Le deepfake et son métadiscours : l'art de montrer que l'on ment

The deepfake and its metadiscourse: The art of showing lies

< Vivien LLOVERIA [1] >

1. CeReS - Université de Limoges, MICA - Université Bordeaux Montaigne
vivien.lloveria@unilim.fr

DOI : 10.25965/interfaces-numeriques.4876

< RÉSUMÉ >

Comment peut-on montrer le deepfake dans les médias ? Notre hypothèse est que toute communication sur le deepfake, textuelle ou visuelle, s'accompagne nécessairement d'un métadiscours plus ou moins volontaire et plus ou moins explicite pour « montrer que l'on ment ». À travers l'étude d'écrits journalistiques et d'exemplaires visuels, nous allons tenter d'explorer et de rendre compte de la variété de ces métadiscours, tant sur le plan de l'expression que sur celui du contenu.

< MOTS-CLÉS >

Deepfake, métadiscours, sémiotique, rhétorique visuelle, design d'information

< ABSTRACT >

How can deepfake be shown in the media? Our hypothesis is that any communication on deepfake, whether textual or visual, is necessarily accompanied by a more or less deliberate and more or less explicit metadiscourse to 'show lies'. Through the study of journalistic writings and visual examples, we will try to explore and account for the variety of these metadiscourses, both in terms of expression and content.

< KEYWORDS >
Deepfake, Metadiscourse, Semiotics, Visual rhetoric, Information design

"La nature a des perfections pour montrer qu'elle est l'image de Dieu, et des défauts pour montrer qu'elle n'en est que l'image."
Blaise Pascal, *Pensées*.

1. Introduction

Comment peut-on montrer le deepfake dans les médias ? Notre hypothèse est que toute communication sur le deepfake, textuelle ou visuelle, s'accompagne nécessairement d'un métadiscours plus ou moins volontaire et plus ou moins explicite pour « montrer que l'on ment ». À travers l'étude d'écrits journalistiques et d'exemplaires visuels, nous allons tenter d'explorer et de rendre compte de la variété de ces métadiscours, tant sur le plan de l'expression que sur celui du contenu.

2. Technologie, vérité des images et métadiscours du deepfake

Pour les informaticiens, le deepfake est d'abord le produit d'un deep learning, c'est-à-dire une forme d'apprentissage machinique exploitant des réseaux de neurones artificiels (Giles 2018) – les réseaux adverses génératifs ou GAN (en anglais generative adversarial networks) – capables de traiter de larges échantillons d'exemples en vue notamment d'imiter les expressions faciales, corporelles et vocales d'une personne (Westerlund 2019). Ce processus technologique engendre une rhétorique des images dès lors qu'il se nourrit des données visuelles numérisées de deux entités afin de créer des croisements (swapping en anglais) dont l'un devient la plupart du temps le manipulé de l'autre. On parle plus précisément de manipulation faciale (ou facial re-enactment en anglais) lorsqu'il s'agit d'un échange de visages (Kumar, Vatsa, et Singh 2020). Si l'on se réfère aux propos de Vincent Claveau (2020), plusieurs variantes composent la famille du deepfake :

- La simple synchronisation du mouvement des lèvres avec le discours d'une autre personne (lip-syncing en anglais) ;

- L'application des expressions du visage (face swapping en anglais) d'un acteur (le marionnettiste) sur le visage de la cible (la marionnette).
- La production d'images entièrement artificielles (des visages inédits).

La virtuosité du processus s'améliorant de jour en jour, la perception humaine n'est désormais plus en capacité de faire la distinction entre le vrai et le faux. Citons Vincent Claveau :

> *En vérité, il faut admettre que nous sommes entrés dans une ère où l'on ne peut plus croire la preuve vidéo. C'est un changement de paradigme car maintenant, n'importe qui peut truquer une vidéo. La démarche la plus sûre se résume à la vérification des sources. Qui envoie cette vidéo, puis-je lui faire confiance… ? (Claveau 2020).*

Eu égard à la littérature foisonnante sur la fabrique du faux dans les images, nous constatons que cette question n'est pas nouvelle dans l'horizon de la sémiotique visuelle. Ce qui l'est davantage, c'est la très grande difficulté voire l'impossibilité d'authentifier et de faire la vérité à l'œil nu. Le deepfake diffère donc du simple « montage » photographique ou vidéo dans son degré de véracité pour la perception humaine. A priori, seule une machine et son œil numérique détiennent la compétence de détection du faux au sein d'une production également artificielle.

La nature indétectable du deepfake nous permet ainsi de poser notre première question : si le deepfake ne saute pas aux yeux, comment les médias peuvent-ils rendre compte de son fonctionnement et de ses effets sur ses observateurs ? Si l'hypertrucage est invisible à l'œil humain, comment peuvent-ils parler d'une image qui ment parfaitement ? S'il n'y a pas de prises sensibles et de critères, au sens de Châteauraynaud et Bessy (1995), pour authentifier et dénoncer l'imposture, quelles sont donc les ressources mobilisées par les communicants pour sensibiliser à ce phénomène du deepfake ?

