instances, notamment législatives, qui tentent, au meilleur de leur connaissance, de cerner les pourtours de ce terme multifacettes.

À travers ces définitions, nous souhaitons d'emblée mettre en surbrillance le caractère dichotomique de l'IA, laquelle a à la fois vocation à ouvrir un monde de possibilités, mais qui tend aussi, tout à la fois, à créer des occasions « d'imiter » les humains, avec les travers que cela suppose...

Si, à ce jour, une définition unique de l'IA n'a pu être répertoriée comme faisant consensus, en revanche, la plupart des acteurs gravitant autour de ce domaine semblent s'entendre pour dire que l'IA peut être utilisée à diverses fins. Ainsi présente-t-elle certains avantages et potentiels de percées (ex. : accès à des services moins coûteux, à de l'information et à des formations, sécurisation accrue de lieux de travail, développement de nouvelles générations de produits, etc.) (Parlement européen, 2023a). Toutefois, utilisée à mauvais escient, l'IA peut entraîner des risques de dérapages nécessitant un encadrement juridique approprié.

Les hypertrucages (connus sous le nom de « deepfakes » en anglais), renvoient par exemple à des contenus générés ou modifiés à l'aide de l'IA, tels que des images, des fichiers audio ou vidéo (Parlement européen, 2023b), lesquels, à défaut d'être clairement et lisiblement identifiés comme tels, peuvent semer la confusion auprès d'une personne raisonnable qui consulterait ce type de contenu. Du côté des Philippines, on le décrit comme « tout enregistrement audio, visuel ou audiovisuel créé ou modifié par des moyens techniques, tels que l'enregistrement vidéo, le film cinématographique, l'enregistrement sonore, l'image électronique ou la photographie, qui sont si convaincants qu'une personne raisonnable le prendrait pour une représentation authentique du discours ou de la conduite d'un individu » (Congress of the Republic of the Philippines, 2023, section 3 – traduction libre).

L'hypertrucage, selon une définition retenue par le Parlement du Canada, « met à profit les algorithmes d'apprentissage automatique pour insérer des visages et des voix dans des enregistrements vidéo et audio de personnes réelles, et permet la création d'impostures réalistes » (Chesney et Citron, cité dans Siekierski, 2019, p. 1). En effet, le procédé de manipulation que supposent les hypertrucages permet de créer des contenus très convaincants, d'éléments ou d'événements n'ayant pourtant pas eu lieu tels quels (« Hypertrucage », 2019). Ainsi, un hypertrucage audio, par exemple, « réalisé à partir d'enregistrements audio et de sons de synthèse [...] permet l'imitation ultraréaliste de la voix et la modification à volonté du discours d'une personne » (« Hypertrucage », 2019). Autrement dit, la technologie « fait qu'il est possible de créer des enregistrements audio et vidéo de vraies personnes en train de dire et de faire certaines choses qu'elles n'ont jamais dites ou faites » (Chesney et Citron, cité dans Siekierski, 2019, p. 1).

De son côté, pour ajouter à l'éventail définitionnel déjà varié, la Californie a opté pour l'usage du terme « digital replica », afin de décrire un élément pourtant très ressemblant au « deepfake ». En effet, digital replica est défini dans le « *Assembly Bill 1836 (Act to amend Section 3344.1 of the Civil Code, relating to intellectual property)* » comme « une simulation numérique de la voix ou de l'image d'un individu qui ressemble si étroitement à la voix ou à l'image de l'individu qu'un profane ne serait pas en mesure de distinguer facilement la simulation numérique de la voix ou de l'image authentique de l'individu » (California Assembly, 16 janvier 2024 – traduction libre). Pourtant, le mot « hypertrucage » fait aussi partie du vocabulaire apparaissant actuellement à l'intérieur de propositions législatives à l'étude en sol californien, notamment dans le « *Assembly Bill 1856 (Act to amend Section 647 of the Penal Code, relating to crimes)* », où il désigne « tout support audio ou visuel dans un format électronique, y compris, mais sans limitation, toute image, film cinématographique ou enregistrement vidéo, qui est créé ou modifié d'une manière qui semblerait faussement à un observateur raisonnable être un enregistrement authentique du discours ou du comportement réel de la personne représentée dans l'enregistrement » (California Assembly, 26 juin 2024 – traduction libre). On peut alors se demander ce qui a pu pousser les rédacteurs législatifs à opter pour deux termes différents afin de décrire une réalité similaire.

Quant à la désinformation, elle peut se définir comme étant une « information erronée ou déformant la réalité, qui est transmise au moyen des médias de masse ou des médias sociaux, dans le but de manipuler l'opinion publique » (« Désinformation », 2019) ou une « utilisation des techniques de l'information de masse pour induire en erreur, cacher ou travestir les faits » (Le Robert Dico en ligne, s.d.).

Notons que certains font état d'une nuance entre « désinformation » et « désinformation » en anglais. Par exemple, le « *The Strengthened Code of Practice on Disinformation* » s'en remet aux définitions du « *European Democracy Action Plan* » et énonce que « désinformation is false or misleading content that is spread with an intention to deceive or secure economic or political gain and which may cause public harm » et que « désinformation is false or misleading content shared without harmful intent though the effects can be still harmful, e.g. when people share false information with friends and family in good faith » (Commission européenne, 2022, p. 1). On saisit donc que la désinformation se fonde sur un dessein malicieux tandis que le mot « désinformation », n'est pas teinté d'intention préjudiciable, même si ses effets peuvent l'être. Pour certains, le fait de présenter la désinformation comme un contenu potentiellement préjudiciable plutôt qu'illégal entraîne son lot d'enjeux (Ó Fathaigh et al., 2021, p. 2). Néanmoins, ce choix peut s'expliquer par certains défis définitoires que nous abordons plus loin.

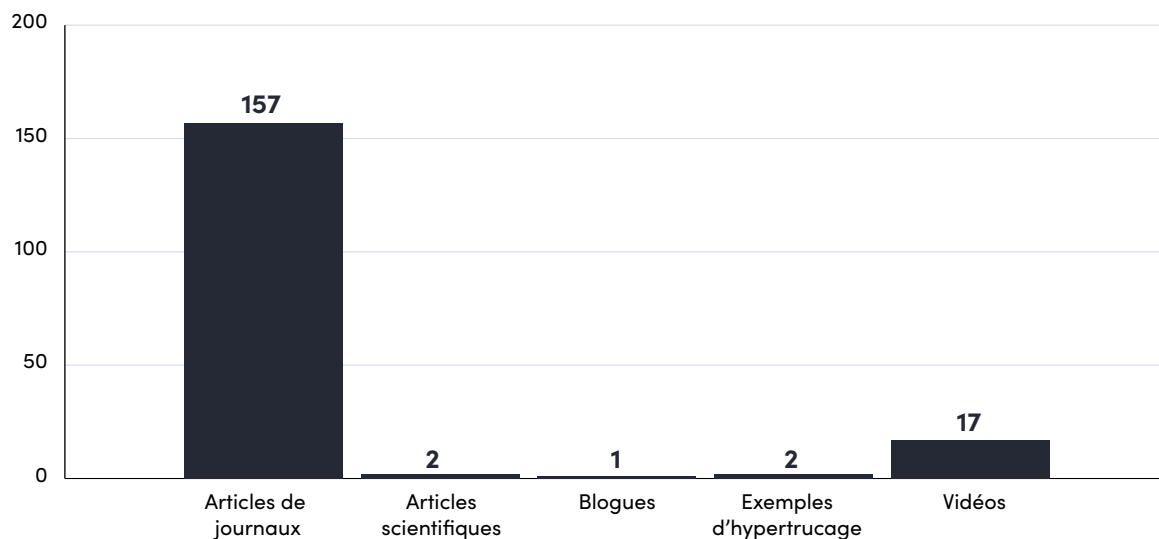
1.2 Méthodologie

Pour réaliser cette recension, une veille informationnelle libre a été menée dans les principaux médias internationaux et spécialisés. L'objectif était de collecter des ressources hautement partagées sur les réseaux sociaux accessibles et virales sur les utilisations de l'hypertrucage. Les sources incluent des articles de presse, des rapports d'organisations, des études académiques et des analyses de cas. Les ressources ont été ajoutées dans un tableau pour faciliter leur analyse et leur classification.

Une attention particulière a été portée aux ressources traitant des impacts de l'hypertrucage sur les aînés et les femmes, deux groupes particulièrement vulnérables aux abus de cette technologie. Les articles ont été sélectionnés en fonction de leur pertinence, de leur crédibilité et de leur contribution à la compréhension des enjeux liés à l'hypertrucage. Les données ont ensuite été analysées pour identifier les principaux domaines touchés et les objectifs recherchés par les utilisateurs de cette technologie. 125 ressources hautement partagées sur les réseaux sociaux ont été trouvées en tout. L'ensemble de ces ressources se trouve dans les Annexes A, B et C.

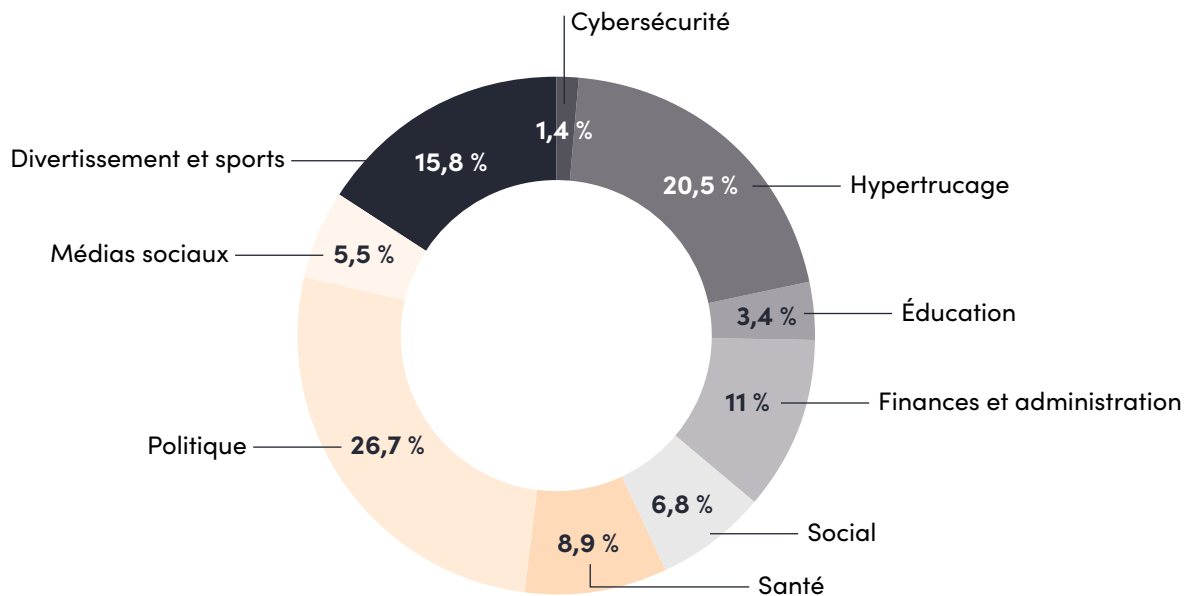
1.3 Résultats

A Types de ressources hautement partagées sur les réseaux sociaux



Les résultats des types de ressources hautement partagées sur les réseaux sociaux sur l'hypertrucage montrent une diversité de sources et de cas d'utilisation (Annexe A). Les articles de journaux, qui représentent la majorité des ressources avec 157 ressources, couvrent des sujets variés allant des fraudes financières aux manipulations politiques. Parmi les principales sources, on trouve la BBC, CNN, CBC News et NBC News, qui ont toutes rapporté des incidents impliquant des hypertrucages de toutes sortes. Les articles scientifiques, bien que moins nombreux avec seulement deux ayant été publiés par Nature et IEEE Spectrum, fournissent des analyses approfondies sur les impacts des hypertrucage et les technologies pour les détecter. Une seule ressource provenant d'un blogue offre quant à elle des perspectives détaillées sur des cas spécifiques, comme ceux analysés par le *Deepfakes Analysis Unit*. Les exemples d'hypertrucage illustrent concrètement comment cette technologie est utilisée pour tromper et manipuler, avec des cas rapportés par CNN et *The New Yorker*. Les vidéos, au nombre de 17, sont des ressources visuelles qui ont pu démontrer l'impact des hypertrucages, incluant des reportages et des analyses produits par des plateformes comme YouTube et Global News. Ces ressources montrent l'étendue et la diversité des utilisations des hypertrucages, ainsi que les défis juridiques et technologiques qu'ils posent. Elles soulignent également l'importance de la sensibilisation et de la réglementation pour contrer les dérapages de l'intelligence artificielle et de l'hypertrucage.

B Domaines des ressources hautement partagées sur les réseaux sociaux



Hypertrucage

60 ressources hautement partagées sur les réseaux sociaux portent principalement sur l'hypertrucage. Les effets pernicioeux de cette technologie y sont principalement présentés dans ces articles. Par exemple, un article de journal de Rao (2023) dans le IEEE Spectrum démontre comment le recours à l'hypertrucage peut contribuer au phénomène conspirationniste. Plus précisément, en mars 2022, l'auteur explore cet enjeu à l'aide d'une vidéo prétendant montrer le président ukrainien Volodymyr Zelenskyy conseillant la reddition aux forces russes a émergé. Bien que cette vidéo ait eu un impact limité sur le moral des troupes, elle a suscité des inquiétudes quant à la confiance dans les médias en raison des vidéos d'hypertrucage. Une étude sur les tweets liés aux hypertrucages générés par l'IA pendant la guerre russo-ukrainienne a révélé que même des hypertrucages peu convaincants ont semé le doute et les théories du complot en ligne. De nombreuses personnes ont accusé à tort des vidéos réelles d'être des hypertrucages, ce qui a alimenté le scepticisme et les théories du complot. Il est difficile de s'attaquer aux hypertrucages, car ils varient en termes d'objectif et de technologie. Les vidéos d'hypertrucage sont un problème mondial, utilisées à des fins de désinformation et de propagande. Des recherches supplémentaires demeurent toutefois nécessaires pour développer des stratégies efficaces pour les combattre.



Éducation

Dans le domaine de l'éducation, six ressources hautement partagées sur les réseaux sociaux ont été trouvées. Certaines présentent les hypertrucages comme pouvant être utilisés à des fins pédagogiques. Par exemple, Coffey (2023) dans Inside Higher Ed rapporte le cas d'un message d'intérêt public sur la cybersécurité, dans lequel figure la voix d'Astrid Tuminez, présidente de l'université Utah Valley, est raconté par un robot doté d'une intelligence artificielle qui imite sa voix. L'annonce met en garde contre les menaces de cybersécurité telles que l'hameçonnage et les escroqueries téléphoniques. Plus tard, Mme Tuminez apparaît pour révéler l'utilisation d'un clone IA de sa voix. L'article traite de l'utilisation croissante de l'IA dans les universités, y compris des avatars d'IA des présidents d'université, à des fins telles que l'engagement des étudiants. Il met en évidence les risques potentiels et les considérations éthiques associés à l'IA, en soulignant l'importance de la transparence et de l'éducation dans ce paysage technologique en constante évolution.

Cependant, l'hypertrucage pose également des risques de désinformation et d'atteinte à la vie privée. Par exemple, une vidéo d'hypertrucage a permis une escroquerie impliquant un clonage vocal à l'université Cal Poly. Des escrocs ont utilisé la technologie d'IA pour créer une imitation vocale convaincante, prétendant être un membre de la famille de la victime. Cette manipulation a été utilisée pour tenter d'extraire des informations personnelles ou de l'argent. L'incident souligne les risques croissants liés à l'utilisation malveillante du clonage vocal par IA dans les fraudes téléphoniques.



Finances et administration

Les hypertrucages représentent une menace croissante pour le secteur financier et administratif, où ils sont utilisés pour des fraudes sophistiquées. 21 ressources hautement partagées sur les réseaux sociaux ont été trouvées dans ces domaines. Par exemple, un article de journal de Flitter et Cowley (2023) dans le New York Times présente le cas de fraudeurs ayant utilisé l'hypertrucage pour frauder un client d'une banque aux États-Unis. En 2023, Clive Kabatznik, un investisseur de Floride, a appelé Bank of America, l'une des plus grandes banques américaines, pour effectuer un important transfert d'argent. La banque a ensuite reçu un deuxième appel d'une voix qui ressemblait à celle de M. Kabatznik, mais il s'agissait en réalité d'un hypertrucage qui tentait de transférer l'argent sur un autre compte. Bien que cette escroquerie ait été détectée, les données des clients, telles que les coordonnées bancaires, volées par des pirates informatiques, facilitent les attaques de ces escrocs, en particulier contre des clients fortunés ayant des apparitions publiques, y compris des discours, qui facilitent la création de fausses voix. Les fausses voix créées par des programmes informatiques ne sont pas fréquentes et n'ont fait leur apparition que l'année dernière, mais on s'attend à ce qu'elles se multiplient. Certains escrocs utilisent également des logiciels de synthèse vocale pour imiter la voix de leurs clients lorsqu'ils appellent la banque, mais cette méthode est plus facilement identifiée comme une fraude.

Un autre exemple du recours à l'hypertrucage à des fins malveillantes en finance s'est déroulé en février 2024 à Hong Kong (Chen et Magramo, 2024). En effet, des escrocs ont utilisé de l'hypertrucage pour créer une vidéo simulant le directeur financier de l'entreprise londonienne. Cette vidéo a été présentée lors d'une vidéoconférence avec l'employé basé à Hong Kong. Le faux directeur financier a demandé de manière urgente des transferts d'argent, ce qui a conduit l'employé à effectuer 15 transferts vers 5 comptes bancaires locaux. Les vidéos d'hypertrucage ont été créées à partir d'enregistrements de vraies conférences en ligne. Les escrocs ont utilisé l'application *WhatsApp*, des courriels et des appels vidéo individuels pour ajouter de la crédibilité à l'arnaque. Les voix synthétiques et les images ont été manipulées pour imiter parfaitement le directeur financier et d'autres collègues.



Social

Les hypertrucages ont un impact significatif sur des interactions sociales négatives comme le harcèlement ou l'intimidation d'individus. 14 ressources hautement partagées sur les réseaux sociaux relevant de tels cas ont été trouvées. Un exemple d'un tel usage est celui reporté par AFP (2024) dans *Le Parisien*. Selon le journaliste, des photos dénudées de plusieurs jeunes femmes ont été créées grâce à un logiciel d'hypertrucage et diffusées sur des applications de messagerie instantanée. Au total, ce sont 21 jeunes filles qui ont vu de supposées photos dénudées d'elles partagées sur les réseaux sociaux. Le réalisme des images rend très difficile de faire la différence entre celles qui étaient authentiques et celles qui avaient été modifiées, ce qui a pour effet d'accentuer les préjudices émotionnels et sociaux. Les auteurs de ces méfaits, cinq jeunes hommes, majeurs et mineurs, originaires de Séville dans le sud-ouest de l'Espagne, ont été arrêtés.



Santé

Dans le domaine de la santé, les hypertrucages peuvent être utilisés pour améliorer les diagnostics médicaux en créant des images synthétiques pour l'entraînement des algorithmes d'IA. *Retrace* (2022), leader de l'IA dentaire et de l'infrastructure de santé américaine, a publié une étude présentant un nouveau réseau adversarial génératif (GAN), proche de l'hypertrucage pour améliorer le diagnostic des maladies parodontales. De nombreux algorithmes d'IA médicale peinent à se généraliser au-delà de leurs données d'apprentissage. Cette étude GAN, portant sur 10 687 radiographies, démontre son efficacité dans l'amélioration de la précision de l'IA pour les maladies parodontales. Les avantages potentiels de la technologie s'étendent aux transactions complexes de soins de santé, améliorant les expériences patient-fournisseur et la précision du diagnostic parodontal.

Cependant, ils peuvent aussi être utilisés de manière malveillante, comme dans le cas d'une vidéo deepfake d'un médecin promouvant un faux médicament contre le dysfonctionnement érectile (Deepfakes Analysis Unit, 2024). Ce deepfake montre Rajat Sharma, un présentateur indien de nouvelles, promouvant un remède contre le diabète. La vidéo utilise des clips vidéo non liés, accompagnés d'un audio généré par IA, pour créer une fausse promotion. Une analyse a montré que le contenu vidéo et audio avait été manipulé pour donner l'impression que Sharma soutenait le produit, ce qui n'est pas le cas.



Politique

Les hypertrucages sont de plus en plus utilisés dans le domaine politique pour manipuler l'opinion publique. Par exemple, une vidéo deepfake de Kamala Harris tenant des propos controversés a circulé sur les réseaux sociaux, influençant potentiellement les perceptions des électeurs (Binder, 2024). Ce deepfake partagé par Elon Musk sur sa page X semble montrer Kamala Harris faisant des commentaires sur la diversité, Joe Biden et les politiques frontalières. La voix de Kamala Harris a été synthétisée pour faire des déclarations qu'elle n'a jamais faites. Des clips réels de Harris ont été édités et combinés avec l'audio manipulé pour créer une fausse publicité de campagne. Nous tenons à rappeler que cette vidéo a été vue plus de 150 millions de fois.



Médias sociaux

Les plateformes de médias sociaux sont des vecteurs majeurs de diffusion des hypertrucages. Ces vidéos peuvent rapidement devenir virales, comme dans le cas d'une vidéo trafiquée de Greta Thunberg qui circule en ligne, dans laquelle on la voit soutenir à tort un livre intitulé « Vegan Wars » et promouvoir l'utilisation de « grenades végétaliennes » et de « chars et armes durables » pendant les guerres (Check, 2023). Dans la vidéo, son discours ne correspond pas aux mouvements de sa bouche, et il semble qu'elle plaide en faveur d'un équipement militaire et d'armes respectueux de l'environnement. La BBC a confirmé que la vidéo est un hypertrucage et qu'elle a été créée à partir d'une interview de Greta Thunberg datant de novembre 2022. Rien ne permet d'affirmer que Greta Thunberg est l'auteur d'un livre intitulé « Vegan Wars ». La vidéo a apparemment été produite par Snicklick, un média humoristique allemand, et marquée comme « satire » sur ses pages YouTube et Instagram. Bien que l'intention soit satirique, cette désinformation pourrait nuire aux messages et à la réputation de Thunberg, ce qui peut nuire à sa cause pour le développement durable, étant donné que la vidéo satirique n'est pas étiquetée comme de l'hypertrucage.



Divertissements et sports

Dans le domaine du divertissement et des sports, les hypertrucages sont souvent utilisés pour créer des contenus sensationnalistes qui s'avèrent faux. Par exemple, une vidéo hypertrucquée impliquant Emma Watson, la vedette d'Harry Potter, dans laquelle son visage a été superposé sur le corps d'une autre personne dans une vidéo sexuellement explicite via l'utilisation de technologies recourant à l'IA (NBC News, 2023). Cet exemple démontre comment les hypertrucages peuvent nuire à la réputation de célébrités, particulièrement à celle des femmes de l'industrie.

C Objectifs de l'hypertrucage

Les hypertrucages, ou deepfakes ont plusieurs objectifs principaux, souvent malveillants. Entre autres, les hypertrucages sont souvent utilisés pour réaliser des escroqueries. L'objectif est de tromper les individus pour obtenir des informations sensibles ou de l'argent. Par exemple, des escrocs en Chine ont utilisé de l'hypertrucage pour frauder 4,3 millions de yuans (622 000 USD) à un homme en se faisant passer pour un ami lors d'un appel vidéo (Saxena, 2023).

Un autre objectif des hypertrucages est de nuire à la réputation des personnes. Cela se fait en diffusant de faux contenus pour ternir l'image publique d'une personne. Un exemple notable est une vidéo d'hypertrucage de l'actrice indienne Sonakshi Sinha la montrant lors d'un défilé de mode en maillot de bain créée pour nuire à sa réputation (Deepfakes Analysis Unit, 2024).

Les hypertrucages posent également des menaces en matière de cybersécurité. Ils sont utilisés pour des cyberattaques sophistiquées, comme l'*hameçonnage*, c'est-à-dire une tentative d'obtenir des informations sensibles ou d'accéder à un système informatique en envoyant des messages frauduleux qui semblent authentiques. Par exemple, un hypertrucage audio imitant la voix du PDG de *LastPass*, Karim Toubba, a été utilisé pour tromper un employé et obtenir des informations sensibles (Ahmed, 2024).

La désinformation et la manipulation de l'opinion publique sont d'autres objectifs courants des hypertrucages. Ils sont utilisés pour diffuser de fausses informations et manipuler l'opinion publique ou créer un climat politique conflictuel. Par exemple, une vidéo d'hypertrucage de la reine Rania de Jordanie soutenant Israël a été créée pour manipuler l'opinion publique pendant le conflit Israël-Hamas (Niranjankumar & Boom, 2023).

Les hypertrucages sont également utilisés pour créer et diffuser des contenus pornographiques sans le consentement des personnes concernées. Un exemple est celui des hypertrucages pornographiques de jeunes filles créés et diffusés en Espagne, impliquant 21 victimes (AFP, 2024).

Enfin, les hypertrucages peuvent être utilisés pour sensibiliser le public aux dangers potentiels de cette technologie et lutter contre la désinformation (Etkin, 2024 ; Voicepods, 2024). Des initiatives existent pour détecter et contrer les hypertrucages, comme celles menées par des entreprises de cybersécurité.

Pour conclure cette partie, les exemples montrent la diversité des objectifs des hypertrucages, principalement malveillants, et les dangers qu'ils représentent dans différents contextes.

D L'hypertrucage chez les personnes âgées

Les personnes âgées sont particulièrement vulnérables aux attaques par hypertrucage, en raison de leur moindre familiarité avec les technologies numériques et de leur confiance souvent plus élevée envers les communications qu'elles reçoivent (Annexe B). C'est pourquoi, dans le cadre de cette recension, une attention particulière a été apportée aux ressources hautement partagées sur les réseaux sociaux traitant de cette tranche de la population. En tout, quatre ressources ont été retrouvées qui seront ici précisées.

D'abord, Keren Etkin (2024), gérontologue, souligne dans une vidéo que les personnes âgées sont parmi les plus exposées aux hypertrucages, surtout avec la croissance des outils alimentés par l'intelligence artificielle. Elle y propose notamment des solutions pour protéger les personnes âgées, qui n'ont pas les mêmes compétences en littératie numérique que les générations plus jeunes. Cette sensibilisation est cruciale pour prévenir les abus et les escroqueries qui ciblent cette population vulnérable.

Un exemple concret de cette vulnérabilité est l'escroquerie subie par un homme de 76 ans en Inde (Kumar, 2023). Des criminels ont créé une vidéo d'hypertrucage utilisant le visage et la voix d'un ancien officier de police pour extorquer de l'argent à cet homme âgé. Effrayé par les menaces de représailles pour une fausse image où il semblait avoir sollicité des services sexuels, il a effectué plusieurs paiements aux escrocs. La police a rapidement pris une déposition et lancé une enquête, mais cet incident illustre bien les dangers auxquels les personnes âgées peuvent être confrontées.

Les escroqueries par deepfake ne se limitent pas à des vidéos. Un homme du Kerala a perdu 40 000 roupies après avoir été trompé par des cybercriminels qui ont utilisé la technologie de l'IA pour se faire passer pour un ancien collègue lors d'un appel vidéo sur WhatsApp (ET NOW, 2023). Ils ont demandé de l'argent pour une opération chirurgicale fictive.

De même, à Terre-Neuve-et-Labrador, au Canada, des personnes âgées ont été arnaquées de plus de 200 000 dollars par des fraudeurs qui ont imité la voix de leurs petits-enfants pour demander de l'argent en prétendant être en difficulté (Radio-Canada, 2023). Ces incidents montrent l'importance de sensibiliser les personnes âgées aux risques des deepfakes et de mettre en place des mesures de protection pour les aider à se défendre contre ces nouvelles formes de cybercriminalité.

E L'hypertrucage chez les femmes

Les femmes sont une autre tranche de la population particulièrement touchées par les hypertrucages, souvent utilisés pour les harceler, les humilier ou nuire à leur réputation (Annexe C). Sept des ressources hautement partagées sur les réseaux sociaux trouvées traitent de ce cas d'usage auprès des femmes.

Un exemple frappant est celui de Kgomotso Mathabe, une urologue sud-africaine, qui a découvert une vidéo d'hypertrucage d'elle-même promouvant un faux médicament contre le dysfonctionnement érectile (Nordling, 2024). Cette vidéo, suffisamment réaliste pour semer le doute parmi ses proches, dirigeait les utilisateurs vers un site Internet où ils étaient invités à entrer leurs coordonnées bancaires. Ceux qui l'ont fait ont vu leur argent siphonné de leurs comptes sans recevoir de médicament en retour. Cet incident montre comment les hypertrucages peuvent être utilisés pour des escroqueries financières, en exploitant l'image et la crédibilité des femmes pour tromper les victimes.

Les hypertrucages sont également utilisés pour créer des contenus pornographiques non consensuels, particulièrement ceux mettant en scène des femmes, causant des préjudices émotionnels et sociaux importants. Le documentaire « Another Body » suit la quête de justice d'une étudiante qui découvre une fausse pornographie d'elle-même en ligne. Ce film montre comment de jeunes femmes sont victimes de vidéos manipulées où leur image est falsifiée pour apparaître dans des contenus pornographiques. Ces femmes se battent pour reprendre le contrôle de leur identité numérique et dénoncent l'impact dévastateur de ces pratiques sur leur vie.

De même, en Espagne, des photos dénudées de plusieurs jeunes femmes ont été créées grâce à un logiciel d'intelligence artificielle et diffusées sur des applications de messagerie instantanée (AFP, 2024). Le réalisme des images rend très difficile de faire la différence entre celles qui étaient authentiques et celles qui avaient été modifiées, accentuant ainsi les préjudices émotionnels et sociaux.

Ces exemples montrent la diversité des utilisations malveillantes des hypertrucages contre les femmes, allant de l'escroquerie financière à la création de contenus pornographiques non consensuels, en passant par l'usurpation d'identité. Il est crucial de sensibiliser le public aux dangers des hypertrucages et de mettre en place des mesures de protection pour aider les victimes à se défendre contre ces nouvelles formes de cybercriminalité.





2.

Contrer les dérapages de l'IA du point de vue juridique

Dans cette partie du rapport, nous ferons état du mandat de recherche qui nous fut confié et qui consiste, essentiellement, à répertorier des textes juridiques, adoptés ou en cours d'analyse, qui visent à encadrer diverses dérapes que peut engendrer l'utilisation de l'IA à mauvais escient. D'emblée, nous exposerons en quoi a consisté le projet de recension de textes juridiques que nous avons mené et comment celui-ci s'est déployé, méthodologiquement (section 2.1). Nous continuerons en présentant ces deux grands risques de dérapages découlant de l'IA que sont les hypertrucages et la désinformation en dans la section des résultats (section 2.2). Nous terminerons avec une discussion des principales tendances et angles remarquables dans cette recension des textes juridiques pour contrer l'hypertrucage.

2.1 Méthodologie

Au cours de l'été 2024, il fut procédé à un recensement, essentiellement de lois et projets de loi en cours, afin de mieux cerner l'encadrement juridique jusqu'à présent offert et en processus d'élaboration, dans le but de contrer diverses problématiques engendrées par un usage questionnable et malveillant de l'IA. Bien que d'autres problématiques puissent s'inscrire en tant que dérives possibles de l'usage malveillant de l'IA, telles que la discrimination en contexte de travail (Saleh et Vincelette, 2024; Conseil de l'Europe, s.d.; Couillard, 2023 et Radio Canada, 2023) et des préoccupations en matière de protection de données personnelles et de la vie privée (Commission Nationale de l'Informatique et des Libertés (CNIL), s.d. et Gauthier, 2023), par choix, le présent rapport se concentre essentiellement sur les enjeux relatifs aux hypertrucages et à la désinformation.

Cette étude n'a pas la prétention d'être exhaustive. Elle brosse néanmoins un portrait de textes qui, au moment d'écrire ces lignes, s'inscrivent en tant que guides à valeur juridique afin d'orienter les actions à être prises en vue de combattre ces enjeux, et ce, à travers le monde. Les risques spécifiquement ciblés par notre recensement (hypertrucage et désinformation) sont plus amplement définis à la section suivante. La section 4 « Thématiques prépondérantes traitées législativement », ci-dessous, expose de manière synthétique le fruit de notre recensement législatif, par grands thèmes.

Dans cette partie du rapport, nous ferons état du mandat de recherche qui nous fut confié et qui consiste, essentiellement, à répertorier des textes juridiques, adoptés ou en cours d'analyse, qui visent à encadrer diverses dérapes que peut engendrer l'utilisation de l'IA à mauvais escient.

Afin de réaliser cette analyse, de nombreux documents furent consultés et analysés. Cela dit, certains sites Internet consultés ont facilité nos activités de recensement. Tel est le cas des références suivantes, lesquelles contiennent une mine de renseignements :

- National Conference of State Legislatures. (2024, 7 mai). Deceptive Audio or Visual Media ('Deepfakes') 2024 Legislation. <https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation>
- Thomson Reuters (2024). Deepfakes: Federal and state regulation aims to curb a growing threat. <https://www.thomsonreuters.com/en-us/posts/government/deepfakes-federal-state-regulation/>
- Public Citizen. (2024). Tracker: State Legislation on Deepfakes in Elections. <https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/>
- Multistate. (2024). Combating Sexual Deepfakes. <https://www.multistate.ai/deepfakes-sexual>
- Fairly. (2024). Map of Global AI Regulations - A New Era. <https://www.fairly.ai/blog/map-of-global-ai-regulations>

Plusieurs mots clés furent d'usage courant dans le cadre de nos recherches, mais ceux étant revenus les plus souvent sont entre autres : « hypertrucage », « images intimes », « pornographie », « politique », « élection », « désinformation », et leurs combinaisons et déclinaisons en anglais.

Notons également au passage que plusieurs site Web dédiés à la recherche de textes législatifs, adoptés ou possible-ment en voie de l'être, furent consultés, notamment :

- Legiscan : <https://legiscan.com/>
- Gouvernement du Canada (ministère de la Justice) : <https://www.justice.gc.ca/>
- Congrès des États-Unis (United States Congress) : <https://www.congress.gov/>
- Parlement européen : <https://www.europarl.europa.eu/plenary/fr/texts-adopted.html>

Cela dit, telle une toile d'araignée aux fils interconnectés afin de tisser un ensemble cohérent, les renseignements répertoriés sur un site Internet nous ont conduits à d'autres sources dont l'énumération, dans cette section, serait fastidieuse. Cela dit, nous avons pris grand soin de préciser nos sources documentaires, notamment dans les tableaux se trouvant en annexes. Mais avant de plonger dans les résultats de cette recherche, quelques remarques liminaires nous semblaient nécessaires.

2.2 Résultats

De nos recherches, il appert que certains grands thèmes ont émergé de manière prépondérante. Ceux-ci sont ressortis comme faisant l'objet d'initiatives d'encadrements juridiques de manière plus fréquente et relativement structurée.

A Au regard des volets politiques/électoraux

Certaines histoires impliquant des personnalités politiques témoignent du fait que ces dernières ne soient pas exemptées d'hypertrucages les impliquant ou les impactant. Pensons notamment au fait qu'en janvier 2024, « lors de la primaire aux États-Unis, des électeurs du New Hampshire ont reçu des appels téléphoniques de « Joe Biden » [ou plutôt un appel automatisé utilisant l'IA pour se faire passer pour le président Biden] leur enjoignant de ne pas voter » (Gagné, 2024 et Bond, 2024).

Songons aussi au fait qu'en septembre 2023, des hypertrucages audio ont visé Monika Tódová. En effet, « une fausse conversation générée par IA mettait en scène la journaliste et [Michal Šimečka,] le dirigeant du parti progressiste slovaque, fomentant une fraude électorale ». Tandis que ces faussetés ont été diffusées « sur les réseaux sociaux à l'avant-veille du scrutin, ce contenu a clairement profité à la désinformation [...] » (Frau-Meigs, 2024 et Reporters sans frontières, 2024). Keir Starmer, alors chef de l'opposition britannique, fit aussi, bien malgré lui, l'objet d'un clip audio semblant le montrer injurier ses employés, lequel clip serait devenu viral (Frau-Meigs, 2024 ; Bristow, 2023 et Meaker, 2023).

Marine Le Pen, femme politique française, n'a pas non plus échappé aux hypertrucages. En effet, dans un montage généré par l'IA et publié sur la plateforme X (ex-Twitter), on peut y voir la dame tenir un discours en russe (Bourgeois 2024 et Herrero, 2024). Du côté de l'Inde, peu avant les élections, des vidéos semblant montrer Manoj Tiwari, président de l'unité du Bharatiya Janata Party (BJP), à Delhi, furent envoyées à des millions d'électeurs. Ce dernier « exhortait les citoyens à voter pour le BJP, en anglais et en haryanvi », mais il s'agissait en fait d'hypertrucages (Xavier, 2020 et Jee, 2020 – traduction libre).

Et enfin, pour donner ce dernier exemple, parlons également de cette vidéo à fort déploiement, sur les réseaux sociaux, prétendant montrer le président ukrainien Volodymyr Zelensky en train de danser en tenue moulante ; un hypertrucage qui, en réalité, fut généré en remplaçant le visage d'un danseur par celui de Zelensky (Fenau, 2024 et McCarthy, 2024). Ces subterfuges ont essentiellement pour objectif d'influencer potentiellement les résultats électoraux ou alors de discréditer une personnalité politique. La désinformation et les techniques d'hypertrucage que peut engendrer une utilisation de l'IA à mauvais escient peuvent causer des dommages non seulement aux individus, mais mettre aussi à mal des processus démocratiques en cours.

Afin de tenter de dissuader ce type de pratiques, des législations furent adoptées en vue de prévoir notamment des sanctions en cas de diffusion illicite d'hypertrucages concernant des personnalités politiques. Ainsi, certaines initiatives législatives entendent criminaliser la diffusion de contenu « hypertrucages politiques » sur les plateformes de médias sociaux si la diffusion a lieu, par exemple, dans les 90 jours précédant une élection. Des lois visent aussi à prohiber de telles publications diffusées sans le consentement de la personne représentée, dans l'intention de nuire à un candidat, d'influencer les résultats des élections ou de dissuader le vote, en érigeant cela comme une infraction pénale. Les sanctions vont d'une amende à une peine d'emprisonnement possible (Bleisch, 2024).

Cela dit, certaines exceptions sont parfois prévues, sous réserve de respecter certaines conditions. Tel peut être le cas entre autres : 1) si la personne visée par l'hypertrucage politique donne son consentement à ce dernier ; 2) si l'hypertrucage est clairement identifié comme tel (c'est-à-dire qu'il contient une divulgation spécifique et visible indiquant qu'il s'agit bel et bien d'un hypertrucage) ; 3) s'il s'agit d'un hypertrucage qui constitue une parodie ou de la satire et 4) si les médias diffusent un hypertrucage aux fins d'un reportage informatif légitime (Bleisch, 2024).

L'Annexe D présente une recension des textes législatifs élaborés et traitant de cette thématique.

B En ce qui a trait aux images à caractère intime, diffusées de manière non consensuelle, incluant les images de pornographie juvénile

D'après ce qu'on peut lire dans un projet de loi philippine, il appert qu'« une proportion importante du contenu hypertrucage en ligne, environ 95 à 96 %, sont de nature pornographique et non consensuelle » (Congress of the Republic of the Philippines, 2023 – traduction libre). D'autres sources confirment cette donnée, dont l'organisme End violence against woman, au Royaume-Uni, qui déclare que « la pornographie non consensuelle constitue 96 % de tous les hypertrucages trouvés en ligne, dont 99,9 % représentent des femmes » (End violence against woman, 2024).

En début d'année 2024, le réseau social X « a bloqué les recherches sur Taylor Swift des jours après que le site eut été inondé d'images pornographiques générées par l'IA et non consensuelles de la chanteuse américaine » (Tran, 2024).

L'Annexe D présente une recension des textes législatifs élaborés et traitant de cette thématique.

C Présentation effectuée devant une cour de justice (tour d'horizon à l'échelle canadienne seulement)

Les instances judiciaires canadiennes semblent préoccupées par l'intégrité des observations susceptibles d'être présentées aux tribunaux lorsque les avocats, notamment, font usage de l'IA, et pour cause. En effet, l'affaire Zhang v. Chen, 2024 BCSC 285 (CanLII) (CanLII, 2024), qualifiée de décision historique par le cabinet d'avocats McCarthy Tétrault, constitue un exemple criant d'« utilisation abusive des outils d'IA dans une procédure judiciaire en Colombie-Britannique » (Frame et Rullmann, 2024 – traduction libre). Dans ce dossier, « l'avocate a déposé par erreur une requête contenant des autorités juridiques inexistantes qui avaient été « hallucinées » (c'est-à-dire fabriquées) par ChatGPT. L'avocate a ensuite retiré les citations fabriquées et a témoigné à l'audience qu'elle ne savait pas que ChatGPT pouvait générer de fausses autorités » (Frame et Rullmann, 2024 – traduction libre et Proctor, 2024). Le juge Masuhara a, en outre :

Ordonné un examen obligatoire par un avocat de tous ses dossiers devant la Cour suprême de la Colombie-Britannique et a déclaré qu'à l'avenir, il serait prudent pour [l'avocate concernée] d'informer le tribunal et les parties adverses lorsque les documents qu'elle soumet au tribunal comprennent du contenu généré par des outils d'IA tels que ChatGPT. (Frame et Rullmann, 2024 – traduction libre)

Comme l'expliquent les auteurs Frame et Rullmann :

Cette décision constitue un rappel important des obligations éthiques des avocats lorsqu'ils utilisent de tels outils, notamment le devoir de compétence et le devoir de vérifier les résultats produits par l'IA. Cela souligne également le manque de directives officielles sur l'utilisation des outils d'IA dans les salles d'audience de la Colombie-Britannique, en particulier sur la question de savoir si la divulgation de leur utilisation est requise. (2024 – traduction libre)

Soulignons qu'en écho à ce constat alarmant voulant que l'IA puisse induire en erreur en créant des références juridiques inexistantes, un certain nombre de juridictions exigent désormais « non seulement la divulgation du fait que des outils d'IA générative ont été utilisés pour préparer des documents déposés devant les tribunaux, mais également la divulgation de la manière dont ces outils ont été utilisés » (Frame et Rullmann, 2024 – traduction libre). L'Annexe D présente une récession des tribunaux judiciaires canadiens ayant émis des directives en ce sens.

Cela dit, notons au passage qu'en Californie, le « *Senate Bill 970 (Act to add Chapter 22.7 (commencing with Section 22650) to Division 8 of the Business and Professions Code, to amend Section 3344 of the Civil Code, to add Article 2.5 (commencing with Section 1425) to Chapter 1 of Division 11 of the Evidence Code, and to add Chapter 9 (commencing with Section 540) to Title 13 of Part 1 of the Penal Code, relating to artificial intelligence technology)* » semble faire état de préoccupations similaires :

Il exige que le Conseil judiciaire, au plus tard à la date spécifiée, examine l'impact de l'IA sur l'introduction de preuves dans les procédures judiciaires et élabore toutes les règles de procédure nécessaires pour aider les tribunaux à évaluer les allégations selon lesquelles les preuves introduites ont été générées par ou manipulé par l'IA. (California Senate, 2024 et National Conference of State Legislatures, 2024)





3.

Contrer les dérapages de l'IA et de l'hypertrucage du point de vue technologique

La montée en puissance des hypertrucages, ces vidéos, images, et enregistrements audio manipulés de manière réaliste grâce à l'IA, représente une préoccupation majeure dans les domaines de la cybersécurité, de la confidentialité, et de la confiance numérique. Alors que la technologie évolue à un rythme effréné, les hypertrucages ne sont plus seulement des curiosités technologiques, mais sont devenus des outils potentiels de désinformation, de manipulation de l'opinion publique, et de chantage. Ils constituent désormais des menaces tangibles pour les individus, les entreprises, et les gouvernements, capables de porter atteinte à la réputation, de créer des conflits, et de déstabiliser des sociétés entières.

Cette section vise à explorer en profondeur les techniques utilisées pour créer et détecter les hypertrucages, en les classant en quatre grandes catégories : l'accessibilité et l'efficacité des efforts déployés par les entreprises et les développeurs, les initiatives des gouvernements, les actions des organismes à but non lucratif, ainsi que les avancées réalisées dans le domaine de la recherche académique. En regroupant ces différentes techniques selon leurs origines et objectifs, ce rapport fournira une vue d'ensemble des efforts mondiaux pour comprendre, maîtriser, et neutraliser les hypertrucages.

D'abord, nous présenterons la méthodologie utilisée pour la réalisation d'une recension portant des technologies utilisées pour créer et détecter les hypertrucages (section 3.1). Ensuite s'ensuivra un approfondissement des principaux résultats de cette recension (section 3.2).

Cette section vise à explorer en profondeur les techniques utilisées pour créer et détecter les hypertrucages, en les classant en quatre grandes catégories : l'accessibilité et l'efficacité des efforts déployés par les entreprises et les développeurs, les initiatives des gouvernements, les actions des organismes à but non lucratif, ainsi que les avancées réalisées dans le domaine de la recherche académique.

3.1 Méthodologie

Ce projet de revue de littérature sur les techniques de création et de détection des hypertrucages a été mené en plusieurs étapes rigoureuses.

Nous avons commencé par définir ce qu'est un hypertrucage. Un hypertrucage est une technique de manipulation de contenu numérique utilisant l'IA, en particulier des réseaux de neurones profonds (d'où le terme « deep »), pour créer des images, vidéos ou enregistrements audio falsifiés qui paraissent authentiques. Cette technologie permet, par exemple, de superposer le visage d'une personne sur celui d'une autre dans une vidéo ou de synthétiser la voix d'une personne pour lui faire dire des choses qu'elle n'a jamais prononcées. Les hypertrucages sont devenus une source de préoccupation croissante en raison de leur potentiel à diffuser de la désinformation, tromper le public, ou porter atteinte à la réputation des individus.