Notre hypothèse sera donc la suivante : tout discours médiatique sur le deepfake, image comme vidéo, s'accompagnera nécessairement d'un métadiscours destiné à rendre visible et compréhensible la falsification opérée par cette technologie (une forme de « dénonciation » textuelle et

visuelle). Cette franchise inévitable du deepfake au sein de l'espace médiatique nous permettrait, paradoxalement, de la considérer comme l'image fausse la plus honnête jamais observée. Condition de son existence médiatique, le deepfake rejoue visuellement le paradoxe du menteur en passant d'un « je dis que je mens » à un « je montre que je mens ».

Pour un sémioticien, ces dispositifs de présentation éclairée du deepfake constituent autant de manières d'énoncer et de « dénoncer » le trucage visuel. Autant de plans d'expression qui configurent, sur le plan du contenu, certains rapports à la vérité de l'image, articulant diverses interprétations et insérant cet acte de mentir dans un espace de valeurs coprésentes plus ou moins conflictuelles. Ainsi, notre perspective centrée sur la signification cherchera, par l'analyse des manifestations médiatiques du deepfake, à rendre compte de la variété des procédés métadiscursifs qui organisent à la fois les configurations du visible et les univers interprétatifs de ces images singulières.

3. La sémiotique des images et la véridiction

Pour la sémiotique Greimassienne, la vérité se situe dans les apparences et résulte de la *véridiction*, c'est-à-dire d'un ensemble d'opérations destinées à produire du vraisemblable, du « faire vrai » à l'intérieur du discours (Entrée Véridiction, 1979). Les combinaisons des modalités véridictoires de l'*être* et du *paraître* au sein du dispositif du carré sémiotique permettent alors, selon la définition donnée par Jacques Fontanille dans le second volume du dictionnaire de la théorie du langage de Greimas et Courtés (Entrée véridiction, 1986), de distinguer le « vrai » (le « dire ce qui est » et le « ne pas dire ce qui n'est pas ») du faux (le « dire ce qui n'est pas » et le « ne pas dire ce qui est »). Toutefois, le chercheur émet une critique intéressante pour notre étude à propos de la qualification de « mensonge » généralement attribuée au terme complexe *paraître* + *non-être*. Selon celui-ci, pour qu'il y ait « mensonge », la situation énonciative supposerait un écart volontairement introduit par l'instance d'énonciation entre un savoir présupposé (connu) et un savoir débrayé (montré). Ces dernières réflexions sémiotiques confirment ainsi la nécessité de recourir à un discours « tiers » venant dénoncer le mensonge en révélant un écart (entre le dire et les choses), sa nature (le

contenu de cet écart) et la localisation de sa source (l'instance énonçante ou plus simplement le « menteur »). C'est précisément ce métadiscours de « dénonciation visuelle » (Lloveria 2014) que nous placerons au centre de notre attention dans la médiatisation du deepfake.

Une fois la vérité définie comme un processus de véridiction par la sémiotique, se pose la question de sa transposition au sein d'une sémiotique des images. Selon Jean-François Bordron (2007), les formes de l'expression que génèrent les images n'échappent pas au jugement véridictoire dès lors qu'elles sont porteuses de sens. Le visible dans son ensemble n'est *a priori* ni vrai, ni faux, mais il le devient par les jugements portés sur ses contenus sémiotiques. Pour décrire les processus qui conduisent au statut véridictoire d'une image, Jean-François Bordron (2009) a eu recours à trois scénographies des images distinguant trois expériences cognitives et donnant lieu à trois statuts de la vérité qui leur sont propres.

La première scène réfère à l'expérience cognitive d'une *image horizon* qui, partant des indices d'existence d'une chose dans l'image, nous mène à la production de ses formes (phase d'iconisation) et débouche sur son identification (phase de symbolisation). Ce qui est évalué comme vrai dans cette image repose sur « ce qui, par elle, est fait image », c'est-à-dire son horizon. Le référentiel utilisé est alors celui de la perception quotidienne (le « monde naturel » de Greimas comme intersémioticité) ou celui plus spécialisé des rendus technologiques (par exemple dans le cas des images scientifiques, une autre image servira de référence à la chose évaluée).

La seconde expérience cognitive est celle de l'*image écriture* qui déplace le centre de gravité non pas sur la chose présentée ou représentée, mais sur l'image comme plan de l'expression codifié et plus ou moins conforme à des attentes. Le « vrai » n'est plus ici identifié mais *déchiffré* dans la présence d'écarts, dans le jeu des formants visuels qui font tenir l'image comme une totalité cohérente (la « textualité » du visible). Anne Beyaert-Geslin (2007) a déjà démontré que le statut de vérité des images pouvait dépendre de données plastiques telles que le flou ou le décadrage de la photographie et non plus des seuls contenus figuratifs, formulant l'hypothèse d'une *esthétique de la vérité*.

Enfin, la troisième expérience cognitive est celle de l'*image événement* qui situe la vérité de l'image dans la scène même de son effectuation. L'image n'est plus conforme à la chose (son horizon), ni à sa manière de l'exprimer (son écriture) mais à l'origine de sa production : « l'image est-elle bien ce qu'elle dit être ? ». Le jugement de vérité se focalise alors sur la recherche des *artefacts*, ces productions visuelles accidentelles qui n'ont finalement rien à voir avec le phénomène dont elles prétendent rendre compte visuellement.

Cette reprise des scénographies de l'image proposées par Jean-François Bordron nous a permis de faire une distinction plus claire au sein du métadiscours dénonçant le deepfake entre un premier rapport de conformité avec les choses du monde (accord/désaccord), un second avec l'écriture des images (cohérent/incohérent) et un troisième avec la source de la manifestation visuelle (légitime/illégitime).

4. Corpus et méthode d'analyse

Pour cette étude, notre corpus est organisé en deux grands ensembles :

Le premier ensemble est un sous-corpus textuel constitué de 159 articles journalistiques extraits de la base Europress avec les mots clefs « deepfake » et « deep fake » (en deux mots), de langue française entre 2018 et 2021. L'objectif de ce corpus était de décrire les figures, les thèmes et les axiologies qui structurent l'univers métadiscursif du deepfake, ainsi que les verbalisations des pratiques visuelles d'authentification de l'hypertrucage ;

Le second ensemble est un sous-corpus visuel rassemblant les images fournies par le moteur de recherche Google à la requête « deepfake » et « deep fake » (en deux mots) qui, une fois catégorisées, ont permis d'exemplifier plusieurs classes de métadiscours visuels.

L'usage de deux corpus a permis d'établir des comparaisons, de faire dialoguer, se confronter et se compléter les choix visuels et les jugements décrits par la textualité avec ceux montrés par les images.

5. Métadiscours volontaire/involontaire et implicite/explicite

Pour commencer, reprenant les commentaires de Jacques Fontanille sur la nécessité de créer un écart entre *savoir présupposé* et *savoir débrayé* dans la production d'un « mensonge », nous avons opéré quelques distinctions initiales afin d'articuler un peu plus notre corpus. La première a opposé un métadiscours *volontaire* et *involontaire*. La seconde, subdivision du métadiscours volontaire, opposa un métadiscours *explicite* et *implicite*.

5.1. Métadiscours volontaire et involontaire : deep fake vs cheap fake

La distinction entre métadiscours *volontaire* et *involontaire* mobilise la modalité du vouloir. Reconnaître une image comme un deepfake peut être le résultat d'un acte volontaire de construction d'un métadiscours ou celui d'un métadiscours qui s'installe *malgré-lui* dans l'image, lorsque « l'image résiste à la ressemblance » pour reprendre l'expression d'Anne Beyaert-Geslin (2004). D'un côté nous trouvons des images qui s'annoncent clairement comme un deep fake par un métadiscours assumé par l'instance d'énonciation, c'est le cas du deepfake annoté par des informations visuelles ou encore des situations impertinentes qui exposent clairement l'image comme un montage. De l'autre, nous trouvons le deepfake qui se veut indétectable et qui le devient malgré lui par la somme des indices figuratifs et plastiques qu'il laisse dans sa matière visuelle. Ces impertinences icono-plastiques, un métadiscours propre à *l'image écriture*, trahissent l'hypertrucage par des manifestations visuelles qui ne sont pas assumées par son énonciateur.

Cette distinction apparaît dans les articles de presse, notamment dans la différence que propose Monica Bickert, spécialiste chez Facebook, entre le deepfake et le « cheapfake » (Bickert 2020). Dans cette dernière catégorie, les images laissent involontairement traîner des indices qui trahissent leur appartenance à la catégorie du deepfake. Conséquemment, une quantité de métadiscours textuels et visuels, présents dans les médias, cherchent à expliquer au public la manière de *saisir* ces métadiscours de la falsification nichés dans les imperfections de l'image. Ces discours de second ordre fonctionnent comme des « méta-métadiscours » destinés à interpréter les traces de l'énonciation

mensongère à même l'image. Citons quelques exemples textuels de cette forme métadiscursive :

> *« Au premier coup d'œil, tout est là. […] Mais en y regardant d'un peu plus près, le doute s'installe […] Il faut avancer image par image pour voir quelques raccords ratés ou bugs visuels. Au moment où le golfeur retire ses lunettes de soleil, son visage est, le temps d'une fraction de seconde, différent, et sa main légèrement transparente. On voit également un nez derrière les lunettes. » (20 minutes, 27/02/2021)*
> *« Les plus élémentaires d'entre elles [deepfake] ne sont par exemple pas capables de simuler un clignement naturel des yeux. » (Les Echos, 11/03/2021).*
> *« D'autres indices doivent attirer l'attention, comme les incohérences d'éclairage entre le visage du personnage et le reste de son corps » (Les Echos, 11/03/2021).*
> *« […] regardez bien les oreilles ou les cheveux de la personne, ainsi que le profil du visage lorsque la tête tourne. S'il y a un décalage ou changement de forme ou de couleur, il y a des possibilités que ce soit un montage […] » (Le journal de Québec, 26/01/2020).*
> *[Réflexes pour repérer un deepfake] « Parfois le rendu est un peu flou alors que la vidéo est nette. Regardez attentivement les yeux ou la bouche de la personne qui parle, des zones traditionnellement complexes à reproduire pour un ordinateur » (Le Figaro, 29/06/2019)*

Inversement et paradoxalement, l'exemple de Dieudonné invoquant un deepfake pour se défendre lors d'un jugement pour outrage permet de démontrer que l'absence de signes de trucage visibles ne permet pas à la justice de statuer sur la présence d'un deepfake.