Après avoir établi cette définition, nous avons entrepris une recherche approfondie pour identifier les types de sources disponibles concernant les outils de détection et de création d'hypertrucages. Cette recherche nous a permis de classer ces sources en quatre catégories : les entreprises technologiques, les gouvernements et entités étatiques, la communauté des chercheurs, et les organisations non gouvernementales. Pour chacune de ces catégories, nous avons procédé à la collecte d'informations sur les outils de détection de hypertrucages, en compilant des données sur leur fonctionnement, leurs technologies sous-jacentes, et leurs spécificités.

Nous avons ensuite testé ces outils un par un. Pour ceux qui ne proposaient pas de démonstration gratuite, nous avons sollicité des accès de test via des demandes spécifiques. Dans certains cas, lorsque la technologie utilisée n'était pas suffisamment documentée ou qu'une démonstration n'était pas disponible, nous avons organisé des réunions avec les développeurs ou les entreprises responsables.

Les outils ont été filtrés selon plusieurs critères, parmi lesquels : l'accessibilité à une démonstration gratuite, la qualité de la documentation fournie, et la technologie employée. Nous avons privilégié les outils les plus performants, notamment ceux qui se distinguaient malgré l'utilisation de technologies similaires à d'autres outils.

Par la suite, nous avons entamé la rédaction d'une fiche Excel pour les outils sélectionnés, en détaillant divers aspects tels que l'équipe de développement, le public cible, la description de la technologie utilisée, ainsi que les avantages et les limites de chaque outil. Cette étape de rédaction a été complétée par un second test approfondi de chaque outil (Annexe F).

Dans le cadre de cette recherche, nous avons également exploré les outils de création d'hypertrucages, car certains d'entre eux proposent à la fois des fonctionnalités de création et de détection, un phénomène qui a particulièrement retenu notre attention.

Enfin, nous avons identifié et recensé les chercheurs influents dans le domaine des hypertrucages, en nous concentrant sur ceux dont les travaux sont les plus cités et qui sont activement engagés dans cette thématique. Nous avons aussi mis l'accent sur les revues et les technologies développées dans le cadre de la recherche.

Cette section résume et critique les résultats obtenus. Chaque sous-section qui suit se concentre sur une des catégories que nous avons explorées, fournissant une analyse détaillée des outils et technologies identifiés.

3.2 Résultats

A L'accessibilité et l'efficacité des efforts déployés par les entreprises et les développeurs

La plupart des outils de création d'hypertrucages offrent non seulement des démonstrations gratuites, mais permettent également une utilisation complète gratuite via une licence de base. Ces versions gratuites sont souvent suffisamment robustes pour produire des hypertrucages convaincants, bien que des licences payantes soient disponibles pour ceux qui recherchent une précision accrue, des performances améliorées ou la possibilité de tests plus poussés. Des logiciels tels que ElevenLabs (ElevenLabs, 2024) et Synthesia ([PULL FROM H1], n.d; Montage vidéo IA, n.d; Réalisez des vidéos face caméra en 5 minutes, n.d) en sont des exemples notables. Ils présentent généralement une interface conviviale et intuitive, rendant la création d'hypertrucages accessible même à des utilisateurs sans compétences techniques avancées. Ces outils, souvent disponibles en version web, sont compatibles avec tout type de système d'exploitation sans nécessiter d'installation, ce qui abaisse encore la barrière d'entrée pour un large public.

L'accessibilité de ces outils est également renforcée par la disponibilité d'API ou de modèles open source, comme ceux partagés sur des plateformes telles que GitHub. Ces ressources sont particulièrement précieuses pour les chercheurs, les développeurs ou toute personne ayant des compétences en programmation et un intérêt pour la technologie des hypertrucages. Grâce à ces outils, il est possible de créer des hypertrucages de qualité sans nécessiter d'investissements substantiels en logiciels ou en matériel, ce qui a contribué à l'augmentation rapide de la création de ces contenus.

En revanche, les outils et initiatives dédiés à la détection des hypertrucages sont moins souvent disponibles en version gratuite et leur accès est plus limité. Les solutions gratuites qui existent tendent à être moins performantes, ce qui réduit leur efficacité dans la détection précise et rapide des hypertrucages. Par exemple, certains outils comme Resemble.ai (Resemble AI, 2024) offrent des démonstrations sur demande, mais cela nécessite généralement une approbation préalable, limitant ainsi leur accessibilité immédiate.

Les grandes entreprises technologiques investissent massivement dans le développement de technologies de détection des hypertrucages. Des initiatives telles que FakeCatcher d'Intel (Intel fake catcher, n.d), (Clayton, B. J., 2023) ou Microsoft Video Authenticator (Burt, T, 2021; The Deepfake Detection Challenge (DFDC) dataset, n.d; Kelion, B. L, 2020; Microsoft, n.d), développées par Microsoft Research, illustrent ces efforts. De même, Google DeepMind a mis au point Google SynthID (SynthID, 2024; Netgrows Tech, 2023; DeepMind, n.d), une technologie pionnière dans la lutte contre les hypertrucages. Cependant, ces technologies ne sont pas destinées au grand public et restent majoritairement intégrées seulement dans les produits et services des entreprises qui les développent. Pour bénéficier de ces technologies, les utilisateurs finaux doivent généralement acquérir des matériels spécifiques (ordinateurs, téléphones, etc.) qui les intègrent, ce qui peut limiter leur diffusion.

En parallèle, des initiatives collaboratives comme la Content Authenticity Initiative (CAI) (Ducarf, J, 2024; Parsons, A, 2024; Deschaseaux, C, 2024) voient le jour. Cette initiative regroupe plus de 2000 entreprises, dont des géants de la photographie comme Nikon, Canon, et Leica, ainsi que des leaders technologiques tels que Microsoft et Intel. Elle inclut également des médias de renom comme l'AFP, la BBC, et le New York Times, ainsi que des experts en traitement d'image comme Adobe. Le but de cette initiative est de créer un écosystème de confiance où l'authenticité des contenus est garantie à travers une coopération étroite entre médias et entreprises technologiques. Ce type de collaboration vise à établir des normes et des technologies communes pour identifier et signaler les contenus manipulés.

Les plateformes de réseaux sociaux, conscientes des risques posés par les hypertrucages, ont commencé à prendre des mesures pour limiter leur impact. Elles mettent en place des initiatives et des politiques d'utilisation plus strictes pour protéger les utilisateurs tout en préservant la liberté d'expression. Par exemple, Meta (anciennement Facebook) a lancé des défis pour inciter les développeurs à se concentrer sur la détection des hypertrucages. Ces plateformes utilisent également des algorithmes internes pour étiqueter les contenus potentiellement générés par IA, tout en s'appuyant sur les signalements des utilisateurs pour identifier les manipulations organisées ou intentionnelles.

B Pays et gouvernements

Certaines entreprises développent des logiciels spécifiquement destinés à un usage gouvernemental, que ce soit dans le cadre de collaborations formelles ou à titre bénévole. Par exemple, la plateforme Sentinel, spécialisée dans la détection des hypertrucages, a été créée pour l'Estonie. Principalement utilisée par des entités gouvernementales, des agences de défense et des organisations médiatiques, Sentinel exploite des réseaux neuronaux avancés pour analyser les médias et identifier les signes de manipulation. Ces analyses portent sur les expressions faciales, les clignements des yeux et les manipulations audios, utilisant des modèles d'apprentissage profond pour examiner une large gamme de signaux biologiques, de données de reconnaissance faciale et de données audio. Les réseaux de neurones convolutifs (CNN) sont notamment employés pour détecter avec précision les images et vidéos manipulées, permettant ainsi une reconnaissance efficace des visages et des objets altérés.

La mission de Sentinel (Sentinel – defending against deepfakes and information warfare, n.d) est de protéger les démocraties contre la guerre de l'information, les campagnes de désinformation, et la manipulation des médias. Fondée par d'anciens experts en IA et en cybersécurité de l'OTAN (Organisation du traité de l'Atlantique Nord), Sentinel bénéficie du soutien d'investisseurs renommés tels que Jaan Tallinn (cofondateur de Skype et investisseur précoce de DeepMind) et Taavet Hinrikus (cofondateur de TransferWise). Leur vision est de devenir une « couche de confiance » sur Internet, vérifiant l'ensemble de la chaîne d'approvisionnement en informations critiques et protégeant un milliard de personnes contre la guerre de l'information.

Sentinel maintient également une vaste base de données des hypertrucages vérifiés pour améliorer la précision de ses détections. Cette base de données permet d'identifier les modèles courants de manipulation utilisés dans les hypertrucages. La plateforme utilise une classification sophistiquée pour distinguer les visages réels des visages manipulés et fournit des visualisations des manipulations détectées, permettant aux utilisateurs de voir précisément où et comment le contenu a été altéré. Son approche de détection s'inspire de la norme de défense en profondeur (DiD) en cybersécurité.

D'autres initiatives rassemblent plusieurs pays, comme l'Union européenne avec le projet WeVerify (Vlinkinfo, n.d), lancé en décembre 2018 dans le cadre du programme Horizon 2020. Développé par un consortium de partenaires européens, WeVerify est principalement utilisé par des institutions et des organisations à travers l'Europe, bien que ses outils et ressources soient accessibles mondialement, en particulier pour les professionnels des médias, les chercheurs et les vérificateurs de faits. Le projet, initié en 2020 au Centre de recherche et technologie Hellas, utilise une combinaison de techniques de vérification des faits et de détection des manipulations médiatiques, y compris la détection des hypertrucages. La plateforme intègre des technologies de vision par ordinateur, de traitement du langage naturel, et d'apprentissage automatique pour analyser et authentifier le contenu visuel et textuel. Parmi les fonctionnalités de WeVerify, on trouve des outils de vérification d'image, de traçage de l'origine des contenus et de détection des manipulations vidéo, avec des algorithmes spécifiques pour identifier les altérations subtiles caractéristiques des hypertrucages. Ces outils permettent aux utilisateurs de vérifier l'authenticité des images et des vidéos en analysant les métadonnées, les incohérences visuelles, et les anomalies dans les fichiers médias.

Aux États-Unis, des mesures sont en place depuis 2019 pour contrer les hypertrucages, en raison des menaces croissantes qu'ils représentent pour la sécurité nationale, la vie privée, et l'intégrité des informations. Le gouvernement fédéral et les États ont adopté diverses lois, technologies, et initiatives de sensibilisation pour réguler ces contenus manipulés. Parmi les lois fédérales, le « *National Defense Authorization Act (NDAA) 2020* » exige que les agences de renseignement américaines évaluent les menaces potentielles posées par les hypertrucages. Bien que le « *Communications Decency Act (CDA) Section 230* » ne traite pas spécifiquement des hypertrucages, il joue un rôle crucial en protégeant les plateformes en ligne de la responsabilité concernant le contenu généré par les utilisateurs. Certains États, comme la Californie, le Texas, et la Virginie, ont adopté des lois pour contrer l'utilisation malveillante des hypertrucages, notamment dans un contexte électoral ou pornographique.

Sur le plan technologique, le National Institute of Standards and Technology (NIST) et la Defense Advanced Research Projects Agency (DARPA) développent des normes et des outils pour détecter les hypertrucages. La DARPA, à travers le programme Semantic Forensics (SemaFor) (Forensic, n.d), se concentre sur l'identification des altérations sémantiques dans les contenus multimédias. Le Department of Homeland Security (DHS) a également lancé plusieurs initiatives pour lutter contre les hypertrucages, notamment via la Cybersecurity and Infrastructure Security Agency (CISA), qui développe des technologies de détection et propose des formations aux professionnels de la cybersécurité.

En Chine, le gouvernement a mis en place des lois strictes, développé des technologies avancées, et lancé des initiatives de sensibilisation pour réguler l'utilisation des hypertrucages. En janvier 2020, le Cyberspace Administration of China (CAC) a publié une réglementation exigeant que les contenus générés par l'IA soient clairement marqués comme tels, et que les plateformes en ligne soient tenues de supprimer les contenus illégaux. La Chine investit également dans le développement de technologies de détection des hypertrucages, avec des algorithmes avancés développés par des institutions comme l'Université de Tsinghua. Depuis janvier 2023, des réglementations strictes concernant la technologie de synthèse profonde, émises par la CAC, le ministère de l'Industrie et des Technologies de l'Information (MIIT), et le ministère de la Sécurité Publique (MPS), sont en vigueur (AFP Le Journal De Montréal, 2023).

En Allemagne, des projets tels que DeFaktS, lancé en 2022 par le Centre de recherche sur l'informatique (FZI), et NoFake, développé par la Ruhr-Universität Bochum et CORRECTIV, utilisent l'IA pour détecter la désinformation propagée sur les réseaux sociaux et les plateformes numériques. NoFake, qui se termine en décembre 2024, exploite des modèles basés sur les transformateurs, comme BERT, pour analyser et classifier les informations en ligne (noFake – Vernetzung und Sicherheit digitaler Systeme, n.d).

Au Canada, le gouvernement a également pris des mesures pour limiter l'impact des technologies d'hypertrucage. Le Code criminel canadien comprend des dispositions qui peuvent s'appliquer aux hypertrucages nuisibles, et la Législation sur la Protection de la Vie Privée protège contre l'utilisation non autorisée des informations personnelles. Diverses initiatives technologiques, comme l'Initiative du Citoyen Numérique (ICN), soutenue par le ministère du Patrimoine canadien, visent à renforcer la résilience contre la désinformation en ligne, y compris les hypertrucages (Hypertrucage* : Que peut-on faire à propos du contenu audio et vidéo de synthèse? , n.d).

Cette section illustre les efforts déployés par différents pays et gouvernements pour combattre la menace croissante des hypertrucages, en développant des technologies avancées de détection et en adoptant des réglementations spécifiques. Alors que des plateformes comme Sentinel en Estonie et des initiatives telles que WeVerify en Europe montrent l'engagement de divers acteurs dans la lutte contre la manipulation des médias, les États-Unis, la Chine, l'Allemagne et le Canada adoptent des approches variées en combinant législation, recherche technologique et initiatives de sensibilisation. Ces efforts, bien que distincts, visent tous à renforcer la résilience contre la désinformation, et à protéger les institutions et les citoyens contre les impacts négatifs des hypertrucages.

C Organisme à but non lucratif

Plusieurs organismes à but non lucratif se mobilisent pour lutter contre la désinformation, notamment à travers la détection et la prévention des hypertrucages. Parmi eux, l'initiative Ban Deepfakes (The solution | Ban deepfakes, s.d), lancée par l'organisation Future of Life Institute en février 2024, se présente comme une réponse directe à la menace croissante des hypertrucages (The solution | Ban deepfakes, s.d), (Media | Ban deepfakes, s.d), (FAQ | Ban Deepfakes, s.d). Cette initiative cherche à interdire la création et la diffusion des hypertrucages à des fins malveillantes. Cependant, bien qu'elle adresse un enjeu critique, elle soulève également des questions sur son accessibilité, son efficacité, et les ressources qu'elle mobilise.

Sur le plan de l'accessibilité, Ban Deepfakes ne fournit pas de détails précis sur les plateformes compatibles, limitant ainsi la compréhension de sa portée technologique (The solution | Ban Deepfakes, s.d). De plus, l'absence d'informations claires sur les coûts complique l'évaluation de son accessibilité pour divers groupes d'utilisateurs. Traditionnellement, ce type de campagne vise une large diffusion et est souvent proposé gratuitement, mais sans confirmation explicite, il est difficile d'en juger.

Concernant les ressources offertes, Ban Deepfakes semble davantage orienté vers une campagne de sensibilisation qu'un outil technologique autonome. Bien que cela réduise la nécessité d'infrastructures techniques complexes, cela limite également sa capacité à fournir des solutions techniques robustes. La documentation disponible, bien que pertinente, reste insuffisante pour évaluer pleinement la mise en œuvre et l'efficacité des stratégies proposées.

D'autre part, l'organisation WITNESS dirige l'initiative Human Rights Campaigns & Projects, qui utilise les technologies, y compris la vidéo, pour documenter et exposer les violations des droits humains à travers le monde (Human Rights Campaigns & Projects from WITNESS, 2018). Ce projet, qui s'inscrit dans un cadre plus large d'activisme numérique, se distingue par sa capacité à autonomiser les individus pour documenter ces violations de manière crédible et persuasive. L'utilisation de la vidéo comme preuve visuelle, combinée à une formation rigoureuse, renforce l'impact des campagnes menées par WITNESS. Toutefois, l'efficacité de cette initiative dépend en grande partie des compétences des utilisateurs à appliquer ces technologies de manière éthique et sécurisée.

Ces deux initiatives, Ban Deepfakes et WITNESS, pourraient bénéficier d'un échange mutuel de pratiques : Ban Deepfakes pourrait intégrer une dimension éducative pour sensibiliser davantage à l'importance de la véracité dans les médias numériques, tandis que WITNESS pourrait s'inspirer des mesures préventives de Ban Deepfakes pour renforcer la crédibilité et la sécurité des contenus produits. Ensemble, elles pourraient offrir une réponse plus complète aux défis complexes posés par les technologies de manipulation et de désinformation.

Une autre initiative notoire est BioID Connect, lancée en collaboration avec le gouvernement allemand depuis 2020 (Atsüren, A, 2024). Cet outil technologique se concentre sur la sécurité des accès via l'authentification biométrique, offrant une protection robuste contre la fraude et l'usurpation d'identité. BioID Connect se distingue par sa compatibilité avec diverses plateformes et dispositifs, rendant la solution accessible via des applications mobiles, des navigateurs web, et intégrable dans des systèmes existants grâce à des API. Cette flexibilité est un atout majeur, bien que l'accessibilité soit conditionnée par la disponibilité de dispositifs capables de capturer des données biométriques, et par une infrastructure cloud fiable, ce qui peut poser des problèmes dans les régions où l'accès à Internet est limité.

Au Royaume-Uni, l'organisation Full Fact a développé l'initiative Full Fact AI, axée sur la vérification automatique des faits grâce à l'IA. Cet outil analyse et vérifie les informations circulant dans les médias, offrant des vérifications rapides et précises des affirmations publiques (The Full Fact Team – Full Fact, s. d) (All of the facts we have checked – Full Fact, s.d) (Full Fact Training – Full Fact, s.d). Conçu pour être intégré dans diverses plateformes médiatiques et sociales, Full Fact AI vise à être utilisé par les journalistes, chercheurs, et potentiellement le grand public. Cependant, son efficacité dépend de la qualité des données d'entraînement et de la complexité des algorithmes utilisés, tandis que son accessibilité pourrait être limitée par des barrières linguistiques et technologiques.

Enfin, Info Finder Facts from Africa in one place, une initiative d'Africa Check, se concentre sur la vérification des faits en Afrique, en offrant une plateforme centralisée où les utilisateurs peuvent accéder à des informations vérifiées sur divers sujets (Comment nous vérifions les faits Africa Check, s.d). Bien que cette initiative joue un rôle essentiel dans la lutte contre la désinformation en Afrique, elle repose principalement sur une vérification manuelle des faits, nécessitant une équipe dédiée de fact-checkers et de chercheurs. Info Finder pourrait bénéficier d'un partenariat avec Full Fact AI pour automatiser la vérification des faits et exploiter le potentiel de l'IA (Première organisation indépendante de vérification des faits en Afrique. Africa Check., s. d.).

D Recherches autour de la détection des hypertrucages

Le domaine scientifique et académique des hypertrucages est en pleine expansion, reflétant à la fois les avancées technologiques rapides et les préoccupations croissantes liées à leur utilisation. Les hypertrucages, qui exploitent l'IA pour créer des vidéos, images, ou sons hyperréalistes, posent des défis significatifs dans divers secteurs, notamment la sécurité, l'éthique, et la confiance du public. Les recherches actuelles se concentrent sur la détection, la régulation, et les implications sociétales de cette technologie, tout en explorant les possibilités d'innovation qu'elle offre dans des contextes créatifs et industriels. L'analyse des articles académiques révèle des contributions notables, mais aussi des limitations propres à chaque étude.

L'article intitulé « Deepfake Video Detection: Challenges and Opportunities » offre une analyse approfondie des enjeux liés aux jeux de données. Il classe les ensembles de données en trois générations : les premières générations comprenant moins de 1 000 vidéos, les deuxièmes générations avec environ 10 000 vidéos de haute qualité, et les troisièmes générations qui augmentent encore le nombre de vidéos et de frames (Kaur, A,2024). Les jeux de données spécifiques mentionnés incluent UADFV, Deepfake TIMIT, Face Forensics++, Google-DFD, Celeb-Df, DeeperForensics-1.0, le DeepFake Detection (DFD) Challenge, Wild-Deepfake, DF-W, OpenForensics, et DeepFakeFace (DFF). Chaque ensemble de données est évalué en fonction de sa diversité, de son réalisme, et de sa disponibilité, soulignant l'importance de disposer de jeux de données de haute qualité pour améliorer les méthodes de détection des hypertrucages (Kaur, A,2024). Cependant, il est important de noter que dans notre propre recherche, nous avons constaté que les outils mis à la disposition du grand public ne mentionnent souvent pas les jeux de données utilisés, ce qui a rendu difficile l'analyse de la performance de ces outils en fonction des jeux de données consommés.

L'article « Mapping the Deepfake Landscape for Innovation » propose une feuille de route articulée autour de plusieurs axes. Il met l'accent sur l'amélioration des techniques de détection des hypertrucages, le développement de cadres réglementaires pour encadrer leur utilisation, et l'exploration d'applications positives dans des domaines tels que le divertissement et l'éducation (Whittaker, L,2023). L'article souligne également l'importance de la collaboration interdisciplinaire entre chercheurs, régulateurs, et industries, ainsi que la sensibilisation et l'éducation du public pour garantir une utilisation responsable de cette technologie en pleine évolution (Whittaker, L,2023). Toutefois, bien que cette perspective multidisciplinaire soit précieuse, elle risque de devenir rapidement obsolète en raison de l'évolution rapide des technologies liées aux hypertrucages.

L'article intitulé « What We Know and Don't Know About Deepfakes: An Investigation Into the State of the Research and Regulatory Landscape » offre une analyse complète du phénomène des hypertrucages, en se concentrant sur l'état actuel de la recherche empirique et les réponses réglementaires émergentes (Birrer, A,2024). L'étude souligne que, bien que les hypertrucages soient souvent perçus comme une menace majeure, beaucoup de ces préoccupations reposent davantage sur des spéculations que sur des preuves empiriques solides. Le document passe en revue les défis spécifiques posés par les hypertrucages, tels que la désinformation, la pornographie non consensuelle, et la difficulté pour le public de détecter ces falsifications.