> *« Jugé pour outrage, Dieudonné invoque un "deep fake" » […] « La qualité et le décor de la vidéo montrent que ce n'est pas un trucage. Il n'y a aucun pixel, aucun flou, aucune décoloration, détails courants avec les deepfake […] » (Aujourd'hui en France, 15/03/2021).*

Le *cheapfake*, comme tentative ratée, vient finalement se placer dans une zone intermédiaire entre la catégorie du « montage » et celle du « deepfake ». D'un point de vue tensif, il exprime une forme de concession, se présentant comme l'imprévu (un événement) qui le fait choir de son statut de deepfake : bien que réalisé par un processus de deep learning, c'est un montage, car perceptible à l'œil nu.

L'ensemble des commentaires et annotations visuels communiqués par les médias engagent alors l'observateur dans une véritable chasse au cheapfake, se focalisant essentiellement sur les lieux possibles de la détection, les zones fragiles de la production du simulacre numérique sur lesquels il doit s'arrêter et prendre le temps d'une analyse détaillée. Le cheapfake interroge par exemple la totalité du visage, le « tout se tient » de son *image écriture*, dans la production de la forme et la cohésion de cette figure complexe.

5.2. Métadiscours volontaire implicite et explicite : rhétorique iconique et design d'information

Différemment, lorsque le deepfake est mis en scène pour être reconnu comme tel, et donc assumé par son énonciateur, deux formes de métadiscours peuvent alors être distinguées :

- La première est un métadiscours implicite car il se présente indirectement sous la forme d'une impertinence, d'un effet rhétorique. Il y a dans ces images une allotopie qui transgresse les règles habituelles de notre culture visuelle. Des visages qui ne sont pas les bons, des corps qui n'ont pas les gestes conformes aux visages, des visages engendrés par des célébrités fusionnées. Autant d'impertinences qui autorisent, par la rupture d'isotopie, la détection du deepfake et lui attribuent une intention particulière, celle d'obliger l'observateur à une réévaluation rhétorique, une relecture de l'image sous un angle humoristique, militant ou artistique par exemple.
- La seconde est un métadiscours explicite généralement présenté sous la forme d'apports de type infographique sur le support que devient le deepfake indétectable. Nous verrons en deuxième partie les différentes mises en scène que nous avons répertoriées telles que le masque, le cadre, le réseau qui expriment, chacune à leur manière, des contenus sémantiques et axiologiques différents.

5.3. Métadiscours implicite orienté n° 1 : la célébrisation

Le premier contact entre le deepfake et le grand public, du moins dans sa création, passe le plus souvent par des applications sur smartphone permettant d'insérer un visage familier dans des films pour remplacer des acteurs célèbres. Ici, l'isotopie s'élabore avec le contexte du film qui prend la place du fond, du degré conçu, à partir duquel va se dégager la figure impertinente d'un autre visage, familier mais moins célèbre. Cette rhétorique iconique basée sur une opération de substitution (Groupe μ et al. 1992) du visage familier à celui de la star inaccessible produit un contenu euphorique, un plaisir, de nature humoristique et ludique selon les quelques commentaires journalistiques qui suivent :

> *« Doublicat, l'application "deepfake" qui vous fait entrer dans la peau de votre star préférée » (Le Figaro, 20/07/2020).*
> *« […] les possesseurs de smartphones peuvent désormais échanger leur visage avec celui de leur star favorite et figurer dans leurs clips les plus célèbres » (Le Figaro, 20/07/2020).*
> *« Les résultats, particulièrement réalistes, provoquent l'allégresse des internautes sur les réseaux sociaux […] » (Le Figaro, 20/07/2020).*

Cette forme de deepfake, même parfaitement exécutée, n'a donc pas besoin de déléguer son analyse à une intelligence artificielle pour être détectée puisqu'elle se révèle elle-même, par le « devenir célèbre » (la *célébrisation*[1] du deepfake), cette figure rhétorique qui s'appuie sur le contraste entre un contenu fictionnel célèbre et l'incrustation d'un visage familier.