L'article met en lumière plusieurs aspects cruciaux pour renforcer les outils de lutte contre les hypertrucages, en établissant des liens clairs avec des initiatives existantes telles que Ban Deepfakes, WITNESS, et Full Fact AI. D'abord, l'amélioration de la détection des hypertrucages, y compris ceux de faible résolution, rappelle l'importance de technologies avancées comme celles utilisées par Full Fact AI, qui automatisent la vérification des faits et pourraient être adaptées pour identifier les caractéristiques subtiles des hypertrucages. Ensuite, l'accent mis sur le renforcement de la confiance du public dans les contenus authentiques souligne l'importance des initiatives éducatives, similaires à celles de WITNESS, qui sensibilisent les utilisateurs à reconnaître les manipulations tout en évitant une méfiance généralisée. L'article propose également des ajustements réglementaires, alignés sur l'approche de Ban Deepfakes, pour instaurer des fonctionnalités de signalement plus robustes et garantir une transparence accrue sur la provenance des contenus. Enfin, en traitant des impacts psychologiques des hypertrucages, notamment dans des cas de pornographie non consensuelle, l'article recommande l'intégration de mécanismes rapides de signalement et de suppression, appuyés par un soutien aux victimes, une approche qui pourrait inspirer des outils comme BioID Connect pour offrir une protection supplémentaire contre les abus en ligne. Ces insights fournissent une feuille de route pour développer des outils plus complets, efficaces, et responsables dans la lutte contre les hypertrucages.

Enfin, l'article intitulé « Deepfake Detection Using Deep Feature Stacking and Meta-Learning » explore une méthode novatrice pour détecter les hypertrucages en utilisant une approche basée sur l'empilement de fonctionnalités profondes (deep feature stacking) combinée à un apprentissage méta (meta-learning) (Naskar, G., 2024). L'étude utilise des modèles CNN, notamment Xception et EfficientNet-B7, pour extraire les caractéristiques des images, puis sélectionne les meilleures fonctionnalités avant de les classer avec un perceptron multicouche (MLP). Les résultats montrent une amélioration significative de la précision de la détection par rapport aux méthodes existantes, atteignant jusqu'à 98% d'exactitude sur les ensembles de données Celeb-DF et FaceForensics++. Le modèle se révèle robuste face aux manipulations des niveaux de luminosité et surpasse les méthodes de détection antérieures en termes de performance générale et de résistance aux attaques adversariales, démontrant ainsi la robustesse de la plateforme Sentinel.

3.3 Liste non-exhaustive des chercheurs et chercheuses travaillant sur le phénomène de l'hypertrucage

Nom	Coordonnées-réseaux sociaux	Pays	Centre de recherche	Spécialisation non exhaustive
Agarwal, Shruti	Shruti Agarwal (agarwalshruti15.github.io) Shruti Agarwal LinkedIn	États-Unis	Research Scientist for Content Authenticity Initiative at Adobe	Reconnue pour ses travaux sur la détection des hypertrucages et les techniques d'analyse multimédia, qui visent à identifier et contrer les manipulations numériques dans les images et les vidéos.
Amerini, Irene	Irene Amerini LinkedIn Irene Amerini (google.com)	Italie	Sapienza Université de Rome	Chercheuse spécialisée en vision par ordinateur, intelligence artificielle et médico-légale numérique. Elle est particulièrement connue pour ses travaux sur la détection des manipulations d'images et de vidéos, y compris les hypertrucages.
Badard, Thierry	Thierry Badard	Canada	Centre de recherche en données et intelligence géospatiale, Chaire de recherche industrielle en bases de données géospatiales décisionnelles	Professeur à l'Université Laval, il est également Directeur et chercheur régulier au Centre de recherche en données et intelligence géospatiale. Il s'intéresse notamment à l'hypertrucage en géographie.
Bestagini, Paolo	Paolo Bestagini Assistant Professor (polimi.it) (1) Paolo Bestagini LinkedIn	Italie	Politecnico di Milano	Spécialiste de l'authentification des contenus multimédias et de l'identification des hypertrucages.
Bourassa, Renée	Renée Bourassa	Canada	Centre de recherche sur les arts, les lettres et les techniques	Professeure à l'Université Laval, elle est vice-Directrice et membre du comité scientifique du Centre de recherche sur les arts, les lettres et les techniques. Elle fait de la recherche sur la désinformation par l'art et à la culture numérique.
Brin, Colette	Colette Brin	Canada	Centre d'études sur les médias, Obvia	Professeure à l'Université Laval, elle est aussi directrice du Centre d'études sur les médias et co-responsable de l'Axe Arts et Médias. Elle s'intéresse entre autres aux transformations des pratiques journalistiques entre autres causées par l'IA.
Chang, Ming-Ching	(albany.edu) Ming-Ching Chang LinkedIn	États-Unis et Taiwan	University at Albany Computer Vision and Machine Learning Lab (CVML) University at Albany	Reconnu pour ses travaux sur la surveillance vidéo, la détection d'objets, et l'analyse en temps réel de séquences vidéo.

Chen, Yu	Yu-Chen Chen LinkedIn	États-Unis	Researcher at York Centre for Asian Research	Doctorant en sciences politiques spécialisé dans l'autoritarisme transnational numérique, les politiques ethniques et d'immigration, les droits de l'homme, et la désinformation en ligne. Haut du formulaire Bas du formulaire
Farid, Hany	Hany Farid (berkeley.edu)	États-Unis	Université de Californie à Berkeley	Expert en criminalistique numérique et en vision par ordinateur, connu pour ses travaux sur la détection des images et vidéos truquées, y compris les hypertrucages.
Gerasimov, Artem	Artem Gerasimov	Émirats arabes unis	-	IA, Cybersécurité. PDG et fondateur de Oz Forensics.
Goodfello, Ian	Ian Goodfellow LinkedIn Ian Goodfellow – Wikipédia (wikipedia.org) Ian Goodfellow --- Research Page	États-Unis	Researcher at DeepMind	Chercheur en IA largement reconnu pour ses contributions dans le domaine de l'apprentissage profond (deep learning). Il est surtout connu pour avoir inventé les réseaux antagonistes génératifs (GANs, Generative Adversarial Networks), une innovation majeure dans le domaine de l'IA.
Groh, Matthew	Matt Groh	États-Unis	Human-AI Collaboration Lab (Northwestern University)	IA et éthique de l'IA. Travaille sur la collaboration entre humains et IA, l'éducation et le design centré sur l'humain.
Gupta, Rijul	Rijul Gupta LinkedIn	États-Unis	-	Rijul Gupta est le PDG de <i>DeepMedia</i> et un expert renommé en médias synthétiques, titulaire d'un diplôme en apprentissage automatique de l'Université de Yale. Forbes l'a désigné comme un leader d'opinion dans le domaine des hypertrucages. Il a créé DeepMedia dans le but de protéger la technologie d'hypertrucage contre les usages destructeurs et de promouvoir une utilisation éthique.
Lalonde, Jean-François	Jean-François Lalonde	Canada	Institut Intelligence et Données	Professeur à l'Université Laval et directeur scientifique adjoint de l'Institut Intelligence et Données, il a des intérêts de recherche pour l'apprentissage automatique et l'hypertrucage
Li, Hao	Hao Li - about me (hao-li.com) Hao Li LinkedIn	États-Unis	CEO & Co-Founder of Pinscreen et Associate Professor at MBZUAI	Chercheur, entrepreneur et professeur reconnu pour ses contributions dans le domaine de la vision par ordinateur, de l'infographie et de l'intelligence artificielle, en particulier dans le domaine des technologies liées aux hypertrucages et à la modélisation 3D des visages.
Li, Yuezun	Yuezun Li's Homepage	Chine	Groupe de recherche sur l'analyse visuelle et la sécurité de l'Université de l'océan de Chine	Hypertrucages, vision par ordinateur. Spécialiste de la détection des manipulations dans les médias numériques.

Lyu, Siwei	Siwei Lyu home page (buffalo.edu)	États-Unis	Center for information Integrity (CII) de l'Université de Buffalo	Hypertrucages, Vision par ordinateur. Dirige le UB Media Forensics Lab, spécialisé dans la détection des hypertrucages et co-directeur du <i>Center for Information Integrity</i> .
Michelot, Florent	Florent Michelot	Canada	Chercheur associé du GRIIPTIC Membre de l'Obvia	Professeur associé à l'université Concordia, il a un intérêt pour l'impact des technologies sur l'éducation et comment éduquer pour se prémunir face à la désinformation par l'IA.
Naffi, Nadia	Nadia Naffi	Canada	Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale, Obvia	Professeure agrégée à l'Université Laval, elle est titulaire de la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale et co-responsable de l'Axe Éducation et Capacitation à l'Obvia, Elle étudie entre autres l'éducation à l'hypertrucage et les impacts sociétaux de l'IA en éducation.
Patel, Yogesh	Yogesh Patel - IEEE Xplore Author Profile	Inde	Department of Computer Science and Engineering Institute of Technology Nirma University	Ses domaines de recherche incluent la blockchain, l'apprentissage fédéré, l'apprentissage automatique, et l'apprentissage profond.
Roy-Chowdhury, Amit	Amit K. Roy-Chowdhury Video Computing Group (ucr.edu) Amit Roy-Chowdhury LinkedIn	États-Unis	University of California, Riverside. Chair, Robotics Program Director, Center for Robotics and Intelligent Systems	Chercheur renommé dans le domaine de la vision par ordinateur, de l'intelligence artificielle (IA), et de l'apprentissage automatique. Il est particulièrement connu pour ses travaux sur l'analyse de la vidéo, la reconnaissance d'objets et de personnes, ainsi que sur la surveillance automatisée et la compréhension de scènes complexes à partir de séquences vidéo. Il dirige aussi un laboratoire de recherche qui se concentre sur divers aspects de la vision par ordinateur, de la robotique et de l'intelligence artificielle.
Shahriyari, Ali	Ali Shahriyari LinkedIn	États-Unis	Co-Founder and CTO – Reality Defender	Plus de 25 ans d'expérience dans l'industrie du développement de logiciels
Tariq, Shahroz	Shahroz Tariq (researchgate.net) (1) Shahroz Tariq, PhD LinkedIn	Australie	CSIRO's Data61	Hypertrucage, apprentissage machine. Expert en médico-légal numérique et détection de l'hypertrucage.
Theobalt, Christian	Christian Theobalt :: MPI Informatik (mpg.de)	Allemagne	Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence	Chercheur dans le domaine de l'infographie, de la vision par ordinateur et de l'intelligence artificielle, avec un accent particulier sur la reconstruction 3D, la capture de mouvement, et la modélisation des visages et du corps humain.
Warin, Thierry	Thierry Warin	Canada	CIRANO	Professeur à HEC Montréal, il est chercheur principal en Science des données et co-responsable du grand thème Économie mondiale au CIRANO. Il fait entre autres de la recherche sur la désinformation à l'ère du numérique.

Wu, Y. Kelly	Y. Kelly Wu LinkedIn Y. KELLY WU KellyWuuu	États-Unis	Rochester Institute of Technology	Doctorant en informatique. Champs d'intérêt : Détection des hypertrucages, Intelligence Artificielle Explicable (XAI), Interaction Homme-Machine (HCI), Vision par Ordinateur, Désinformation.
Wu, Zhizheng	Zhizheng Wu (drwuz.com)	Chine	Chinese University of Hong Kong	Zhizheng est le créateur de Merlin, un kit d'outils de synthèse vocale open-source. Il a initié et co-organisé le premier défi de vérification d'identité vocale et les mesures contre les falsifications en tant que session spéciale à Interspeech 2015, ainsi que le <i>Voice Conversion Challenge</i> 2016 et le <i>Blizzard Challenge</i> 2019.
Zhang, Weiming	Le groupe de Weiming Zhang (ustc.edu.cn)	Chine	University of Science and Technology of China	Il est actuellement professeur à l'Université des sciences et technologies de Chine, vice-président de l'École de sécurité du cyberspace et secrétaire général adjoint du Comité d'experts en criminalistique et sécurité multimédia de la Société pour l'image et le graphisme. Ses principaux intérêts de recherche comprennent la dissimulation d'informations, la protection de la confidentialité des données et la sécurité de l'IA.





Discussion

Une discussion sera ici entamée pour chacun des trois volets de ce rapport : les ressources hautement partagées sur les réseaux sociaux concernant les cas d'utilisation de l'hypertrucage, une analyse des défis juridiques liés à leur régulation, et une exploration des techniques de création et de détection des hypertrucages.

Volet des ressources hautement partagées sur les réseaux sociaux

Les hypertrucages sont devenus une menace croissante dans divers domaines, touchant particulièrement les femmes et les personnes âgées. Les tendances observées montrent que cette technologie est utilisée à des fins variées, allant de l'escroquerie financière à la création de contenus pornographiques non consensuels, en passant par l'usurpation d'identité et la manipulation de l'opinion publique.

Les hypertrucages sont utilisés dans plusieurs domaines, notamment la finance, la politique, le social ainsi que le divertissement et le sport. Dans le domaine financier, les hypertrucages sont souvent utilisés pour des escroqueries, visant à tromper les individus afin d'obtenir des informations sensibles ou de l'argent. Dans le domaine politique, les hypertrucages permettent de manipuler l'opinion publique et de nuire à la réputation de certaines personnalités politiques. Le domaine social est également fortement touché, notamment par la création de contenus pornographiques non consensuels, causant des préjudices émotionnels et sociaux importants. Enfin, dans le domaine du divertissement, les hypertrucages sont utilisés pour créer des contenus trompeurs, souvent à des fins de harcèlement ou de diffamation.

Les personnes âgées et les femmes sont particulièrement vulnérables aux hypertrucages. Les personnes âgées sont souvent ciblées par des escroqueries financières utilisant des hypertrucages. Ces escroqueries exploitent la faible littératie numérique des aînés et leur confiance souvent plus élevée envers les communications qu'ils reçoivent. Les femmes, quant à elles, sont souvent ciblées par des hypertrucages à des fins de harcèlement, de diffamation et de création de contenus pornographiques non consensuels. Ces attaques exploitent souvent l'image publique des femmes pour créer des contenus trompeurs et nuisibles.

Les enjeux liés aux hypertrucages sont nombreux et complexes. La difficulté de distinguer les contenus authentiques des contenus manipulés pose un défi majeur pour la protection des individus et la lutte contre les abus. Les angles morts actuels incluent le manque de réglementation spécifique pour encadrer l'utilisation des hypertrucages et la nécessité de développer des technologies de détection plus avancées, comme nous l'approfondirons dans les volets 2 et 3 de ce rapport. De plus, il est crucial de sensibiliser le public aux dangers des hypertrucages et de mettre en place des mesures de protection pour les personnes les plus vulnérables, un rôle essentiel que devra jouer l'éducation dans le futur.

Les hypertrucages sont devenus une menace croissante dans divers domaines, touchant particulièrement les femmes et les personnes âgées.

Pour pallier ces limites, plusieurs pistes de recherche et d'action doivent être explorées. Il est essentiel de développer des outils de détection des hypertrucages plus robustes et accessibles. La recherche doit également se concentrer sur l'élaboration de cadres réglementaires adaptés pour encadrer l'utilisation des hypertrucages et protéger les individus contre les abus. Il est important de noter que les cas documentés, principalement rapportés par les médias et visibles sur les réseaux sociaux ne représentent que la partie émergée de l'iceberg. De nombreux incidents restent méconnus ou non signalés, soulignant l'importance d'une surveillance accrue. Dans ce contexte, il est crucial que des organisations spécialisées documentent activement la désinformation liée aux hypertrucages et mènent des efforts de sensibilisation et de débunkage. Une approche proactive est nécessaire, impliquant non seulement la réaction aux cas existants, mais aussi l'anticipation des potentielles instances de désinformation. Cette vigilance constante doit s'accompagner d'une communication claire auprès du public sur les risques émergents et les menaces potentielles. Enfin, des campagnes de sensibilisation et d'éducation doivent être mises en place pour informer le public des risques associés aux hypertrucages et des moyens de se protéger.

Volet juridique

A Mouvement législatif plus rapide à plus petite échelle et hésitation des hautes instances

Au cours de ce projet, nous avons réalisé que la plupart des lois adoptées portant sur les hypertrucages l'ont été à une échelle plus « locale » que « nationale ».

Au regard des hypertrucages à caractère intime par exemple, diffusés de manière non consensuelle, alors que le projet de loi C-63 (Loi édictant la Loi sur les préjudices en ligne, modifiant le Code criminel, la Loi canadienne sur les droits de la personne et la Loi concernant la déclaration obligatoire de la pornographie juvénile sur Internet par les personnes qui fournissent des services Internet et apportant des modifications corrélatives et connexes à d'autres lois), au Canada, n'a toujours pas été adopté au moment d'écrire ces lignes, il appert que, de leur côté, l'Alberta, l'Île-du-Prince-Édouard, le Manitoba, le Nouveau-Brunswick et la Saskatchewan ont procédé à l'adoption de dispositions législatives portant sur ce sujet. Les provinces semblent donc plus à même de légiférer rapidement que ne l'est le fédéral sur ce point.

Aux États-Unis, le même constat s'observe : tandis que le « *Preventing Deepfakes of Intimate Images Act* » et le « *Bill to require covered platforms to remove nonconsensual intimate visual depictions, and for other purposes* » ne sont toujours pas adoptés présentement, le Dakota du Sud « *Act to revise provisions related to the possession, distribution, and manufacture of child pornography* » et le Texas « *Act relating to the unlawful production or distribution of sexually explicit videos using deep fake technology; creating a criminal offense* », pour ne mentionner que ces deux exemples, ont choisi d'aller de l'avant en adoptant des lois contenant des dispositions au sujet des hypertrucages à caractère sexuel.

Cette réflexion vaut également en ce qui a trait à l'action des tribunaux. Ainsi, si plusieurs cours de justice provinciales ont émis au cours de l'année 2023 des directives dans le but de tenter de sensibiliser les avocats quant à la garantie d'intégrité, à la véracité et à l'exactitude des observations judiciaires et des jurisprudences et autres documentations dont ils font usage devant les tribunaux s'ils ont recours à l'IA, il aura fallu attendre au 20 décembre 2023 pour que la cour fédérale du Canada fasse de même.

B Divergences conceptuelles

Le concept d'IA, en soi, n'est pas unanimement défini, bien que certains points communs semblent ressortir de définitions proposées. Toutefois, l'IA n'est pas le seul terme qui laisse transparaître une certaine difficulté définitionnelle. Tel est aussi le cas du mot « désinformation » par exemple. Ainsi, « la politique de l'Union européenne en matière de désinformation repose sur l'idée que la désinformation n'est pas illégale en soi, mais qu'elle est nuisible » (Ó Fathaigh et al., 2021, p. 2 – traduction libre). Suivant la même ligne de pensée, « un groupe d'experts indépendants de haut niveau de l'Union européenne sur les informations fausses et la désinformation en ligne (HLEG) a souligné que la désinformation n'est pas nécessairement illégale, mais qu'elle peut néanmoins être préjudiciable aux citoyens et à la société dans son ensemble ».

En contrepartie, des auteurs observaient en 2021 que certains « États membres de l'Union européenne pourraient faire de la désinformation une infraction illégale » et déploraient qu'il y ait « un manque d'études approfondies parmi les études juridiques de l'Union européenne axées sur les définitions actuelles de la désinformation et sur la question de savoir si la législation nationale des États membres de l'Union européenne peut effectivement appliquer ces définitions de la désinformation » (Ó Fathaigh et al., 2021, p. 2 – traduction libre). On sentait donc, au sein même de l'Union européenne par exemple, qu'il existait une volonté des États membres de traiter la désinformation comme étant illégale, mais une hésitation de l'Union à emboîter le pas en ce sens. À notre connaissance, cette situation ne semble pas avoir significativement évolué depuis cette publication, bien qu'une analyse plus approfondie et actualisée de la législation européenne pourrait révéler des développements récents en la matière.

C Considération relative à la liberté d'expression

La note explicative accompagnant le projet de loi « *Act defining deepfake and the imposition of penalties therefore for acts constituting harmful deepfakes* », émanant des Philippines, présente en parallèle l'hypertrucage et la liberté d'expression en mentionnant notamment que :

« La section 4, article III de la Constitution de 1987 garantit le droit à la liberté de parole et d'expression. Toutefois, l'État doit s'engager à agir efficacement pour condamner et combattre l'utilisation inappropriée et malveillante des technologies, en particulier dans les cas de diffusion de désinformation, de harcèlement, d'incitation à la violence et de fraude à l'identité » (Congress of the Republic of the Philippines, 2023 – traduction libre).

Pour compléter le point précédent portant sur les « embûches conceptuelles » et comme l'indiquent certains auteurs, le fait d'ériger la désinformation, au sens général, au rang de crime peut « conduire à une autocensure et soulever des inquiétudes particulières en ce qui concerne la liberté d'expression » (Ó Fathaigh et al., 2021, p. 12 – traduction libre). De l'avis de certains experts, « les mesures de lutte contre la désinformation doivent être formulées avec soin afin qu'elles ne puissent pas être utilisées abusivement par les gouvernements pour justifier la censure du journalisme ou la critique » (Woolf, 2022). Qui plus est, tel qu'indiqué dans le projet de loi philippine susmentionné :

Les deepfakes, en eux-mêmes, ne sont pas intrinsèquement illégaux. En fait, ils offrent de précieux potentiel d'applications positives. Par exemple, Cereproc utilise des deepfakes pour créer des voix numériques pour les enfants handicapés ou ceux qui ont perdu la capacité de parler en raison d'une maladie. (Congress of the Republic of the Philippines, 2023 – traduction libre).

Conséquemment, proscrire en bloc la désinformation et les hypertrucages, de manière large et englobante, engendrerait des préoccupations juridiques importantes. Dans les faits, bien que l'intention première consiste possiblement à limiter la propagation de « fake news » aux conséquences potentiellement importante ou grave, il n'en demeure pas moins qu'« il est parfois compliqué de démontrer qu'une information est vraie ou fausse ou de distinguer une fausse information d'une simple opinion », voire même d'une plaisanterie (Lequeux, 2024). La plupart conviendra que « certaines [« fake news »] peuvent être anodines [...], [mais] d'autres peuvent être bien plus dangereuses » (Lequeux, 2024). De fait, prohiber de façon claire certains éléments de manière circonscrite tout en tolérant d'autres « types de désinformation » qui apparaissent d'importance moindre n'est pas une mince affaire, d'un point de vue législatif et définitionnel. Il faut être conscient, en fait, que « la légalité des deepfakes dépend de leur contenu, car ils peuvent enfreindre les droits d'auteur, violer la protection de données, diffamer des individus et empiéter sur la vie privée » (Congress of the Republic of the Philippines, 2023 – traduction libre). Or, faire dépendre la légalité des contenus entraîne une appréciation au cas par cas qui peut être à la fois exigeante pour les tribunaux, engendrer des exemptions possibles, en plus de donner lieu à une part de subjectivité dans l'appréciation des faits. Tel que réalistement décrit dans un projet de loi philippine : « Résoudre les complexités posées par les deepfakes nécessite un équilibre délicat. À ce titre, ce projet de loi introduit plusieurs exceptions relatives à des actes qualifiés de deepfakes nuisibles » (Congress of the Republic of the Philippines, 2023 – traduction libre).

D Innovation versus risques liés à l'IA

Le caractère innovant de l'IA a, à quelques reprises, été mis de l'avant dans nos lectures. D'ailleurs, à ce sujet, le Parlement européen estime que :

« La sous-exploitation de l'IA est perçue comme un risque majeur [, notamment] si l'UE venait à rater certaines opportunités [...]. Cela pourrait aussi entraîner une perte d'avantages concurrentiels par rapport à d'autres régions du monde ou mener à une stagnation économique » (Parlement européen, 2023).