5.4. Métadiscours implicite orienté n° 2 : le marionnettisme

L'effet rhétorique peut également fonctionner en inversant la transformation. Un contexte inconnu, un corps inconnu supporte le visage d'une célébrité. L'effet est bien différent puisqu'il s'agit de prêter aux célébrités des comportements qu'elles n'ont pu avoir pour diverses

[1] Ce charmant néologisme date du 19ème siècle. Pour cela cf. Mercier (1801, 104)

raisons. Ce *marionnettisme*2 du deepfake, comme miniaturisation en une version manipulable de la célébrité, n'est pas sans rapport avec une certaine prise de pouvoir sur celle-ci. Citons pour cela Claude Levi Strauss à propos de la poupée comme outil de transformation de la personne en sujet :

> *« [La] transposition quantitative accroît et diversifie notre pouvoir sur un homologue de la chose : à travers lui, celle-ci peut être saisie, soupesée dans la main, appréhendée d'un seul coup d'œil. La poupée de l'enfant n'est plus un adversaire, un rival ou même un interlocuteur ; en elle et par elle, la personne se change en sujet » (Lévi-Strauss 1962, 38).*

Un premier exemple vidéo, celui de la Reine Elizabeth II, a besoin de peu d'explication. Réalisée par la chaîne d'information Channel Four, cette communication était justement destinée à sensibiliser le public sur les dangers du deepfake. La Reine s'y présente d'abord dans une posture conforme à ses habitudes, mettant en garde ses interlocuteurs du danger de la désinformation et achevant subitement son propos par une reprise endiablée de la danse de TikTok, debout sur son bureau. Cette franche rupture d'isotopie dans son comportement suffit à révéler l'hypertrucage. Les articles des journaux qui suivent également cet effet de surprise/révélation, insistent sur l'impertinence et justifient la production du faux par de louables intentions :

> *« Deepfake. La reine d'Angleterre apparaît dans un clip de sensibilisation… mais c'est un deepfake. » (Ouest France, 28/12/2020).*
> *« La reine reprend une danse de TikTok » (Ouest France, 28/12/2020).*
> *« Dans son faux discours, Elizabeth II s'exprime justement sur la véracité des images qui circulent sur les réseaux sociaux et dans nos smartphones » (Ouest France, 28/12/2020).*

D'autres utilisations de ce marionnettisme relèvent plus clairement du genre de la caricature en proposant des images où les puissants sont

[2] Ce terme de « marionnettisme », introduit par une citation de Claude Levi Strauss, a été emprunté à Pascal Le Maléfan dans Marty et Marie-Grimaldi (2004, 74)

tournés en dérision. Cette forme de deepfake, comme la précédente, trouve sa légitimité dans une visée informative et humoristique mais également critique.

Une autre forme d'incrustation de visages célèbres présente un caractère plus inquiétant et ambivalent : le deepfake qui donne vie à une personne « absente ». Selon les motifs de cette absence, les contenus axiologiques sont variables, allant de l'euphorie associée à l'animation de sujets artistiques célèbres dans le cadre de la promotion d'un musée, en passant par la simple neutralité lorsqu'il s'agit de remplacer une actrice malade de la COVID19 dans la série « Plus belle la vie », pour terminer sur l'effet plus dysphorique que peut engendrer la réanimation des photographies de personnes décédées comme le propose l'application en ligne DeepNostalgia (sur le site internet My Heritage).

> *« L'équipe du musée s'est amusée à faire passer les photos des portraits des collections par la moulinette de Womba Ai, une application de deepfake (ou hypertrucage) qui, par intelligence artificielle, anime et fait chanter un visage. Bluffant ! » (Ouest France, 24/03/2021)*
> *« Trucage à la rescousse. Pour pallier l'absence de l'actrice, "Plus belle la vie" a eu recours au "deepfake", une technique de montage novatrice » (Aujourd'hui en France, 17/11/2020).*
> *« MyHeritage, un site de généalogie en ligne, a mis au point un outil du nom de Deepnostalgia, qui permet grâce à des algorithmes de donner vie à vos photos en créant une vidéo d'hypertrucage (deepfake) qui anime le visage de vos proches, et même de vos ancêtres, de quoi vous donner la chair de poule. » (ICI Radio Canada, 26/02/2021)*

Enfin, l'assujettissement créé par ce marionnettisme du deepfake peut conduire à la mise en danger d'un individu, voire de la démocratie tout entière, dès lors qu'il met en scène de fausses conduites jugées immorales ou dégradantes ou encore en produisant de faux discours aux conséquences désastreuses.

C'est le cas de la journaliste indienne Rana Ayyub qui s'est retrouvée insérée dans un film pornographique, diffusé dans tout le pays après sa critique du parti au pouvoir :

> *« On imagine les harcèlements, pressions ou chantages possibles... La journaliste indienne Rana Ayyub en a fait l'amère expérience. En avril, après sa critique du parti au pouvoir, elle a vu son image intégrée à un film X. Elle raconte, mortifiée : "Il a fini sur presque tous les téléphones du pays. Moi, qui me disais journaliste et féministe, je n'arrivais pas à endurer cette humiliation. La seule façon que mes détracteurs ont trouvée pour me discréditer, c'est la misogynie et la diffamation." » (L'Obs, 12/01/2019)*

Ou encore celui de Joe Biden lui prêtant des propos racistes :

> *« Un montage vidéo comme celui qui prêtait au candidat démocrate Joe Biden des propos racistes en début d'année ne devrait plus avoir droit de cité sur Twitter. L'un des tweets de partage de cette vidéo avait été vu plus d'un million de fois. » (20 Minutes, 08/02/2020)*

5.5. Métadiscours volontaires implicites et réversibles : les croisons-les

Une autre forme de rhétorique du deepfake souligne le pouvoir de création par la recomposition et l'hybridation des visages. À la manière des mots valises dans le domaine littéraire, les « croisons-les » sont reconnaissables à la fois par un visage et un contexte célèbre. Ils deviennent des métadiscours réversibles dès l'instant où la direction de la transformation devient moins assurée.

Dans nos exemples, s'agit-il finalement d'un Trump devenu Kim Jung-Un ou de l'inverse ? Proche de l'interpénétration iconique décrite par le Groupe μ (1992), les « croisons-les » proposent en général de fusionner des figures possédant une certaine « intersection », soit dans l'apparence, soit dans le rôle thématique. C'est pourquoi les originalités capillaires nord-coréennes et américaines méritent, par le deepfake, une exploration particulièrement humoristique tout autant que l'inquiétante folie d'un Jim Carrey, digne héritier et remplaçant attendu de Jack Nicholson. Dans notre troisième exemple, c'est assez naturellement que Sylvester Stallone met sa musculature à disposition pour remplacer Arnold Schwarzenegger dans le film Terminator. Coiffure excentrique, visage de joker, corps musculeux sont autant de lieux d'intersection tant sur le plan

de l'expression (un lieu de substitution) que sur le plan du contenu (une amusante « presqu'équivalence »).

La couleur axiologique de cette forme de deepfake est globalement euphorique et le ton donné se veut ludique et humoristique. Ce dispositif icono-plastique, mobilisant tant l'image horizon que l'image écriture, fonctionne comme un jeu de langage visuel, une réécriture sur un mode hypothétique : « si Trump avait la même coiffure que Kim Jong-Un… », ou « si Sylvester Stallone avait joué dans Terminator… » avec l'objectif de faire ressentir cette « intersection » des images.

5.6. Métadiscours volontaires implicites et réversibles : les anonymés

Pour compléter les cases laissées libres par notre système d'analyse, il a fallu imaginer un croisement de personnes inconnues, autrement dit l'incrustation d'un visage inconnu sur un corps inconnu. Si le dispositif expressif est simple à décrire, les effets sur l'observateur nous paraissent d'un moindre intérêt dès lors que sont mis en scène de vrais inconnus. Sans indices dans les formants de l'image, le croisement entre des personnes inconnues de l'observateur peut tout à fait passer inaperçu (principe de non-pertinence) et ne pas conduire au calcul d'une figure rhétorique.

Pourtant, un exemplaire de cette forme de croisement nous indique un effet réel et puissant, permettant à la fois d'effacer l'identité de la personne initiale tout en sauvegardant ce qui fait son humanité : la force des émotions que peuvent dégager son corps et son visage en particulier. Le deepfake offre alors un anonymat numérique qui, tout en préservant l'identité source de l'humain manipulé, permet d'exprimer, de la manière la plus authentique possible, son intériorité. C'est le cas relevé dans notre corpus du reportage « » Welcome to Chechnya » (littéralement « Bienvenue à Tchétchénie »), un film documentaire sur la persécution des homosexuels en Tchétchénie, tourné à la fin des années 2010. Il se compose de témoignages réels de réfugiés tchétchènes LGBT dont les visages ont été anonymés mais non cachés, car les expressions faciales demeurent par l'incrustation de visages d'acteurs.

> *« […] garantir l'anonymat de quelqu'un souhaitant témoigner pour une raison ou une autre, tout en conservant ses émotions faciales. Sorti sur HBO il y a quelques semaines, le documentaire Welcome to Chechnya, qui traite des persécutions subies par les personnes LGBT en Tchéchénie, a adopté cette technique » (Le Figaro, 20/07/2020).*

6. Métadiscours volontaires explicites : le design d'information

À côté de ce métadiscours implicite du deepfake exprimé par une rhétorique icono-plastique des visages et des corps, nous trouvons une autre grande catégorie de métadiscours dits explicites car manifestés par tout un dispositif de mise en scène et de valorisation du processus d'hypertrucage. Dans ces dispositifs visuels, le deepfake devient le support d'un apport authentifiant et explicatif de l'image. Ces apports peuvent par exemple prendre l'apparence d'étiquetages, d'infographies ou de mises en comparaison d'images.

6.1. Métadiscours volontaires explicites : les coulisses de l'image

Le premier dispositif de métadiscours explicite est celui que nous dénommons « les coulisses de l'image ». Une fréquente forme de deepfake s'accompagne d'une série d'images « sources » et « cibles » qui en expliquent sa réalisation par un effet de contrastes engageant chez l'observateur un jeu de comparaison.