De fait, au cours de nos travaux, il est arrivé que nous ressentions une forme de « dualité » entre, d'une part, l'aspiration au progrès et à l'innovation que peut supposer l'utilisation de l'IA et, d'autre part, l'encadrement juridique des technologies afin de tenter de diminuer les risques en découlant.

De fait, le Parlement européen en vient à écrire que la « réglementation [concernant l'IA] pourrait devenir trop stricte et étouffer toute innovation » dans certains cas (Parlement européen, 2023), et, de son côté, Monsieur Clyde Shavers (parti démocrate), représentant de l'État à l'Assemblée législative de l'État de Washington, défend le projet de loi H.B. 1951 -2024 (Act Relating to promoting ethical artificial intelligence by protecting against algorithmic discrimination [...]) en mentionnant notamment : « L'IA est une question complexe que nous ne comprenons pas entièrement. Vous ne voulez pas arrêter l'innovation. Vous ne voulez pas arrêter le commerce. Mais vous ne voulez pas entraver les droits civiques des gens. ... En ce moment, nous sommes à un moment où l'IA est partout » (Shavers, cité dans Stang, 2024 – traduction libre).

Au Canada, le projet de loi C-27 sur l'IA et les données (LIAD) (Parlement du Canada, 2022) s'est aussi fait critiquer notamment parce que « une [de ses] mesures phares [...] [consiste] à laisser à l'industrie de l'IA au Canada le rôle de s'autoréguler »; une orientation législative que certains contestent en se demandant si ce projet de loi est « teinté pour soutenir l'innovation » (Normandin, 2023).

Soulignons, par ailleurs, qu'un même outil doté d'IA (ex. : technologie d'OpenAI) peut à la fois être générateur d'avancées notoires (Dupuis, 2024) et de dérives inquiétantes (De Rosa, 2024).

Ces propos tendent à fournir un exemple attestant de cette présentation souvent dualiste entre le désir de ne pas se priver des avancées que peut générer l'IA et ce ressenti imposant d'appuyer sur les freins pour considérer les risques relatifs à son usage. C'est donc à la croisée de ces chemins que s'inscrit le souhait de légiférer ou d'encadrer juridiquement l'emploi de l'IA, ce qui ajoute au caractère complexe du travail du législateur.

E Agir « en réponse à »

Comme l'exposent certains cas de figure (ex. : hypertrucages ayant visés des personnalités politiques ou des vedettes et l'affaire Zhang v. Chen, 2024 BCSC 285 (CanLII) (CanLII, 2024)), les mesures juridiques jusqu'à présent adoptées semblent être prises davantage dans un dessein réactif que proactif. Cela nous mène à nous demander si le droit est, en quelque sorte, à « la remorque » des incidents susceptibles d'être rencontrés en lien avec un usage malveillant de l'IA. Par exemple, dans l'affaire Zhang v. Chen, 2024 BCSC 285 (CanLII), il aura fallu qu'une avocate dépose une requête contenant des autorités juridiques inexistantes ayant été « hallucinées » et étant le fruit de la « créativité » de l'IA avant que les cours de justice canadiennes ne proposent publiquement un encadrement juridique dans le but de faire contrepoids à ce type de problématique et d'encadrer l'usage que peuvent faire les professionnels du droit de l'IA dans le cadre de la représentation de leurs clients.

F Le caractère évolutif et donc incertain des usages malicieux futurs possibles

Dans un document complémentaire relatif à la Loi sur l'IA et les données (LIAD), le gouvernement du Canada écrit notamment :

« À mesure que la technologie évolue, de nouvelles capacités et utilisations des systèmes d'IA émergeront, et nous avons besoin d'un cadre capable de s'adapter à l'évolution du paysage. Le gouvernement adopte une approche agile sur la réglementation de l'IA basé sur l'élaboration et l'évaluation de règlements et de lignes directrices selon un cycle régulier en étroite collaboration avec les parties prenantes, ainsi que l'adaptation de l'application aux besoins de l'environnement dynamique ».

De fait, au fil de nos recherches, une préoccupation constante consistant à encadrer juridiquement l'usage de l'IA, la diffusion d'hypertrucages et la désinformation, tout en fournissant un cadre évolutif fut ressentie. Ainsi, les instances législatives nous apparaissent conscientes que les textes de loi doivent être en mesure d'évoluer de manière, idéalement, à ne pas accuser de retard et à ne pas souffrir de « zones grises » ou d'« angles morts » au regard du développement technologique rapide. Or, quiconque connaît la rigidité que peut supposer parfois la rédaction de certains textes de droit, qui se veulent à la fois encadrants, clairs et efficaces, saura reconnaître que ce désir pourtant légitime d'évolutivité législative représente un défi de taille quant aux thématiques abordées dans le présent rapport et d'autres thèmes et enjeux découlant de l'IA. Comment trouver le parfait équilibre entre un texte de loi sans ambiguïté quant aux infractions qu'il crée, tout en se voulant assez vaste, englobant et un brin volontairement flou et flexible pour couvrir des situations inconnues au moment de son adoption ?

G Voir ce qui est fait ailleurs

Comme le souligne à juste titre le gouvernement canadien : « L'interconnexion mondiale de l'économie numérique exige que la réglementation des systèmes d'IA sur le marché soit coordonnée à l'échelle internationale » (Gouvernement du Canada, 2023). De fait, on sent, à travers les textes juridiques adoptés pour encadrer l'emploi de l'IA, une certaine volonté de ne pas s'inscrire en « porte-à-faux » avec des encadrements juridiques étrangers.

Cela dit, soyons réalistes : outre ce dessein de relative « uniformisation » des pratiques juridiques internationales, plus importante encore est possiblement cette volonté de consulter des textes législatifs adoptés ailleurs dans le but de s'en inspirer dans la législation interne. Sous le poids des regards comparatifs de certains qui diront par exemple qu'au Canada, « le cadre législatif du projet de loi C-27 [Loi sur l'IA et les données et apportant des modifications corrélatives et connexes à d'autres lois] devrait s'inspirer de la proposition européenne » notamment en matière d'hypertrucages, mais aussi en regardant ce qui fonctionne ou non dans les textes de loi des homologues internationaux, les législateurs, où qu'ils soient, bénéficieront sans doute des leçons acquises par ceux qui les auront précédés dans l'adoption de lois encadrant l'IA, les hypertrucages et la désinformation.

Volet technologique

La lutte contre les hypertrucages révèle plusieurs paradoxes et limitations significatives dans les approches actuelles. Bien que les entreprises technologiques comme Intel avec FakeCatcher et Microsoft avec Video Authenticator investissent massivement dans le développement de technologies de détection, ces solutions présentent une contradiction fondamentale : elles sont développées par les mêmes entreprises qui créent les technologies permettant la génération d'hypertrucages. Cette dualité soulève des questions sur les potentiels conflits d'intérêts et la sincérité de leur engagement dans la lutte contre ce phénomène. De plus, la privatisation de ces technologies de détection, souvent inaccessibles au grand public, crée une disparité préoccupante entre ceux qui peuvent se protéger et ceux qui restent vulnérables.

L'approche gouvernementale, bien qu'ambitieuse avec des initiatives comme WeVerify en Europe ou les réglementations de la Cyberspace Administration of China, souffre d'un manque de coordination internationale. Les efforts disparates des différents pays créent des zones grises exploitables par les créateurs d'hypertrucages. Aux États-Unis, malgré l'implication d'agences comme la DARPA et le NIST, l'efficacité réelle des mesures mises en place reste difficile à évaluer. La fragmentation des législations entre les États complique davantage la situation, créant un patchwork de réglementations potentiellement inefficace face à une menace qui ignore les frontières géographiques.

Les organismes à but non lucratif comme Ban Deepfakes et WITNESS, bien que louables dans leurs intentions, se heurtent à des limitations structurelles importantes. Leurs initiatives de sensibilisation et de prévention, bien que nécessaires, risquent d'être insuffisantes face à la sophistication croissante des hypertrucages. Le décalage entre la rapidité d'évolution des technologies de création d'hypertrucages et le temps nécessaire pour former le public à les détecter crée une vulnérabilité chronique.

La recherche académique, souvent présentée comme un pilier essentiel dans la lutte contre les hypertrucages, fait face à plusieurs défis majeurs. Les méthodes de détection développées deviennent rapidement obsolètes face à l'évolution rapide des techniques de création d'hypertrucages. De plus, la course aux publications et aux financements peut parfois conduire à une surenchère d'annonces de « solutions » dont l'efficacité réelle reste à démontrer. La collaboration interdisciplinaire, bien que souhaitable, se heurte souvent à des barrières institutionnelles et des divergences méthodologiques.

Un déséquilibre fondamental existe entre les ressources nécessaires pour créer et pour détecter les hypertrucages. Alors que la création devient de plus en plus accessible et peu coûteuse, la détection nécessite des investissements considérables en infrastructure, en recherche et en expertise. Cette asymétrie crée une situation où les défenseurs sont constamment en retard sur les créateurs, un problème exacerbé par la démocratisation rapide des outils de création d'hypertrucages.

La fragmentation des efforts de lutte pose également un problème de standardisation. L'absence d'une méthodologie universellement acceptée pour évaluer l'efficacité des outils de détection rend difficile la comparaison des différentes solutions proposées. Les entreprises, les gouvernements et les chercheurs utilisent souvent des métriques différentes pour mesurer le succès de leurs initiatives, ce qui complique l'identification des approches réellement efficaces. Cette situation est particulièrement problématique quand on considère que les jeux de données utilisés pour développer et tester les outils de détection peuvent rapidement devenir obsolètes face à l'évolution constante des techniques de création.

La dimension géopolitique mérite également une attention particulière. Les différences d'approche entre les pays créent des opportunités d'arbitrage réglementaire, où les acteurs malveillants peuvent exploiter les juridictions les plus permissives. Cette situation est particulièrement préoccupante dans le contexte des campagnes de désinformation transfrontalières, où l'absence d'une coordination internationale efficace peut compromettre les efforts nationaux de régulation.

Sur le plan sociétal, la prolifération des hypertrucages contribue à une érosion plus large de la confiance dans les médias numériques. Le phénomène du « liar's dividend » – où la simple existence des hypertrucages permet de jeter le doute sur des contenus authentiques – pourrait avoir des implications dévastatrices pour le discours public et la démocratie. Les solutions techniques actuelles ne répondent pas adéquatement à cette dimension sociale du problème.



La question de l'accessibilité des solutions pose également un problème d'équité. Les outils de détection les plus sophistiqués restent largement inaccessibles aux populations les plus vulnérables aux hypertrucages, notamment dans les pays en développement ou parmi les groupes socialement marginalisés. Cette situation crée un cercle vicieux où ceux qui ont le plus besoin de protection sont les moins bien équipés pour se défendre.

Les implications à long terme pour la vie privée sont également préoccupantes. Les systèmes de détection des hypertrucages nécessitent souvent un accès approfondi aux données personnelles et aux métadonnées des contenus numériques, créant ainsi un compromis problématique entre protection et confidentialité. Cette tension n'est pas suffisamment adressée dans les approches actuelles.

Enfin, l'impact environnemental de la course aux armements technologiques contre les hypertrucages mérite d'être souligné. Le développement et le déploiement d'outils de détection de plus en plus sophistiqués nécessitent des ressources computationnelles importantes, contribuant à l'empreinte carbone croissante du secteur numérique. Cette dimension écologique est largement absente des discussions sur les stratégies de lutte contre les hypertrucages.

Ces observations soulignent la nécessité d'une refonte fondamentale de notre approche de la lutte contre les hypertrucages. Au lieu de se concentrer uniquement sur des solutions techniques toujours plus sophistiquées, il serait plus productif d'explorer des approches préventives et systémiques qui s'attaquent aux conditions sociales, économiques et politiques qui facilitent la prolifération des hypertrucages. Cette approche holistique nécessiterait une collaboration internationale renforcée, une démocratisation des outils de protection, et une plus grande attention aux dimensions éthiques et sociales du problème.



Limites

Certaines limites ont été remarquées principalement au niveau juridique et technologique.

Sur le plan juridique, effectuer un recensement international de documents à contenu juridique (projets de loi, lois, directives émises par des tribunaux, etc.), adopté ou en cours d'élaboration, sur une thématique spécifique, s'avère être un défi à certains égards. Plusieurs aspects confirment le caractère ambitieux de ce projet.

D'abord, comme c'est sans doute le cas pour la plupart des recherches à portée internationale, bien qu'Internet puisse faciliter le partage et la consultation de textes de loi pertinents, la non-uniformisation des plateformes de diffusion et des normes de présentation de l'information nécessite une certaine adaptation de la part des recenseurs. Une plateforme peut, par exemple, diffuser un projet de loi sans préciser quel en est l'état d'avancement de manière actualisée : en est-il au stade de sa présentation ? En première ou deuxième lecture ? Son analyse a-t-elle été relayée à un comité ? A-t-il finalement été adopté ? Il s'agit-là d'autant de questions, fort pertinentes, qui ne trouvent pas forcément réponse facilement lors de la consultation de plateformes Web diffusant du contenu juridique.

Qui plus est, il importe de choisir consciencieusement les plateformes Internet consultées. En effet, il est nécessaire, spécialement dans le cadre de ce type de projet où le droit est appelé à évoluer rapidement, de porter attention à la valeur officielle des documents publiés et à l'exactitude et l'actualisation des informations qui s'y trouve. Une personne néophyte du domaine juridique pourrait, par exemple, se sentir perdue à travers les diverses versions amendées d'un même projet de loi auxquelles il est possible d'accéder par Internet. Elle pourrait également se méprendre sur le statut d'un document. Sur ce dernier point, précisons par exemple que ce n'est pas parce qu'un projet de loi fut adopté et est, par le fait même, devenu loi, qu'il est pour autant en vigueur. Il n'est pas rare en effet que la date d'entrée en vigueur d'une loi soit différée dans le temps.

Le volume de documents consultés s'avère également impressionnant, afin d'effectuer un tour d'horizon qui, sans prétention à l'exhaustivité, se veut le plus complet possible. Parallèlement, mentionnons que la limite relative au temps alloué afin de réaliser le projet et le présent rapport sont à considérer au chapitre des contraintes rencontrées.

De même, bien que des éléments clés de contenu soient relativement similaires d'une région géographique à l'autre, il appert que certaines particularités se dégagent, notamment en ce qui a trait aux définitions préconisées pour un même terme, aux sanctions et aux exceptions législativement prévues en lien avec une infraction donnée. Il importe d'être vigilant afin de déceler ces différences qu'il est possible d'observer et de pouvoir mettre celles-ci en relief, dans un dessein comparatif. Cela dit, ajoutons que l'objectif du présent rapport ne tend pas à offrir une perspective de droit comparé, mais plutôt un portrait global de diverses législations répertoriées.

Certaines limites
ont été remarquées
principalement au niveau
juridique et technologique.

Enfin, l'IA ayant un caractère évolutif très rapide, il n'est pas impossible de penser qu'il puisse survenir des cas où un texte consulté au cours de la période estivale, alors qu'il n'était encore qu'un projet de loi, puisse rapidement devenir une loi officielle au cours de la saison automnale 2024 ou hivernale 2025, selon l'état d'avancement des travaux législatifs.

Cela dit, autant le recensement de lois portant sur les hypertrucages et la désinformation présente un certain niveau de difficulté, autant il s'avère être un défi et une occasion de clarifier l'état actuel du droit sur ces thèmes qui font couler beaucoup d'encre en ce moment.

Sur le plan technologique, plusieurs limites ont été constatées. Au niveau de l'accessibilité et l'efficacité des efforts déployés par les entreprises et les développeurs, nos recherches montrent qu'il est souvent plus facile d'accéder à des outils de création des hypertrucages performants qu'à des outils de détection tout aussi efficaces. Cette disparité souligne le besoin urgent de développer et de démocratiser les technologies de détection pour contrer les menaces croissantes posées par les hypertrucages. Il est essentiel que les efforts de recherche et les collaborations industrielles continuent de se concentrer sur l'amélioration de ces technologies, afin de garantir que la société puisse se défendre contre les abus potentiels de cette technologie.

Pour ce qui est des pays et gouvernements, il est à noter que nos recherches n'ont pas identifié d'initiatives spécifiques au niveau du continent africain, en dehors de celles menées par des organisations non gouvernementales et des entités à but non lucratif.

En ce qui concerne les organismes à but non lucratif, ceux-ci jouent un rôle crucial dans la lutte contre la désinformation et les hypertrucages, chacune adoptant des approches distinctes pour répondre à ces défis. Que ce soit par des campagnes de sensibilisation, l'autonomisation des individus, ou le développement de technologies avancées, ces initiatives démontrent la diversité des stratégies disponibles. Cependant, l'intégration de ces différentes approches pourrait renforcer l'efficacité globale des efforts pour combattre la manipulation et la désinformation dans le monde numérique. En combinant éducation, prévention et technologie, ces organisations peuvent proposer une réponse plus complète et adaptée aux enjeux actuels.

En lien avec les recherches sur l'hypertrucage, ce domaine de recherche est à la croisée des chemins entre innovation technologique et défis sociétaux. Les études récentes montrent que, bien que des progrès significatifs aient été réalisés dans la détection et la compréhension des hypertrucages, des défis subsistent, notamment en matière de disponibilité des jeux de données, de réglementation, et d'éducation publique. Les recherches futures devront continuer à s'appuyer sur une collaboration interdisciplinaire et une adaptation rapide aux évolutions technologiques pour relever ces défis de manière efficace et éthique.



Conclusion

La menace des hypertrucages représente plus qu'une simple dérive technologique : elle incarne une crise de confiance sans précédent dans notre société numérique. Notre analyse approfondie, menée selon trois axes – médiatique, juridique et technologique – révèle un constat alarmant : nous sommes en train de perdre la bataille contre la désinformation artificielle, non par manque de moyens techniques, mais par l'absence d'une vision cohérente et équitable de la lutte. Cette défaillance systémique menace non seulement notre capacité à distinguer le vrai du faux, mais aussi les fondements mêmes de notre contrat social à l'ère numérique.

L'analyse de 125 ressources médiatiques dresse un portrait inquiétant où les hypertrucages, loin d'être de simples curiosités technologiques, se transforment en véritables armes de destruction sociale. Nos recherches ont révélé une diversification croissante des usages malveillants, allant des escroqueries financières sophistiquées à la manipulation de l'opinion publique, en passant par la création de contenus pornographiques non consensuels. Plus troublants encore, les cas documentés ne représentent que la partie visible d'un phénomène bien plus vaste. Cette réalité est particulièrement préoccupante pour les populations vulnérables, notamment les femmes et les personnes âgées, qui se retrouvent ciblées de manière disproportionnée. Cette vulnérabilité accrue révèle comment les hypertrucages amplifient et perpétuent les inégalités sociales existantes, créant un cercle vicieux où les plus démunis deviennent les cibles privilégiées d'attaques toujours plus sophistiquées.

Notre analyse juridique expose un paradoxe fondamental qui compromet l'efficacité de toute réponse légale : alors que la technologie ignore les frontières, nos réponses légales restent fragmentées et territorialisées. Les législateurs, prisonniers de cadres juridiques obsolètes, tentent de réguler une technologie qui évolue plus rapidement que leur capacité à légiférer. Cette inadéquation crée des zones grises juridiques où prospèrent les acteurs malveillants. Plus préoccupant encore, la tension permanente entre innovation technologique et protection des individus conduit souvent à des compromis législatifs qui affaiblissent la portée des mesures de protection. Les tentatives de régulation actuelles, bien qu'ambitieuses, souffrent d'un manque de coordination internationale et d'une approche trop souvent réactive plutôt que proactive.

Sur le plan technologique, nous faisons face à une asymétrie dangereuse qui menace l'équilibre démocratique de nos sociétés numériques. La démocratisation rapide des outils de création d'hypertrucages contraste violemment avec l'accès restreint aux technologies de détection efficaces. Cette situation crée une société à deux vitesses où seuls les plus privilégiés peuvent se protéger efficacement, pendant que la majorité reste exposée aux manipulations numériques. Les solutions technologiques actuelles, souvent développées par les mêmes entreprises qui créent les outils de génération d'hypertrucages, soulèvent des questions éthiques fondamentales sur les conflits d'intérêts inhérents à cette situation.

Notre analyse approfondie, menée selon trois axes – médiatique, juridique et technologique – révèle un constat alarmant : nous sommes en train de perdre la bataille contre la désinformation artificielle, non par manque de moyens techniques, mais par l'absence d'une vision cohérente et équitable de la lutte.

Au cœur de ces défis complexes, l'éducation émerge non pas comme une simple solution parmi d'autres, mais comme le fondement nécessaire de toute stratégie efficace. Pourtant, elle reste paradoxalement le parent pauvre des initiatives actuelles, recevant moins d'attention et de ressources que les solutions techniques. Cette mise en arrière-plan de l'aspect humain de la problématique constitue peut-être notre plus grande erreur stratégique. Une approche éducative robuste et systématique doit devenir la pierre angulaire de notre réponse collective, en commençant par le développement précoce de la littératie numérique et de la littératie d'IA, et en s'étendant à la formation continue des personnes professionnelles et à l'éducation citoyenne. Cette dimension éducative doit également intégrer une composante éthique forte, sensibilisant aux impacts sociétaux des hypertrucages et développant une conscience critique et une agentivité numérique face aux contenus numériques.

Les implications de notre recherche sont profondes et multidimensionnelles. Continuer sur la voie actuelle ne fera qu'accélérer la création d'une société numérique profondément inégalitaire, où la capacité à distinguer le vrai du faux devient un privilège plutôt qu'un droit. La véritable réponse aux hypertrucages ne peut se limiter à une course aux armements technologiques ; elle nécessite une refonte fondamentale de notre approche qui doit associer éducation massive, démocratisation des outils de protection, et régulation internationale coordonnée.

Cette refonte doit s'articuler autour de plusieurs axes prioritaires. D'abord, un investissement massif dans l'éducation à tous les niveaux, incluant les jeunes dans les établissements scolaires, depuis la formation initiale jusqu'à la sensibilisation continue des adultes. Ensuite, le développement de cadres juridiques internationaux harmonisés, capables d'évoluer au rythme des innovations technologiques. Enfin, la création de solutions technologiques accessibles et éthiques, conçues dans une optique d'inclusion plutôt que de profit.

L'heure n'est plus aux demi-mesures ni aux solutions parcellaires. Nous avons besoin d'une révolution dans notre façon d'aborder cette menace, une révolution qui place l'humain, et non la technologie, au centre de la réponse. Sans ce changement radical de paradigme, nous risquons de voir émerger un monde où la vérité devient le privilège du plus fort, et où la confiance sociale s'érode jusqu'à disparaître.

La bataille contre les hypertrucages est donc bien plus qu'une lutte technologique : c'est un combat pour l'avenir de notre tissu social et de notre capacité collective à maintenir un espace public basé sur la vérité et la confiance. L'issue de ce combat déterminera non seulement notre capacité à préserver l'intégrité de l'information dans l'ère numérique, mais aussi notre aptitude à construire une société numérique véritablement équitable et résiliente. Le temps presse, et notre réponse actuelle n'est manifestement pas à la hauteur du défi. C'est maintenant qu'il faut agir, avant que la distinction entre réel et artificiel ne devienne le privilège exclusif d'une élite technologique, creusant davantage le fossé numérique qui menace déjà la cohésion de nos sociétés.

Références

- AFP. (2023, 10 janvier). La Chine serre la vis au « deepfake », ces trucages numériques hyperréalistes. *Le Journal de Montréal*. <https://www.journaldemontreal.com/2023/01/10/la-chine-serre-la-vis-au-deepfake-ces-trucages-numeriques-hyperrealistes>
- AFP. (2024, 5 août). Espagne : cinq jeunes mis en cause pour de fausses photos dénudées de jeunes filles générées par l'IA. *leparisien.fr*. <https://www.leparisien.fr/international/espagne-cinq-jeunes-mis-en-cause-pour-de-fausse-photos-denudees-de-jeunes-filles-gene-rees-par-lia-05-08-2024-PUHVRIY2JBOJE3GRQJVCQDUE.php>
- Africa Check. (n.d.). Comment nous vérifions les faits. <https://africacheck.org/fr/comment-nous-verifions-les-faits>
- Africa Check. (n.d.). Première organisation indépendante de vérification des faits en Afrique. <https://africacheck.org/fr/qui-nous-sommes>
- Anthropic. (2024). *Claude* (version 3 Opus) [grand modèle de langage]. <https://www.anthropic.com/>
- Apprentissage automatique. (2020). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8395061/apprentissage-automatique>
- Apprentissage profond. (2020). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26532876/apprentissage-profond>
- Atsüren, A. (2024, 10 février). Playground Online Demos | BioID. *BioID*. <https://www.bioid.com/playground/>
- Ban deepfakes. (n.d.). FAQ. <https://bandeepfakes.org/faq>
- Ban deepfakes. (n.d.). Media. <https://bandeepfakes.org/media>
- Ban deepfakes. (n.d.). The solution. <https://bandeepfakes.org/solution>
- Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*. <https://doi.org/10.1177/14614448241253138>
- Breen, K. (2024, 9 février). New Jersey teen sues classmate for allegedly creating, sharing fake AI nudes. *CBS News*. <https://www.cbsnews.com/news/new-jersey-teen-sues-classmate-for-allegedly-creating-sharing-fake-ai-nudes/>
- Bronskill, J. (2023, 13 décembre). Extremists certain to use deepfake to carry out attacks in coming year: analysis. *Global News*. <https://globalnews.ca/news/10166874/ai-deepfake-terror-report-canada/>
- Brumfield, C. (2023, 5 décembre). Deepfakes emerge as a top security threat ahead of the 2024 US election. *CSO Online*. <https://www.csoonline.com/article/1251094/deepfakes-emerge-as-a-top-security-threat-ahead-of-the-2024-us-election.html>
- Burchardt, D. (2023, 28 octobre). ESPN slammed for 'deepfake' Damian Lillard interview after Milwaukee Bucks debut – which was actually from. . . *The US Sun*. <https://www.the-sun.com/sport/9443068/espn-deepfake-damian-lillard-nba-interview-video-bucks-blazers/>
- Burgess, M. (2023, 16 octobre). Deepfake porn is out of control. *WIRED*. <https://www.wired.com/story/deepfake-porn-is-out-of-control/>
- Burt, T., & Horvitz, E. (2021, 5 mai). New steps to combat disinformation. *Microsoft on the Issues*. <https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/>
- Carillo, L. (2024, 13 mars). Interpol: Criminals increasingly use AI, crypto and online scams. *OCCRP*. <https://www.occrp.org/en/daily/18562-interpol-criminals-increasingly-use-ai-crypto-and-online-scams>
- CBC News: The National. (2023, 4 novembre). *How open-source intelligence is co-opted to spread war misinformation* [Vidéo]. YouTube. https://www.youtube.com/watch?v=eUBaV_55jdo
- Check, R. F. (2023, 30 octobre). FACT CHECK: Greta Thunberg 'Vegan Grenades' TV interview is deepfake. *Reuters*. <https://www.reuters.com/fact-check/greta-thunberg-vegan-grenades-tv-interview-is-deepfake-2023-10-30/#:~:text=A%20BBC%20spokesperson%20said%3A%20%E2%80%9CWe,her%20book%2C%20The%20Climate%20Book>
- Chen, H., & Magramo, K. (2024, 4 février). Finance worker pays out \$25 million after video call with deepfake 'chief financial officer.' *CNN*. <https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>
- Clayton, B. J. (2023, 22 juillet). Intel's deepfake detector tested on real and fake videos. *BBC News*. <https://www.bbc.com/news/technology-66267961>
- Cleo Abram. (2023, 23 octobre). *Why this music is so controversial (feat. Grimes + Spotify CEO!)* [Vidéo]. YouTube. https://www.youtube.com/watch?v=Ey75Xw_ikqs
- CNW Group. (2023, 21 novembre). Deepfakes – Beware! Scammers are profiting from AI! *Yahoo Finance*. https://finance.yahoo.com/news/deepfakes-beware-scammers-profiting-ai-160000887.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLnVbS8&guce_referrer_sig=AQAAAFd50W5iebe1WtSbpWh7J6JLd6m-qnOQNfYkzmZQgnj8FJ4_JHBUP9XsNLXWeEvm6EcG2VT6BTlim_-NNNtyjiOp5BoPbN0eTlJlJVDaodmqAMWsaTdom8X84hCkTe2m3xVTzIMzELLRUNlpWnnxADVW8kq44Cu3gMIlJBrud
- Coffey, L. (2023, 14 novembre). Presidents use AI voice clones and deepfakes for engagement. *Inside Higher Ed | Higher Education News, Events and Jobs*. <https://www.insidehighered.com/news/tech-innovation/artificial-intelligence/2023/11/14/presidents-use-ai-voice-clones-and>
- Cooke, R. (2023, 22 mars). How scammers likely used artificial intelligence to con Newfoundland seniors out of \$200K. *CBC*. <https://www.cbc.ca/news/canada/newfoundland-labrador/ai-vocal-cloning-grandparent-scam-1.6777106>

De la Santé Publique du Canada, A. (2020, 4 juin). Défis et possibilités en matière de santé publique rendus possibles grâce aux progrès du traitement des langues naturelles. *Relevé des maladies transmissibles au Canada*, 46(6). <https://www.canada.ca/fr/sante-publique/services/rapports-publications/releve-maladies-transmissibles-canada-rmtc/numero-mensuel/2020-46/numero-6-4-juin-2020/traitement-langage-naturel-sous-zone-intelligence-artificielle.html>

DeepMind. (n.d.). *SynthID*. <https://deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid/E20z>

Deschaseaux, C. (2024, 4 juillet). Essentiel pour l'Authenticité en Ligne : Intégrer le Watermarking aux Standards C2PA. *Imatag*. <https://www.imatag.com/fr/blog/authenticite-en-ligne-watermarking-standards-c2pa>

Desk, O. W. (2023, 27 novembre). How Deepfake Videos Leave Online Spaces Unsafe For Women. *Outlook*. <https://www.outlookindia.com/national/how-deepfake-videos-leave-online-spaces-unsafe-for-women-news-333131>

Dixit, P. (2023, 10 octobre). Social media platforms swamped with fake news on the Israel-Hamas war. *Al Jazeera*. <https://www.aljazeera.com/news/2023/10/10/social-media-platforms-swamped-with-fake-news-on-the-israel-hamas-war>

Dove, N. (2023, 7 décembre). Generative AI, deepfakes will 'pollute' election campaigns, CSE warns. *Global News*. <https://globalnews.ca/news/10151522/cse-report-artificial-intelligence-election-threats/>

Ducerf, J. (2024, 30 mai). Transparence et confiance : ChatGPT et Dall-E certifient leurs images avec les Content Credentials - Décryptage de la solution. *Medium*. https://medium.com/@julien_ducerf/transparence-et-confiance-chatgpt-et-dall-e-certifient-leurs-images-avec-les-content-c22eae3a933b

Dunne, M. (2023, 9 novembre). Data Dive: Fake news in the age of AI. *Ipsos*. <https://www.ipsos.com/en-ca/data-dive-fake-news-age-ai>

ElevenLabs. (n.d.). *Detect whether an audio clip was created using ElevenLabs*. https://elevenlabs.io/ai-speech-classifier?_ga=2.38257003.994246499.1719518028-43050963.1719518013

ElevenLabs. (2024, 14 août). *We're building AI to shape the future of digital communication, ensuring content is understood by everyone, everywhere*. <https://elevenlabs.io/about>

Elliott, V. (2023, 18 décembre). Worried about deepfakes? Don't forget "Cheapfakes." *WIRED*. <https://www.wired.com/story/meta-youtube-ai-political-ads/>

Emma Watson's face used in sexual deepfake on social media. (2023, 8 mars). *NBC News*. <https://www.nbcnews.com/tech/social-media/emma-watson-deep-fake-scarlett-johansson-face-swap-app-rcna73624>

ET NOW. (2023, 7 août). *Senior citizens fall prey to AI-Based deepfakes | Tips to protect yourself from Cybercrime* [Video]. YouTube. https://www.youtube.com/watch?v=e2A-6_IRiiM

Etimes.In. (2023, 9 octobre). Jamie Lee Curtis deletes post after backlash for claiming Palestinian children were Israelites amidst ongoing conflict. *The Times of India*. <https://timesofindia.indiatimes.com/entertainment/english/hollywood/news/jamie-lee-curtis-deletes-post-after-backlash-for-claiming-palestinian-children-were-israelites-amidst-ongoing-conflict/articleshow/104276309.cms?from=mdr>

Finextra. (2023, 31 octobre). Deepfake fraud: the rising threat in financial crime. *Finextra Research*. <https://www.finextra.com/the-long-read/844/deepfake-fraud-the-rising-threat-in-financial-crime>

Fischer, S., & Snyder, A. (2023, 10 juin). AI's hidden toll on our brains. *Axios*. <https://www.axios.com/2023/06/10/ai-mental-health-risks-misinformation>

Flitter, E., & Cowley, S. (2023, 30 août). Voice deepfakes are the latest threat to your bank balance. *The New York Times*. <https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html>

Full Fact. (n.d.). *All of the facts we have checked*. <https://fullfact.org/facts/>

Full Fact. (n.d.). *Full Fact Training*. <https://fullfact.org/training/>

Full Fact. (n.d.). *The Full Fact team*. <https://fullfact.org/about/our-team/>

Gain, V. (2022, 18 novembre). How deepfakes and AI are being used to find new ways to treat diseases. *Silicon Republic*. <https://www.siliconrepublic.com/innovation/deepfake-ai-healthcare-diseases-insilico-medicine-pharma>

Gerken, B. T. (2023, 4 octobre). MrBeast and BBC stars used in deepfake scam videos. *BBC News*. <https://www.bbc.com/news/technology-66993651>

Gilbert, D. (2023, 10 octobre). Elon Musk is personally undermining X's efforts to curb Israel-Hamas war disinformation. *WIRED*. <https://www.wired.com/story/elon-musk-israel-hamas-war-disinformation-x/>

Good Morning Britain. (2024, 18 juillet). *The Deepfake Epidemic: How AI is Fueling Medical Ad Scams* [Video]. YouTube. <https://www.youtube.com/watch?v=AdbqXQWqO2I>

Hall, J. (2023, 17 juin). Artificial intelligence is coming for seniors: AI targets older adults in scams. *MarketWatch*. <https://www.marketwatch.com/story/ais-dark-side-targets-older-adults-in-scams-3a8ee698>

Horetski, D. (2023, 21 novembre). TikTok model mortified by AI deepfake video showing her getting dressed. *Dexerto*. <https://www.dexerto.com/entertainment/tiktok-model-mortified-by-ai-deepfake-video-showing-her-getting-dressed-2392928/>

Hsu, T., & Thompson, S. A. (2023, 20 juin). Disinformation researchers raise alarms about A.I. chatbots. *The New York Times*. <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html>

Hypertrucage. (2019). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrlinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26552557/hypertrucage>

Hypertrucage : Que peut-on faire à propos du contenu audio et vidéo de synthèse? (n. d.). *Bibliothèque du Parlement*. https://bdp.parl.ca/sites/PublicWebsite/default/fr_CA/ResearchPublications/201911E

IBM. (n.d.). *What is Computer Vision?* <https://www.ibm.com/fr-fr/topics/computer-vision#:~:text=La%20vision%20par%20ordinateur%20est%20un%20domaine%20de,des%20recommandations%20sur%20la%20base%20de%20ces%20informations>

Intelligence artificielle. (2017). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8385376/intelligence-artificielle>

Intel. (n.d.). *Intel fake catcher*. <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake>

Intel. (n.d.). *Intel fake catcher*. <https://www.intel.fr/content/www/fr/fr/stories/deepfake-detection.html>

Issa, M. B. T. M. (2023, 18 mai). Grief tech: Therapists say AI, deepfake replicas of lost loved ones can help cope with loss - Arabian Business. *Arabian Business*. <https://www.arabianbusiness.com/industries/technology/grief-tech-therapists-say-ai-deepfake-replicas-of-dead-help-with-loss-grieving>

Israel-Hamas conflict: Combatting misinformation with open-source intelligence. (n.d.). *Global News*. <https://globalnews.ca/video/10034917/israel-hamas-crisis-combatting-misinformation-with-open-source-intelligence/>

Jain, K. (2023, 26 novembre). Deepfakes: The imminent threat - The Sunday Guardian Live. *The Sunday Guardian Live*. <https://sundayguardianlive.com/opinion/deepfakes-the-imminent-threat>

Jeu de données. (2020). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8369006/jeu-de-donnees>

Johnny Harris. (2023, 18 octobre). *How Facebook became a tool for genocide* [Vidéo]. YouTube. <https://www.youtube.com/watch?v=K8B0bWQ9u3M>

Jolly, J. (2023, 11 juillet). Financial firms must boost protections against AI scams, UK regulator to warn. *the Guardian*. <https://www.theguardian.com/technology/2023/jul/12/financial-firms-must-boost-protections-against-ai-scams-uk-regulator-to-warn>

Kaur, A., Hoshyar, A. N., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review*, 57(6). <https://doi.org/10.1007/s10462-024-10810-6>

Kelion, B. L. (2020, 1 septembre). Deepfake detection tool unveiled by Microsoft. *BBC News*. <https://www.bbc.com/news/technology-53984114>

Keystone-Ats. (2024, 15 mars). «Another Body»: le combat de jeunes femmes victimes de deepfakes. *SWI swissinfo.ch*. <https://www.swissinfo.ch/fre/%22another-body%22%3a-le-combat-de-jeunes-femmes-victimes-de-deepfakes/73727949>

Klar, R. (2023, 12 octobre). The Hill. *The Hill*. <https://thehill.com/policy/technology/4252928-eu-probes-x-disinformation-israel-gaza/>

Klepper, D. (2023, 28 novembre). Deepfakes from Gaza war increase fears about AI's power to mislead | AP News. *AP News*. <https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbb9c9c3aa1e000db47>

Kumar, A. (2023, 30 novembre). Sextortionists target senior citizen with retd IPS officer's deepfake. *The Times of India*. <https://timesofindia.indiatimes.com/city/ghaziabad/retired-ips-officers-deepfake-used-to-blackmail-senior-citizen/articleshow/105603283.cms>

Lanum, N. (2023, 8 novembre). President Biden's AI executive order has « dangerous limitations, » says Deepfake Detection Company CEO. *Fox Business*. <https://www.foxbusiness.com/media/president-bidens-ai-executive-order-dangerous-limitations-deepfake-detection-ceo>

Linehan, C. (s. d.). Deepfakes in warfare: New concerns emerge from their use around the Russian invasion of Ukraine. *The Conversation*. <https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393>

Little.O. (2023, 6 décembre). Media Matters for America. *Media Matters for America*. <https://www.mediamatters.org/tiktok/deepfake-warren-buffett-selling-bitcoin-scams-tiktok-users>

Mapping Media Freedom. (n.d.). *Belgium: VRT presenter Hanne Decoutere targeted by deepfake promoting fraudulent gaming application* (2024-06-22). <https://www.mapmf.org/alert/31910?q=deep>

Mapping Media Freedom. (n.d.). *France: France 24 targeted by deepfake discrediting the organisation of the Olympic Games* (2024-07-06). <https://www.mapmf.org/alert/31895>

Mapping Media Freedom. (n.d.). *France: French investigative journalist Élise Lucet targeted by deepfake promoting cryptocurrency sites* (2024-03-06). <https://www.mapmf.org/alert/31908?q=deep>

McCarthy, B. (2023, 20 octobre). *Fact check*. AFP. <https://factcheck.afp.com/doc.afp.com.33YP34M>

Métadonnée. (2020). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8869869/metadonnee>

Misciagna, V. (2023, 30 novembre). AI is being used to create deepfake images of the war in Gaza. *Scripps News*. <https://scrippsnews.com/stories/ai-is-being-used-to-create-deepfake-images-of-the-war-in-gaza/>

Modèle (IA). (s. d.). CNIL. <https://www.cnil.fr/fr/definition/modele-ia#:~:text=Le%20mod%C3%A8le%20d'IA%20est,%20du%20syst%C3%A8me%20d'IA>

Monde, L. (2024, 26 février). Deepfake de Joe Biden : l'identité du commanditaire dévoilée. *Le Monde.fr*. https://www.lemonde.fr/pixels/article/2024/02/26/deepfake-de-joe-biden-l-identite-du-commanditaire-devoilee_6218633_4408996.html

Murphy, M. (2023, 8 novembre). BNN. *BNN Bloomberg*. <https://www.bnnbloomberg.ca/microsoft-is-offering-to-help-us-politicians-crack-down-on-deepfakes-1.1995486>

Murphy, M., & Bloomberg. (2023, 8 décembre). 'Nudify' apps that use AI to undress women in photos are soaring in popularity. *Fortune*. <https://fortune.com/2023/12/08/nudify-apps-use-ai-popularity-deepfakes/>

Naskar, G., Mohiuddin, S., Malakar, S., Cuevas, E., & Sarkar, R. (2024). Deepfake Detection using Deep Feature Stacking and Meta-learning. *Heliyon*, e25933. <https://doi.org/10.1016/j.heliyon.2024.e25933>

Navlakha, M. (2024, 6 mars). Deepfake ads featuring Jenna Ortega ran on Meta platforms. Big Tech needs to fight this. *Mashable*. <https://mashable.com/article/jenna-ortega-meta-instagram-deepfake>

Netgrows Tech. (2023, 31 août). *Google SynthID, the invisible watermark to detect AI content* [Video]. YouTube. <https://www.youtube.com/watch?v=bZF0sVWZ7DU>

News Mobile. (2024, 3 août). *FACT CHECK: Indian High Commissioner in security meeting with Bangladesh PM? Find truth here*. <https://www.newsmobile.in/nm-fact-checker/fact-check-indian-high-commissioner-in-security-meeting-with-bangladesh-pm-find-truth-here/>

News Mobile. (2024, 27 mai). *Fact Check: Is that Rashmika Mandanna in red bikini or another deepfake? Know the truth*. <https://www.news-mobile.in/nm-fact-checker/fact-check-is-that-rashmika-mandanna-in-red-bikini-or-another-deepfake-know-the-truth/>

Niranjankumar, N., & Boom. (2023, 31 octobre). *BOOM. BOOM Live*. <https://www.boomlive.in/fact-check-video-of-jordans-queen-rania-supporting-israel-is-a-deepfake-23493>

noFake – Vernetzung und Sicherheit digitaler Systeme. (n.d.). <https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/nofake>

Nordling, L. (2024). Scientists are falling victim to deepfake AI video scams – here's how to fight back. *Nature*. <https://doi.org/10.1038/d41586-024-02521-3>

OpenAI. (2024). *ChatGPT* (version 4o with canvas) [grand modèle de langage]. <https://chat.openai.com>

Parsons, A. (2024, 5 février). 2024, l'année idéale pour encourager l'adoption des Content Credentials. *Adobe Blog*. <https://blog.adobe.com/fr/publish/2024/02/05/adobe-2024-encourager-adoption-des-content-credentials>

Pathak, A. (2024, 17 mai). Réseaux neuronaux convolutifs (CNN) : Introduction. *Geekflare France*. <https://geekflare.com/fr/convolutional-neural-networks/>

Press Release. (2023, 8 juin). Top 3 deepfake threat scenarios users face in 2023. *ACE Times*. <https://www.zawya.com/en/press-release/research-and-studies/top-3-deepfake-threat-scenarios-users-face-in-2023-p8qnf0xr>

QUEVRIN, C., & CAPRON, A. (2024, 24 juillet). Attention à cette vidéo menaçant la France avant les JO et attribuée (à tort) au Hamas. *TF1 INFO*. <https://www.tf1info.fr/jeux-olympiques/fake-news-menace-russe-destabilisation-attention-a-cette-video-menacant-la-france-avant-les-jo-et-attribuee-a-tort-au-hamas-2310141.html>

Quinn, M. (2023, 2 avril). N.L. family warns others not to fall victim to the same deepfake phone scam that cost them \$10K. *CBC*. <https://www.cbc.ca/news/canada/newfoundland-labrador/deepfake-phone-scam-1.6793296>

Radio-Canada. (2023, 22 mars). Des fraudeurs auraient utilisé l'intelligence artificielle pour escroquer des aînés. <https://ici.radio-canada.ca/nouvelle/1965288/intelligence-artificielle-fraude-aines-terre-neuve>

Rajan, R. (2023, 14 novembre). Ripple CEO warns of deepfake scams. *Investing.com Canada*. <https://ca.investing.com/news/stock-market-news/ripple-ceo-warns-of-deepfake-scams-93CH-3178269>

Ramirez, N. M. (2023, 18 octobre). Rolling Stone. *Rolling Stone*. <https://www.rollingstone.com/politics/politics-news/misinformation-gaza-hospital-attack-1234856302/>

Rao, R. (2023, 13 novembre). How deepfakes fuel conspiracy theories. *IEEE Spectrum*. <https://spectrum.ieee.org/deepfake-2666142928>

Reed, T. (2023, 14 novembre). Deepfakes could supercharge health care's misinformation problem. *Axios*. <https://www.axios.com/2023/11/14/ai-deepfake-health-misinformation-fake-pictures-videos>

Resemble AI. (2024, 7 août). *Detect Deepfake Audio with Resemble*. <https://www.resemble.ai/detect/>

Retired Top COP Latest victim of deepfake, video used to con Ghaziabad man. (n.d.). *NDTV.com*. <https://www.ndtv.com/ghaziabad-news/retired-top-cop-latest-victim-of-deepfake-video-used-to-con-ghaziabad-man-4621233>

Retrace. (2022, 24 août). Synthetic Medical Imaging: How Deepfakes Could Improve Healthcare. *Cision*. <https://www.prnewswire.com/news-releases/synthetic-medical-imaging-how-deepfakes-could-improve-healthcare-301611438.html>

Reuters. (2023, 16 mai). WHO warns against bias, misinformation in using AI in healthcare. <https://www.reuters.com/technology/who-warns-against-bias-misinformation-using-ai-healthcare-2023-05-16/>