Figure 1. Face swap onto original work of art using Neural Net, 12 juin 2019

CC BY SA Stephen Wolfram

Ce dispositif comparatif permet la focalisation du regard sur les zones imitées telles que la bouche et certaines autres expressions du visage (à la manière d'un jeu des différences). Dans l'exemple mettant en scène la Reine Elizabeth II cité précédemment, la chaîne, Channel Four a voulu lever définitivement toute ambiguïté en présentant les coulisses du deepfake à la fin de son spot. À la manière d'un bug informatique, qui n'est pas sans rappeler les erreurs de la matrice dans le film Matrix, la scène initiale se « dégrade » visuellement, laissant progressivement se manifester les artefacts numériques qui trahissent les coulisses du deepfake.

6.2. Métadiscours volontaires explicites : les masques et la disponibilité du visage

Un second métadiscours explicite fréquent met en valeur le deepfake par la surimpression d'un masque plus ou moins transparent sur les visages présents dans l'image. Cette coprésence visuelle masque/visage, plus ou moins adversative, produit une troublante impression de mobilité, de possible désolidarisation, voire d'indépendance de la zone du visage humain, accréditant la possibilité de son interchangeabilité.

Par le deepfake, l'individu peut potentiellement perdre son identité et cette nouvelle forme d'anonymat endosse, selon les perspectives, des valeurs plus ou moins positives. Les communications médiatiques sur le deepfake présentent souvent son énonciateur négativement comme quelqu'un qui se cache derrière quelqu'un d'autre : usant du visage et des corps des autres à la manière des marionnettistes. Il n'est ainsi pas étonnant de repérer, dans les illustrations du deepfake, un personnage récurrent, masculin, plutôt jeune, masqué et retiré derrière une capuche. On peut également noter plusieurs exemplaires rejouant la même scène du requin masqué en poisson rouge.

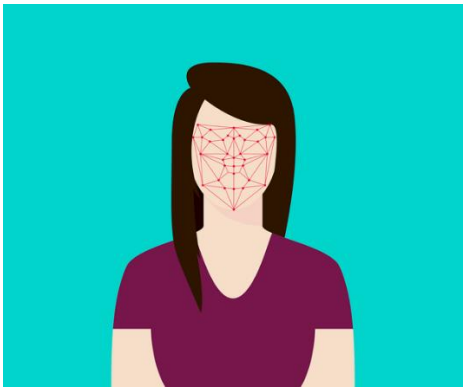
Dans les versions les plus neutres de ces dispositifs, la plupart des images présentent le deepfake sous la forme d'un masque qui, tout en laissant transparaître le visage singulier des individus, laisse la zone colorée localiser une certaine forme d'affordance (Morgagni 2011), une « disponibilité » pour un usage futur.

6.3. Métadiscours volontaires explicites : les cadres

Un troisième métadiscours explicite représente le deepfake accompagné de cadres verts et rouges. Il montre en général deux images identiques à l'œil nu dont l'une est réelle, car affublée d'un cadre vert, et l'autre fausse alors affublée d'un cadre rouge. Du point de vue de la sémiologie graphique de Bertin, (2013, 60) nous pourrions dire simplement que ces annotations visuelles utilisent 1) les deux dimensions de l'image pour localiser le deepfake et 2) la troisième dimension de la couleur (rouge ou verte) pour déterminer et communiquer le statut véridictoire de cette image (vraie ou fausse). Un jugement de valeur semble ici s'ajouter à la dimension informative (localisation) par l'utilisation du vert et du rouge, généralement associée aux valeurs éthiques de bien et de mal.

6.4. Métadiscours volontaires explicites : les réseaux

Un quatrième métadiscours explicite présente le deepfake accompagné d'un réseau surajouté sur les visages. Ces petits ensembles de points interconnectés peuvent être interprétés comme le rendu machinique ou numérique de la perception.



Pixabay Licence

Pour penser ses contenus, nous pourrions reprendre les mots de Michaël Fœssel et Antoine Garapon dans leur article dans la Revue Esprit intitulé « Biométrie : les nouvelles formes de l'identité » :

> *« Mais de quel corps s'agit-il ? En aucun cas du corps vécu qui ne fournit aucune base fiable de reconnaissance, ni même du corps physique qui est sujet, lui aussi, aux aléas du temps et à la possibilité de la dissimulation. Le corps dont s'occupent les techniques biométriques est un corps paradoxal puisqu'il est tout à la fois objectivé (réductible à des paramètres informatiques) et naturel (inaltérable). C'est même, en toute rigueur, un corps métonymique comme l'atteste l'importance de l'iris comme organe privilégié par la reconnaissance biométrique : cet organe, dont les paramètres biologiques sont inaltérables, désigne un principe de constance. »(Fœssel et Garapon 2006)*

Ce visage surajouté, traduction du regard machinique porté sur nos visages humains, énonce la dureté d'une version simplifiée de la réalité, dépouillée de son individualité car soumise aux divers filtres numériques, à leurs grilles objectivantes. D'une certaine manière, ces graphes de visage accomplissent à son paroxysme la vision anthropométrique des photographies d'identification judiciaire développée par Alphonse Bertillon au début du 20ème siècle (Maresca 2011). Nous ne sommes plus des visages porteurs d'émotions, mais une surface de valeurs différenciées, localisées, chiffrées et mises en relation. Ce tissage numérique du réseau du visage est à la fois fascinant et inquiétant car il représente la conversion, la traduction, de notre corps propre en son alter ego numérique, ce « principe constant » auquel nous n'avons pas accès et qui définit pourtant notre identité numérique.