Ritchie, D. (2023, 14 décembre). Vladimir Putin AI deepfake momentarily stuns him. *ReadWrite*. <https://readwrite.com/vladimir-putin-ai-deepfake-momentarily-stuns-him/>

Ryan-Mosley, T. (2023, 1 décembre). A high school's deepfake porn scandal is pushing US lawmakers into action. *MIT Technology Review*. <https://www.technologyreview.com/2023/12/01/1084164/deepfake-porn-scandal-pushing-us-lawmakers/>

Saint-Léger, A. (2024, 16 février). Info ou intox – France 24 victime d'un "deepfake" : l'intox continue à circuler sur le web ! *France 24*. <https://www.france24.com/fr/%C3%A9missions/info-ou-intox/20240206-taylor-swift-cible-privil%C3%A9%C3%A9-des-intox-aux-etats-unis>

Saint-Léger, A. (2024, février 16). Info ou intox – France 24 victime d'un "deepfake" : l'intox continue à circuler sur le web ! *France 24*. <https://www.france24.com/fr/%C3%A9missions/info-ou-intox/20240216-france-24-victime-d-un-deepfake-l-intox-continue-%C3%A0-circuler-sur-le-web>

Sarkar, A. (2023, novembre 17). Canada flags concern on AI-generated deepfake disinformation campaigns. *Cointelegraph*. <https://cointelegraph.com/news/canada-ai-generated-deepfake-disinformation-campaigns>

Sato, M. (2024, juillet 9). The UK politician accused of being AI is actually a real person. *The Verge*. <https://www.theverge.com/2024/7/9/24195005/reform-uk-candidate-election-ai-bot-mark-matlock>

Savage, B. M. (2023, 17 avril). AI-generated Drake and the Weeknd song goes viral. *BBC News*. <https://www.bbc.com/news/entertainment-arts-65298834>

Saxena, V. (2023, 22 mai). \$600,000 'Deepfake' fraud heats up AI debate in China. *Asia Financial*. <https://www.asiafinancial.com/600000-deepfake-fraud-heats-up-ai-debate-in-china>

Schwartz, E. H. (2023, décembre 5). Resemble AI releases new Real-Time Deepfake audio Detector - Voicebot.ai. *Voicebot.ai*. <https://voicebot.ai/2023/12/05/resemble-ai-releases-new-real-time-deepfake-audio-detector/>

Sécurité Internet. (2020). Dans *Grand Dictionnaire Terminologique*. Office québécois de la langue française. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8370798/securite-internet>

Sentinel - Defending against deepfakes and information warfare. (n.d.). <https://thesentinel.ai/>

Spring, B. M. (2023, 15 octobre). Who's behind Israel-Gaza disinformation and hate online? *BBC News*. <https://www.bbc.com/news/world-middle-east-67114313>

Spring, B. M. (2024, mars 4). Trump supporters target black voters with faked AI images. *BBC News*. <https://www.bbc.com/news/world-us-canada-68440150>

Staff, F., & Staff, F. (2023, 1 novembre). AI making Gaza situation worse, Hamas, Israel using deepfakes to sway public opinion. *Firstpost*. <https://www.firstpost.com/world/ai-making-gaza-situation-worse-hamas-israel-using-deepfakes-to-sway-public-opinion-13327892.html>

Staff, F., & Staff, F. (2023, décembre 7). Deepfake menace: Ratan Tata latest to get deepfaked, scammers promote investment scam using fake video. *Firstpost*. <https://www.firstpost.com/world/deepfake-menace-ratan-tata-latest-to-get-deepfaked-scammers-promote-investment-scam-using-fake-video-13478102.html>

Staff, N. A. (n.d.). Pro-Israel activists deep fake Bella Hadid Palestine video. *The New Arab*. <https://www.newarab.com/news/pro-israel-activists-deep-fake-bella-hadid-palestine-video>

Standard, M. (2024, juillet 22). 'Pathetic, Laos' Maisug rallyists blasted for "infecting platforms" with deepfake Marcos video. *Manila Standard*. <https://manilastandard.net/news/314473171/pathetic-laos-maisug-rallyists-blasted-for-infecting-platforms-with-deepfake-marcos-video.html>

Swan, D. (2023, novembre 19). Australia launches world-first crackdown on 'deepfake' porn. *The Sydney Morning Herald*. <https://www.smh.com.au/technology/australia-launches-world-first-crackdown-on-deepfake-porn-20231119-p5ellv.html>

Syme, P. (2023, 9 octobre). Meta is paying the celebrity faces behind its AI chatbots as much as \$5 million for 6 hours of work, report says. *Business Insider*. <https://www.businessinsider.com/meta-paying-celebrity-faces-of-ai-chatbots-as-much-as-5-million-2023-10>

SynthID. (2024, août 22). Google DeepMind. <https://deepmind.google/technologies/synthid/>

Team, W. W. (2023, novembre 22). G20 Virtual Summit: PM Modi backs AI boom but cautions society over misuse of tech like deepfake. *WION*. <https://www.wionews.com/india-news/g20-virtual-summit-pm-modi-backs-ai-boom-but-cautions-society-over-misuse-of-tech-like-deepfake-661575>

The DeepFake Detection Challenge (DFDC) dataset. (n.d.). *Ar5iv*. <https://ar5iv.labs.arxiv.org/html/2006.07397>

The Full Fact team - Full Fact. (n.d.). *Full Fact*. <https://fullfact.org/about/our-team/>

The role of deepfake technology in the landscape of misinformation and cybersecurity threats. (n.d.). *ISACA*. <https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2023/the-role-of-deepfake-technology-in-the-landscape-of-misinformation-and-cybersecurity-threats>

The solution | Ban deepfakes. (n.d.). <https://bandeepfakes.org/solution>

TheFootballHub63. (n.d.). *Facebook*. https://www.facebook.com/story.php?story_fbid=122106270794404672&id=61562140167845&mibextid=WC7FNe&rdid=fS6fbc2vofWvz5b3

TheGerontechnologist - Keren Etkin. (2024, février 23). Exposing the dark side of AI: How deepfake threatens our seniors! [Vidéo]. *YouTube*. <https://www.youtube.com/watch?v=MX5So4rdvIU>

TVA Nouvelles. (2024, juillet 23). « C'est moi Megan Lalonde »: la « vraie » fausse avocate passe aux aveux - reportage [Vidéo]. *YouTube*. <https://www.youtube.com/watch?v=IHLDAgCPURa>

U.S. lawmakers question Meta and X over AI-generated political deepfakes ahead of 2024 election. (2023, octobre 5). *PBS NewsHour*. <https://www.pbs.org/newshour/politics/u-s-lawmakers-question-meta-and-x-over-ai-generated-political-deepfakes-ahead-of-2024-election>

VandeHei, J., & Allen, M. (2023, novembre 8). Behind the curtain: What AI architects fear most (in 2024). *Axios*. <https://www.axios.com/2023/11/08/ai-fears-deepfake-misinformation>

Video of Biden botching Ukraine history is a deepfake. (2024, mars 19). *Fact Check*. <https://factcheck.afp.com/doc.afp.com.34LY8TK>

Video of Gautam Adani endorsing a gaming platform is fake | DeepFakes Analysis Unit | A Misinformation Combat Alliance initiative. (n.d.). <https://www.dau.mcaindia.in/blog/video-of-gautam-adani-endorsing-a-gaming-platform-is-fake>

Video of Rajat Sharma promoting diabetes cure is fake | DeepFakes Analysis Unit | A Misinformation Combat Alliance initiative. (n.d.). <https://www.dau.mcaindia.in/blog/video-of-rajat-sharma-promoting-diabetes-cure-is-fake>

Video of Yogi Adityanath endorsing a gaming platform is fake | DeepFakes Analysis Unit | A Misinformation Combat Alliance initiative. (n.d.). <https://www.dau.mcaindia.in/blog/video-of-yogi-adityanath-endorsing-a-gaming-platform-is-fake>

Video peddling A.I. powered trading software by Musk is fake | DeepFakes Analysis Unit | A Misinformation Combat Alliance initiative. (n.d.). <https://www.dau.mcaindia.in/blog/video-peddling-a-i-powered-trading-software-by-musk-is-fake>

Videos featuring Ambanis promoting a gaming app are fake | DeepFakes Analysis Unit | A Misinformation Combat Alliance initiative. (n.d.). <https://www.dau.mcaindia.in/blog/videos-featuring-ambanis-promoting-a-gaming-app-are-fake>

Vlinkinfo. (n.d.). 10 Top AI Deepfake Detector Tools for 2024 & Beyond. <https://www.vlinkinfo.com/blog/top-ai-deepfake-detector-tools/>

Voicepods. (2024, mars 13). Cal Poly AI voice cloning scam case. *Resemble AI*. <https://www.resemble.ai/cal-poly-ai-voice-cloning-scam-case/>

Voicepods. (2024, mars 26). Arcads AI deepfake video. *Resemble AI*. <https://www.resemble.ai/arcads-ai-deepfake-video/>

Warren, B. J. (2023, 11 novembre). Fake audio of Sadiq Khan is not a crime, says Met. *BBC News*. <https://www.bbc.com/news/uk-england-london-67389609>

We're building AI to shape the future of digital communication, ensuring content is understood by everyone, everywhere. | ElevenLabs. (2024, août 14). *ElevenLabs*. <https://elevenlabs.io/about>

Wei, M., MD. (2023, octobre 18). Who thrives in a world of deepfakes and misinformation? *Psychology Today*. <https://www.psychologytoday.com/us/blog/urban-survival/202310/who-thrives-in-a-world-of-deepfakes-and-misinformation>

West, A. (2024, mai 21). YouTuber Brooke Monk responds to sexual deep fake video with Taylor Swift song. *The Mirror US*. <https://www.themirror.com/entertainment/celebrity-news/youtuber-brooke-monk-responds-sexual-497825>

West, B. K. (2024, avril 2). I was deepfaked by my best friend. *BBC News*. <https://www.bbc.com/news/uk-68673390>

What is Computer Vision? | IBM. (n.d.). <https://www.ibm.com/fr-fr/topics/computer-vision#:~:text=La%20vision%20par%20ordinateur%20est%20un%20domaine%20de,des%20recommandations%20sur%20la%20base%20de%20ces%20informations>

Whittaker, L., Mulcahy, R., Letheren, K., Kietzmann, J., & Russell-Bennett, R. (2023). Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation*, 125, 102784. <https://doi.org/10.1016/j.technovation.2023.102784>

Wikipedia contributors. (2024, août 21). ElevenLabs. *Wikipedia*. <https://en.wikipedia.org/wiki/ElevenLabs>

WITNESS. (2018, décembre 5). Human Rights Campaigns & Projects from WITNESS. <https://www.witness.org/our-work/>

Witt, A., & Henry, N. (n.d.). Taylor Swift deepfakes: New technologies have long been weaponised against women. The solution involves us all. *The Conversation*. <https://theconversation.com/taylor-swift-deepfakes-new-technologies-have-long-been-weaponised-against-women-the-solution-involves-us-all-222268>

World Health Organization: WHO. (2023, octobre 19). WHO outlines considerations for regulation of artificial intelligence for health. *WHO*. <https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health>

X, S. (2023, 27 octobre). First-ever study of wartime deepfakes reveals their impact on news media. *Tech Xplore*. <https://techxplore.com/news/2023-10-first-ever-wartime-deepfakes-reveals-impact.html#:~:text=,School%20of%20Applied%20Psychology%20researcher>

Annexes

Volet ressources hautement partagées sur les réseaux sociaux

Annexe A - Tableau des ressources hautement partagées sur les réseaux sociaux - Général					
Title	Type	Source	Link	Industry / context	Date of publication
MrBeast and BBC stars used in deepfake scam videos	Article professionnel	BBC	https://www.bbc.com/news/technology-66993651	Médias Sociaux	4/10/2023
Social media platforms swamped with fake news on the Israel-Hamas war	Article professionnel	Aljazeera	https://www.aljazeera.com/news/2023/10/10/social-media-platforms-swamped-with-fake-news-on-the-israel-hamas-war	Politique	10/10/2023
Jamie Lee Curtis deletes post after backlash for claiming Palestinian children were Israelites amidst ongoing conflict	Article professionnel	Times of India	https://timesofindia.indiatimes.com/entertainment/english/hollywood/news/jamie-lee-curtis-deletes-post-after-backlash-for-claiming-palestinian-children-were-israelites-amidst-ongoing-conflict/articleshow/104276309.cms?from=mdr	Politique	10/10/2023
Elon Musk is shitposting his way through the Israel-Hamas War	Article professionnel	Wired	https://www.wired.com/story/elon-musk-israel-hamas-war-disinformation-x/	Politique	10/10/2023
Meta is paying the celebrity faces behind its AI chatbots as much as \$5 million for 6 hours of work, report says	Article professionnel	Business Insider	https://www.businessinsider.com/meta-paying-celebrity-faces-of-ai-chatbots-as-much-as-5-million-2023-10	Médias Sociaux	9/10/2023
Voice deepfakes are coming for your bank balance	Article professionnel	New York Times	https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html	Finances et administration	30/08/2023
\$600,000 'Deepfake' fraud heats up AI debate in China	Article professionnel	Asia Financial	https://www.asiafinancial.com/600000-deepfake-fraud-heats-up-ai-debate-in-china	Finances et administration	22/05/2023
Financial firms must boost protections against AI scams, UK regulator to warn	Article professionnel	The Guardian	https://www.theguardian.com/technology/2023/jul/12/financial-firms-must-boost-protections-against-ai-scams-uk-regulator-to-warn	Finances et administration	12/7/2023
Deepfake Porn is out of control	Article professionnel	Wired	https://www.wired.com/story/deepfake-porn-is-out-of-control/	Divertissement et sports	16/10/2023
Misinformation runs rampant in aftermath of Gaza hospital attack	Article professionnel	Rolling Stone	https://www.rollingstone.com/politics/politics-news/misinformation-gaza-hospital-attack-1234856302/	Politique	17/10/2023
Israel-Hamas conflict: COmbatting misinformation with open source intelligence	Vidéo	Global News	https://globalnews.ca/video/10034917/israel-hamas-crisis-combatting-misinformation-with-open-source-intelligence/	Politique	18/10/2023
Who's behind Israel-Gaza disinformation and hate online	Article professionnel	BBC	https://www.bbc.com/news/world-middle-east-67114313	Politique	15/10/2023
EU probes X over spread of Israel-Gaza disinformation, 'terrorist' content	Article professionnel	The Hill	https://thehill.com/policy/technology/4252928-eu-probes-x-disinformation-israel-gaza/	Politique	12/10/2023
Tik Tok removes viral video ad of suspected Joe Rogan AI deepfake	Article professionnel	Mashable	https://mashable.com/article/joe-rogan-tiktok-deepfake-ad	Médias Sociaux	15/02/2023
How Facebook Became a tool for genocide	Vidéo	Johnny Harris	https://www.youtube.com/watch?v=K8B0bWQ9u3M	Politique	18/10/2023
ESPN slammed for 'deepfake' Damian Lillard interview after Milwaukee Bucks debut - which was actually from 2020	Article professionnel	The Sun	https://www.the-sun.com/sport/9443068/espn-deepfake-damian-lillard-nba-interview-video-bucks-blazers/	Divertissement et sports	28/10/2023
Pro-Israel activists deep fake Bella Hadid video showing Palestine 'Remorse'	Article professionnel	The New Arab	https://www.newarab.com/news/pro-israel-activists-deep-fake-bella-hadid-palestine-video	Politique	30/10/2023
Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine	Article professionnel	The Conversation	https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393	Politique	26/10/2023

Why This Music Is So Controversial (feat. Grimes + Spotify CEO!)	Vidéo	Cleo Abram	https://www.youtube.com/watch?v=Ey75Xw_ikqs&ab_channel=CleoAbram	Divertissement et sports	23/10/2023
AI-generated Drake and The Weeknd song goes viral	Article professionnel	BBC	https://www.bbc.com/news/entertainment-arts-65298834	Divertissement et sports	17/04/2023
How deepfakes and AI are being used to find new ways to treat diseases	Article professionnel	Silicon Republic	https://www.siliconrepublic.com/innovation/deepfake-ai-healthcare-diseases-insilico-medicine-pharma	Santé	18/11/2022
Grief tech: Therapists say AI, deepfake replicas of lost loved ones can help cope with loss	Article professionnel	Arabian Business	https://www.arabianbusiness.com/industries/technology/grief-tech-therapists-say-ai-deepfake-replicas-of-dead-help-with-loss-grieving	Santé	18/05/2023
Synthetic Medical Imaging: How Deepfakes Could Improve Healthcare	Article professionnel	prnewswire	https://www.prnewswire.com/news-releases/synthetic-medical-imaging-how-deepfakes-could-improve-healthcare-301611438.html	Santé	24/08/2022
How open-source intelligence is co-opted to spread war misinformation	Vidéo	CBC News: The National	https://www.youtube.com/watch?v=eUBaV_55jdo	Politique	4/11/2023
Top 3 deepfake threat scenarios users face in 2023	Article professionnel	Zawya	https://www.zawya.com/en/press-release/research-and-studies/top-3-deepfake-threat-scenarios-users-face-in-2023-p8qnf0xr	Médias Sociaux	8/6/2023
Deepfake fraud: the rising threat in financial crime	Article professionnel	Finextra	https://www.finextra.com/the-long-read/844/deepfake-fraud-the-rising-threat-in-financial-crime	Finances et administration	2/8/2021
The role of Deepfake Technology in the Landscape of Misinformation and Cybersecurity Threats	Article professionnel	ISACA	https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2023/the-role-of-deepfake-technology-in-the-landscape-of-misinformation-and-cybersecurity-threats	Cybersecrurité	9/8/2023
Fact Check: Greta Thunberg 'vegan grenades' TV interview is deepfake	Article professionnel	Reuters	https://www.reuters.com/fact-check/greta-thunberg-vegan-grenades-tv-interview-is-deepfake-2023-10-30/#:~:text=A%20BBC%20spokesperson%20said%3A%20%E2%80%9CWe,her%20book%2C%20The%20Climate%20Book	Médias Sociaux	30/10/2023
First-ever study of wartime deepfakes reveals their impact on news media	Article professionnel	Tech Xplore	https://techxplore.com/news/2023-10-first-ever-wartime-deepfakes-reveals-impact.html#:~:text=School%20of%20Applied%20Psychology%20researcher	Politique	27/10/2023
I executive order has 'dangerous limitations', says deepfake detection company CEO	Article professionnel	Fox Business	https://www.foxbusiness.com/media/pre-sident-bidens-ai-executive-order-dangerous-limitations-deepfake-detection-ceo	Hypertrucage	8/11/2023
Microsoft Is Offering to Help US Politicians Crack Down on Deepfakes	Article professionnel	BNN Bloomberg	https://www.bnnbloomberg.ca/microsoft-is-offering-to-help-us-politicians-crack-down-on-deepfakes-1.1995486	Hypertrucage	8/11/2023
Deepfake video of Indian star leaves nation fuming	Article professionnel	New York Post	https://nypost.com/2023/11/07/tech/deepfake-video-of-indian-star-rashmika-mandanna-leaves-nation-fuming/	Hypertrucage	7/11/2023
AI making Gaza situation worse, Hamas, Israel using deepfakes to sway public opinion	Article professionnel	First Post	https://www.firstpost.com/world/ai-making-gaza-situation-worse-hamas-israel-using-deepfakes-to-sway-public-opinion-13327892.html	Hypertrucage	1/11/2023
Video Of Jordan's Queen Rania Supporting Israel Is A Deepfake	Article professionnel	Boom	https://www.boomlive.in/fact-check/video-of-jordans-queen-rania-supporting-israel-is-a-deepfake-23493	Hypertrucage	31/10/2023
Biden deepfake announcing US draft resurfaces amid Israel-Hamas war	Article professionnel	AFP Fact Check	https://factcheck.afp.com/doc.afp.com.33YP34M	Hypertrucage	20/10/2023
AI's hidden toll on our brains	Article professionnel	Axios	https://www.axios.com/2023/06/10/ai-mental-health-risks-misinformation	Santé	10/6/2023
Generative AI Is Playing a Surprising Role in Israel-Hamas Disinformation	Article professionnel	Wired	https://www.wired.com/story/israel-hamas-war-generative-artificial-intelligence-disinformation/	Politique	30/10/2023
Disinformation Researchers Raise Alarms About A.I. Chatbots	Article professionnel	New York Times	https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html	Désinformation	8/2/2023

WHO warns against bias, misinformation in using AI in healthcare	Article professionnel	Reuters	https://www.reuters.com/technology/who-warns-against-bias-misinformation-using-ai-healthcare-2023-05-16/	Santé	16/05/2023
WHO outlines considerations for regulation of artificial intelligence for health	Article professionnel	WHO	https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health	Santé	19/10/2023
Data Dive: Fake news in the age of AI	Article professionnel	Ipsos	https://www.ipsos.com/en-ca/data-dive-fake-news-age-ai	Désinformation	1/11/2023
Ripple CEO warns of deepfake scams	Article professionnel	Investing.com	https://ca.investing.com/news/stock-market-news/ripple-ceo-warns-of-deepfake-scams-93CH-3178269	Finances et administration	14/11/2023
Deepfakes could supercharge health care's misinformation problem	Article professionnel	Axios	https://www.axios.com/2023/11/14/ai-deep-fake-health-misinformation-fake-pictures-videos	Santé	14/11/2023
Meta requires political advertisers to mark when deepfakes used	Article professionnel	BBC	https://www.bbc.com/news/technology-67366311	Médias Sociaux	9/11/2023
AI Voice Clones and Deepfakes: The Latest Presidents' Engagement Tools	Article professionnel	Inside Higher Ed	https://www.insidehighered.com/news/tech-innovation/artificial-intelligence/2023/11/14/presidents-use-ai-voice-clones-and	Éducation	14/11/2023
Audio deepfake detective developing new sleuthing techniques	Article professionnel	University of Rochester News Center	https://www.rochester.edu/newscenter/audio-deepfake-detective-developing-new-sleuthing-techniques-573482/	Hypertrucage	9/11/2023
TikTok model mortified by AI deepfake video showing her getting dressed	Article professionnel	Dexerto	https://www.dexerto.com/entertainment/tiktok-model-mortified-by-ai-deepfake-video-showing-her-getting-dressed-2392928/	Hypertrucage	20/11/2023
G20 Virtual Summit: PM Modi backs AI boom but cautions society over misuse of tech like deepfake	Article professionnel	Wion	https://www.wionews.com/india-news/g20-virtual-summit-pm-modi-backs-ai-boom-but-cautions-society-over-misuse-of-tech-like-deepfake-661575	Hypertrucage	22/11/2023
Deepfakes – Beware! Scammers are profiting from AI!	Article professionnel	Yahoo! Finance	https://finance.yahoo.com/news/deepfakes-beware-scammers-profiting-ai-160000887.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d-3cuZ29vZ2x1LnNvbS8&guce_referrer_sig=AQAAAFd50W5iebe1WtSbpWh-tZJ6JLd6m-qnQQNFykzmZQgnj8Fj4_JHBUP9XsNLXWeEvm6EcG2VT6BTlim_-NNNtyJiOp5BoPbN0eTJlJVDaodmqAMWsaTdom8X84hCkkTe2m3xVTzIMzELLRUNlpWnnxADVW8kq44Cu3gMlJlBrud	Finances et administration	21/11/2023
Canada flags concern on AI-generated deepfake disinformation campaigns	Article professionnel	Coin Telegraph	https://cointelegraph.com/news/canada-ai-generated-deepfake-disinformation-campaigns	Hypertrucage	17/11/2023
How Deepfakes Fuel Conspiracy Theories	Article professionnel	IEEE Spectrum	https://spectrum.ieee.org/deepfake-2666142928	Hypertrucage	9/11/2023
Australia launches world-first crackdown on 'deepfake' porn	Article professionnel	The Sydney Morning Herald	https://www.smh.com.au/technology/australia-launches-world-first-crackdown-on-deepfake-porn-20231119-p5el1v.html	Hypertrucage	20/11/2023
How Deepfake Videos Leave Online Spaces Unsafe For Women	Article professionnel	Outlook	https://www.outlookindia.com/national/how-deepfake-videos-leave-online-spaces-unsafe-for-women-news-333131	Hypertrucage	27/11/2023
Deepfakes: The imminent threat	Article professionnel	The Sunny Guardian	https://sundayguardianlive.com/opinion/deepfakes-the-imminent-threat	Hypertrucage	26/11/2023
Who Thrives in a World of Deepfakes and Misinformation?	Article professionnel	Psychology Today	https://www.psychologytoday.com/us/blog/urban-survival/202310/who-thrives-in-a-world-of-deepfakes-and-misinformation	Hypertrucage	17/10/2023
U.S. lawmakers question Meta and X over AI-generated political deepfakes ahead of 2024 election	Article professionnel	PBS	https://www.pbs.org/newshour/politics/u-s-lawmakers-question-meta-and-x-over-ai-generated-political-deepfakes-ahead-of-2024-election	Politique	5/10/2023
Behind the Curtain: What AI architects fear most (in 2024)	Article professionnel	Axios	https://www.axios.com/2023/11/08/ai-fears-deepfake-misinformation	Hypertrucage	8/11/2023

Deepfake Elections: How Indian Politicians Are Using AI-Manipulated Media To Malign Opponents	Article professionnel	Outlook	https://business.outlookindia.com/technology/deepfake-elections-how-indian-politicians-are-using-ai-manipulated-media-to-malign-opponents	Politique	24/11/2023
Fake audio of Sadiq Khan is not a crime, says Met	Article professionnel	BBC	https://www.bbc.com/news/uk-england-london-67389609	Politique	11/11/2023
Fake babies, real horror: Deepfakes from the Gaza war increase fears about AI's power to mislead	Article professionnel	AP	https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9c9c3aa1e000db47	Hypertrucage	28/11/2023
Deepfake mocks Alexandria Ocasio-Cortez's call for Gaza ceasefire	Article professionnel	Yahoo! News	https://news.yahoo.com/deepfake-mocks-alexandria-ocasio-cortezs-184920490.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d-3cuZ29vZ2x1LmNvbS8&guce_referrer_sig=AQAAAFd50W5iebe1WtSbpWh-tZJ6Jld6m-qnQQNFykzmZQgnj8FJ4_JHBUP9XsNLXWeEvm6EcG2VT6BTlim_-NNNtyjiOp5BoPbN0eTjLJVDAodmqAMWsaTdom8X84hCkkTe2m3xVTzIMzELLRUNlpWnnxADVW8kq44Cu3gMlIlJBrud	Hypertrucage	29/11/2023
AI is being used to create deepfake images of the war in Gaza	Article professionnel	Scripps News	https://scrippsnews.com/stories/ai-is-being-used-to-create-deepfake-images-of-the-war-in-gaza/	Hypertrucage	29/11/2023
Deepfakes emerge as a top security threat ahead of the 2024 US election	Article professionnel	CSO	https://www.csoonline.com/article/1251094/deepfakes-emerge-as-a-top-security-threat-ahead-of-the-2024-us-election.html	Politique	5/12/2023
Deepfake Menace: Ratan Tata latest to get deepfaked, scammers promote investment scam using fake video	Article professionnel	Firstpost	https://www.firstpost.com/world/deepfake-menace-ratan-tata-latest-to-get-deepfaked-scammers-promote-investment-scam-using-fake-video-13478102.html	Finances et administration	7/12/2023
Generative AI, deepfakes will 'pollute' election campaigns, CSE warns	Article professionnel	Global News	https://globalnews.ca/news/10151522/cse-report-artificial-intelligence-election-threats/	Hypertrucage	7/12/2023
Resemble AI Releases New Real-Time Deepfake Audio Detector	Article professionnel	voicebot.ai	https://voicebot.ai/2023/12/05/resemble-ai-releases-new-real-time-deepfake-audio-detector/	Hypertrucage	5/12/2023
Deepfake Warren Buffett is selling bitcoin scams to TikTok users	Article professionnel	Media Matters	https://www.mediamatters.org/tiktok/deepfake-warren-buffett-selling-bitcoin-scams-tiktok-users	Finances et administration	6/12/2023
A high school's deepfake porn scandal is pushing US lawmakers into action	Article professionnel	MIT Technology review	https://www.technologyreview.com/2023/12/01/1084164/deepfake-porn-scandal-pushing-us-lawmakers/	Politique	1/12/2023
Worried About Political Deepfakes? Beware the Spread of 'Cheapfakes'	Article professionnel	Wired	https://www.wired.com/story/meta-youtube-ai-political-ads/	Politique	18/12/2023
Extremists certain to use deepfake to carry out attacks in coming year: analysis	Article professionnel	Global News	https://globalnews.ca/news/10166874/ai-deepfake-terror-report-canada/	Politique	13/12/2023
'Nudify' apps that use AI to undress women in photos are soaring in popularity	Article professionnel	Fortune	https://fortune.com/2023/12/08/nudify-apps-use-ai-popularity-deepfakes/	Hypertrucage	8/12/2023
REPORT AI Deepfake Incidents Across Africa Increased by 450% in 2023, Crypto Took the Lead	Article professionnel	Bitcoinke	https://bitcoinke.io/2023/12/ai-deepfake-2023-report/	Finances et administration	20/12/2023
Vladimir Putin AI deepfake momentarily stuns him	Article professionnel	ReadWrite	https://readwrite.com/vladimir-putin-ai-deepfake-momentarily-stuns-him/	Politique	14/12/2023
N.L. family warns others not to fall victim to the same deepfake phone scam that cost them \$10K	Article professionnel	CBC	https://www.cbc.ca/news/canada/newfoundland-labrador/deepfake-phone-scam-1.6793296	Hypertrucage	29/03/2023
Artificial intelligence 'amplifying' effect of scams against older adults	Article professionnel	Mcknights	https://www.mcknightsseniorliving.com/home/news/artificial-intelligence-amplifying-effect-of-scams-against-older-adults/	Hypertrucage	17/11/2023
Retired Top Cop Latest Victim Of Deepfake, Video Used To Con Ghaziabad Man	Article professionnel	NDTV	https://www.ndtv.com/ghaziabad-news/retired-top-cop-latest-victim-of-deepfake-video-used-to-con-ghaziabad-man-4621233	Hypertrucage	30/11/2023

Artificial intelligence is coming for seniors: AI's dark side targets older adults in scams	Article professionnel	MarketWatch	https://www.marketwatch.com/story/ais-dark-side-targets-older-adults-in-scams-3a8ee698	Finances et administration	17/06/2023
Interpol: Criminals Increasingly Use AI, Crypto and Online Scams	Article professionnel	OCCRP	https://www.occrp.org/en/daily/18562-interpol-criminals-increasingly-use-ai-crypto-and-online-scams	Hypertrucage	13/03/2024
Is that you, dear? Why seniors may be worse at detecting voice-generated AI — and the scams that use it	Article professionnel	Toronto Star	https://www.thestar.com/opinion/contributors/is-that-you-dear-why-seniors-may-be-worse-at-detecting-voice-generated-ai-and/article_3fa76c2d-63fd-50db-a395-e7c9b0731491.html	Hypertrucage	5/8/2023
How scammers likely used artificial intelligence to con Newfoundland seniors out of \$200K	Article professionnel	CBC	https://www.cbc.ca/news/canada/newfoundland-labrador/ai-vocal-cloning-grandparent-scam-1.6777106	Hypertrucage	22/03/2023
The Deepfake Epidemic: How AI is Fueling Medical Ad Scams	Vidéo	Good Morning Britain	https://www.youtube.com/watch?v=AdbqXQWQ2I	Santé	7/18/2024
The terrifying AI scam that uses your loved one's voice	Exemple deepfake	The New Yorker	https://www.newyorker.com/science/annals-of-artificial-intelligence/the-terrifying-ai-scam-that-uses-your-loved-ones-voice	Social	3/7/2024
Reform UK candidate election AI bot Mark Matlock	Article professionnel	The Verge	https://www.theverge.com/2024/7/9/24195005/reform-uk-candidate-election-ai-bot-mark-matlock https://www.liberation.fr/international/europe/au-royaume-uni-letrange-histoire-du-candidat-quin-croyait-cree-par-ia-20240711_7ZE7S3Y-5ZRH7BCEXZ54QAGWAAA/	Politique	7/9/2024
AI candidate "Steve" runs for UK parliament	Article professionnel	NBC News	https://www.nbcnews.com/tech/tech-news/ai-candidate-running-parliament-uk-says-ai-can-humanize-politics-rcna156991	Politique	6/13/2024
Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'	Exemple deepfake	CNN	https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html	Finances et administration	2/4/2024
Deepfake de Joe Biden : l'identité du commanditaire dévoilée	Article professionnel	Le Monde	https://www.lemonde.fr/pixels/article/2024/02/26/deepfake-de-joe-biden-l-identite-du-commanditaire-devoilee_6218633_4408996.html	Politique	2/26/2024
France 24 targeted by deepfake discrediting the organisation of the Olympic Games	Article professionnel	Mapping media Freedom	https://www.mapmf.org/alert/31895?q=deepfake	Politique	7/6/2024
French investigative journalist Élise Lucet targeted by deepfake promoting cryptocurrency sites	Article professionnel	Mapping media Freedom	https://www.mapmf.org/alert/31908?q=deep	Politique	3/6/2024
La présentatrice de la VRT Hanne Decoutere ciblée par un deepfake faisant la promotion d'une application de jeu frauduleuse	Article professionnel	Mapping media Freedom	https://www.mapmf.org/alert/31910?q=deep	Social	6/22/2024
Video of Biden botching Ukraine history is a deepfake	Blogue	AFP Fact check	https://factcheck.afp.com/doc.afp.com.34LY8TK	Politique	3/22/2024
Trump supporters target black voters with faked AI images	Article professionnel	BBC	https://www.bbc.com/news/world-us-canada-68440150	Politique	3/4/2024
France 24 victime d'un "deepfake" : l'intox continue à circuler sur le web !	Article professionnel	France 24	https://www.france24.com/fr/%C3%A9missions/info-ou-intox/20240216-france-24-victime-d-un-deepfake-l-intox-continue-%C3%A0-circuler-sur-le-web	Politique	2/16/2024
Taylor Swift, cible privilégiée des intox aux États-Unis	Article professionnel	France 24	https://www.france24.com/fr/%C3%A9missions/info-ou-intox/20240206-taylor-swift-cible-privil%C3%A9gi%C3%A9e-des-intox-aux-etats-unis	Politique	2/6/2024
Kamala Harris deepfakes are going viral on TikTok and Elon Musk's X	Article professionnel	Mashable	https://mashable.com/article/kamala-harris-deepfake-twitter-x-tiktok	Politique	7/23/2024
This platform, created by Elon Musk, was previously hidden and is now endorsed by the government.	Vidéo	Facebook (The football-hub63)	https://www.facebook.com/story.php?story_fbid=122106270794404672&id=61562140167845&mibextid=WC7FNe&rid=fS6fbc2vofWvz5b3	Politique	7/24/2024

Attention à cette vidéo menaçant la France avant les JO et attribuée (à tort) au Hamas	Article professionnel	TF1 INFO	https://www.tf1info.fr/jeux-olympiques/fake-news-menace-russe-destabilisation-attention-a-cette-video-menacant-la-france-avant-les-jo-et-attribuee-a-tort-au-hamas-2310141.html	Politique	7/24/2024
Joe Biden call to Kamala Harris seems to be fake	Article professionnel	RESEMBLE.AI	https://www.resemble.ai/joe-biden-call-to-kamala-harris/	Politique	7/23/2024
Arcads AI Deepfake Video	Article professionnel	RESEMBLE.AI	https://www.resemble.ai/arcads-ai-deepfake-video/	Politique	3/25/2024
Cal Poly AI Voice Cloning Scam Case	Article professionnel	RESEMBLE.AI	https://www.resemble.ai/cal-poly-ai-voice-cloning-scam-case/	Éducation	3/12/2024
LastPass CEO Audio Deepfake Incident	Article professionnel	RESEMBLE.AI	https://www.resemble.ai/lastpass-ceo-audio-deepfake-incident/	Cybersécurité	4/14/2024
'Pathetic, laos' Maisug rallyists blasted for 'infecting platforms' with deepfake Marcos video	Article professionnel	Manila Standard	https://manilastandard.net/news/314473171/pathetic-laos-maisug-rallyists-blasted-for-infecting-platforms-with-deepfake-marcos-video.html https://www.youtube.com/watch?v=VPA1pPktJrE&t=10s	Politique	7/26/2024
I was deepfaked by my best friend'	Article professionnel	BBC	https://www.bbc.com/news/uk-68673390	Social	4/2/2024
Video of Yogi Adityanath Endorsing a Gaming Platform Is Fake	Article professionnel	Deepfakes Analysis Unit	Video of Yogi Adityanath Endorsing a Gaming Platform Is Fake	Social	7/24/2024
Video of Gautam Adani Endorsing a Gaming Platform Is Fake	Article professionnel	Deepfakes Analysis Unit	https://www.dau.mcaindia.in/blog/video-of-gautam-adani-endorsing-a-gaming-platform-is-fake	Social	7/23/2024
Videos Featuring Ambanis Promoting a Gaming App Are Fake	Article professionnel	Deepfakes Analysis Unit	https://www.dau.mcaindia.in/blog/videos-featuring-ambanis-promoting-a-gaming-app-are-fake	Politique	7/19/2024
Video Peddling A.I. Powered Trading Software By Musk Is Fake	Article professionnel	Deepfakes Analysis Unit	https://www.dau.mcaindia.in/blog/video-peddling-a-i-powered-trading-software-by-musk-is-fake	Politique	7/11/2024
Actress Sonakshi Sinha Targeted Through Deepfake Video	Article professionnel	Deepfakes Analysis Unit	https://www.dau.mcaindia.in/blog/actress-sonakshi-sinha-targeted-through-deep-fake-video	Divertissement et sports	7/9/2024
Video of Rajat Sharma Promoting Diabetes Cure Is Fake	Article professionnel	Deepfakes Analysis Unit	https://www.dau.mcaindia.in/blog/video-of-rajat-sharma-promoting-diabetes-cure-is-fake	Santé	7/9/2024
Elon Musk made a Kamala Harris deepfake ad go viral, sparking a debate about parody and free speech	Article professionnel	X (ex twitter), NBC News	https://twitter.com/elonmusk/status/1816974609637417112?ref_src=twsrc%5Etfw%7Ctwcamp%5Etwemblem%7Ctwterm%5E1816974609637417112%7Ctwgr%5E9281121040b56dc53fa9cd52d25a-d1bfc213e3e9%7Ctwcon%5Esl_c10&ref_url=https%3A%2F%2Fframe.nbcnews.com%2FQLMZqL%3F_showcaption%3D-trueapp%3D1 https://www.nbcnews.com/tech/misinformation/kamala-harris-deep-fake-shared-musk-sparks-free-speech-debate-rcna164119	Politique	7/26/2024
New Jersey teen sues classmate for allegedly creating, sharing fake AI nudes	Article professionnel	CBS News	https://www.cbsnews.com/news/new-jersey-teen-sues-classmate-for-allegedly-creating-sharing-fake-ai-nudes/	Éducation	2/9/2024
Fact Check: Indian High Commissioner In Security Meeting With Bangladesh PM? Find Truth Here	Article professionnel	NEWS MOBILE	https://www.newsmobile.in/nm-fact-checker/fact-check-indian-high-commissioner-in-security-meeting-with-bangladesh-pm-find-truth-here/	Politique	8/3/2024
Fact Check: Is That Rashmika Mandanna In Red Bikini Or Another Deepfake? Know The Truth	Article professionnel	NEWS MOBILE	https://www.newsmobile.in/nm-fact-checker/fact-check-is-that-rashmika-mandanna-in-red-bikini-or-another-deepfake-know-the-truth/ https://www.youtube.com/watch?v=xpgCqixCdK4	Social	5/27/2024
Nico Williams has finally made his decision on his future	Vidéo	Facebook	https://www.facebook.com/reel/490436156905653	Divertissement et sports	2024/08

Annexe B - Tableau des ressources hautement partagées sur les réseaux sociaux - Aînés

Title	Type	Source	Link(s)	Industry / context	Date of publication
Exposing the Dark Side of AI: How Deepfake Threatens Our Seniors!	Vidéo	TheGerontechnologist - Keren Etkin	https://www.youtube.com/watch?v=MX5So4rdviU	Finances et administration	2/23/2024
Sextortionists target senior citizen with retd IPS officer's deepfake	Article professionnel	THE TIMES OF INDIA	https://timesofindia.indiatimes.com/city/ghaziabad/retired-ips-officers-deepfake-used-to-blackmail-senior-citizen/articleshow/105603283.cms	Finances et administration	11/30/2023
Senior Citizens Fall Prey To AI-Based Deepfakes Tips To Protect Yourself From Cybercrime	Vidéo	ET NOW (via Youtube)	https://www.youtube.com/watch?v=e2A-6_IRiiM	Finances et administration	8/7/2023
Des fraudeurs auraient utilisé l'intelligence artificielle pour escroquer des aînés	Article professionnel	Radio-Canada	https://ici.radio-canada.ca/nouvelle/1965288/intelligence-artificielle-fraude-aines-terre-neuve	Finances et administration	3/22/2023

Annexe C - Tableau des ressources hautement partagées sur les réseaux sociaux - Femmes

Titre	Type	Source	Lien	Domaine	Date de publication
Scientists are falling victim to deepfake AI video scams — here's how to fight back	Article scientifique	nature	https://www.nature.com/articles/d41586-024-02521-3	Éducation	8/7/2024
«C'est moi Megan Lalonde»: la «vraie» fausse avocate passe aux aveux - reportage	Vidéo	TVA Nouvelles (via Youtube)	https://www.youtube.com/watch?v=IHLDAgCPUrA	Médias sociaux	7/22/2024
"Another Body": le combat de jeunes femmes victimes de deepfakes	Article professionnel	SWI swissinfo.ch.	https://www.swissinfo.ch/fre/%22another-body%22%3a-le-combat-de-jeunes-femmes-vic-times-de-deepfakes/73727949 https://www.youtube.com/watch?v=Cs3Wlf9BcEI https://www.radiolac.ch/culture/another-body-le-combat-de-jeunes-femmes-vic-times-de-deep-fakes/	Social	3/15/2024
Espagne : cinq jeunes mis en cause pour de fausses photos dénudées de jeunes filles générées par l'IA	Article professionnel	Le Parisien	https://www.leparisien.fr/international/espagne-cinq-jeunes-mis-en-cause-pour-de-fausse-photos-denudees-de-jeunes-filles-gene-rees-par-lia-05-08-2024-PUHVRI-TY2JBOJE3GRQJVCQDUE.php	Social	8/5/2024
Taylor Swift deepfakes: new technologies have long been weaponised against women. The solution involves us all	Article professionnel	The conversation	https://theconversation.com/taylor-swift-deepfakes-new-technologies-have-long-been-weaponised-against-women-the-solution-involves-us-all-222268	Social	1/26/2024
Deepfake ads featuring Jenna Ortega ran on Meta platforms. Big Tech needs to fight this.	Article professionnel	Mashable	https://mashable.com/article/jenna-ortega-meta-instagram-deep-fake	Social	3/6/2024
YouTuber Brooke Monk responds to sexual deep fake video with Taylor Swift reference	Article professionnel	The Mirror	https://www.themirror.com/entertainment/celebrity-news/youtuber-brooke-monk-res-ponds-sexual-497825	Diversissement	5/21/2024
Hundreds of sexual deepfake ads using Emma Watson's face ran on Facebook and Instagram in the last two days	Article professionnel	NBC NEWS	https://www.nbcnews.com/tech/social-media/emma-watson-deep-fake-scarlett-johansson-face-swap-app-rcna73624	Diversissement	3/7/2023

Volet juridique

Annexe D - Tableau des textes de loi pertinents

[Cliquez ici pour consulter cette annexe](#)

Annexe E - Comment lire une loi ou un projet de loi

A) Comprendre la structure d'une loi

Les projets de loi et les lois sont généralement organisés de manière similaire. Voici les principales sections que vous pouvez rencontrer :

- Titre : Indique le nom de la loi ou du projet de loi.
- Préambule : Explique les raisons et les objectifs de la loi.
- Articles : Les parties principales du texte, divisées en sections et sous-sections, énonçant les dispositions spécifiques.
- Dispositions finales : Incluent les dates d'entrée en vigueur et les éventuelles abrogations d'anciennes lois.

B) Lire attentivement le titre et le préambule

Le titre et le préambule donnent un aperçu général de l'objectif de la loi. Ils sont essentiels pour comprendre le contexte et les intentions des législateurs.

C) Analyser les articles

Les articles contiennent les règles et les prescriptions détaillées de la loi. Pour chaque article :

- Numéro de l'article : Identifiez le numéro et le titre de l'article pour une référence facile.
- Contenu : Lisez attentivement le texte de l'article. Notez les termes clés et les phrases importantes.
- Définitions : Recherchez une section de définitions souvent placée au début de la loi. Cela clarifie les termes spécifiques utilisés dans les articles.

D) Identifier les obligations et les interdictions

Recherchez les sections de la loi qui imposent des obligations (ce que vous devez faire) et des interdictions (ce que vous ne pouvez pas faire). Cela vous aide à comprendre les implications pratiques de la loi.

E) Noter les sanctions et les recours

Les lois incluent souvent des dispositions sur les sanctions en cas de non-respect et les recours possibles pour les personnes affectées. Ces sections sont cruciales pour comprendre les conséquences juridiques.

F) Prendre en compte les dispositions transitoires et finales

Les dispositions transitoires précisent comment la loi sera mise en œuvre et les dates d'entrée en vigueur. Les dispositions finales peuvent également abroger ou modifier d'autres lois existantes.

G) Utiliser des ressources complémentaires

Si le texte est complexe, n'hésitez pas à utiliser des commentaires juridiques, des guides explicatifs ou des analyses de professionnels du droit. Ces ressources peuvent fournir un contexte supplémentaire et des interprétations utiles.



Chaire de leadership en enseignement
sur les pratiques pédagogiques innovantes
en contexte numérique – Banque Nationale



UNIVERSITÉ
LAVAL