6.5. Métadiscours volontaires explicites : les textures de visage

Pour terminer ce tour des manifestations métadiscursives du deepfake, nous voudrions aborder une catégorie un peu à part : la représentation du deepfake sous la forme de grandes collections d'images légèrement dissemblables, constituant de véritables murs de textures de visages. Ce qui saute aux yeux humains, c'est d'abord cette abondance d'images que notre système perceptif ne semble pouvoir embrasser d'un seul regard. C'est ensuite une impression plastique plus qu'une impression iconique qui nous saisit : il nous semble ici avoir davantage affaire à la matière continue de nos visages, à « du » visage, plutôt qu'à « des » visages perçus comme éléments discrets et

dénombrables. Cette catégorie d'image est intéressante dans le processus de médiatisation du deepfake car elle se présente comme une tentative d'illustrer le processus informatique du « deep learning » en cherchant à mettre en scène son pouvoir de rapprochement entre les images.

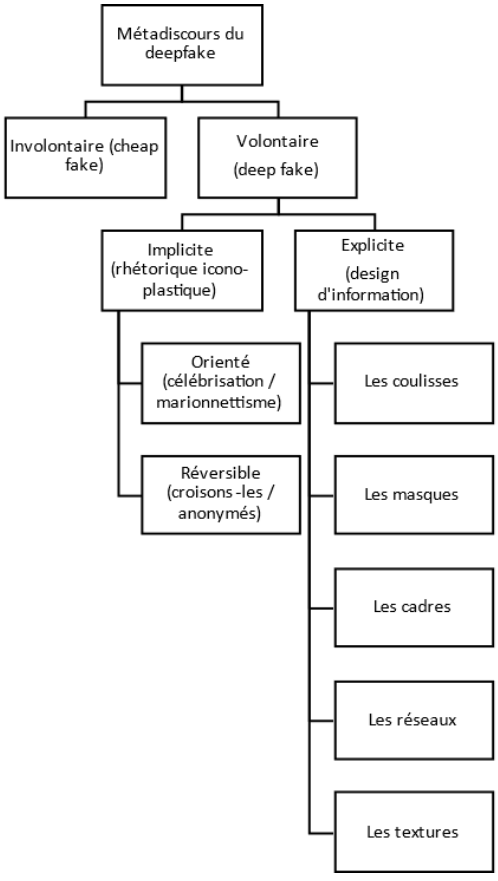
Figure 2. Real Blurred Avatars People Character Many Faces

CC0 Domaine public

Vous partageant une part de mon imaginaire, ce deep learning m'apparaît alors comme un iconophage - mot que j'emprunte à Jérémie Koering (2021) -, un organisme qui ingère et digère nos images, les mémorise, les filtre et n'en laisse que peu de résidus. Ces ensembles d'images ont l'étrangeté d'apparaître comme la nourriture ou le carburant de cet apprentissage profond. Chaque occurrence n'a finalement que peu de valeur car c'est le lissage exercé par la répétition des images qui importe, procurant la fluidité nécessaire aux futures énonciations de l'hypertrucage. L'image devient un agrégat de symboles, s'insérant dans des lignes de codes infinies, sans cesse réactualisées par les nouvelles occurrences et articulées autour de l'unique fonction du simulacre. Pour un sémioticien, cette multiplicité semble vouloir rendre compte du paradigme de nos visages, dont l'exhaustivité garantirait le plus parfait déploiement syntagmatique du deepfake.

7. Conclusion

Cette contribution, sous la forme d'une liste de métadiscours articulée selon le vouloir et le degré d'explicitation du deepfake, ne peut prétendre à une quelconque exhaustivité. Toutefois, elle aura permis de dégager un certain nombre d'exemplaires représentatifs de catégories de métadiscours caractérisés par la récurrence de leur plan d'expression (dispositifs iconiques et plastiques) et une certaine stabilité dans les contenus thématiques et axiologiques. Résultats que nous pourrions synthétiser par le schéma suivant :



Bibliographie

Bertin, Jacques. 2013. *Sémiologie graphique : les diagrammes, les réseaux, les cartes*. Paris : Éd. EHESS.

Bessy, Christian, et Francis Chateauraynaud. 1995. *Experts et faussaires : pour une sociologie de la perception*. 1 vol. Leçons de choses. Paris : Métailié.

Beyaert-Geslin, Anne. 2004. « Crénelage, capiton et métadiscours (où l'image numérique résiste à la ressemblance) ». *Protée* 32 (2) : 75-83. https://doi.org/10.7202/011175ar.

———. 2007. « Une vérité provisoire ». *Actes Sémiotiques*, décembre. https://www.unilim.fr/actes-semiotiques/index.php ?id =3348.

Bickert, Monika. 2020. « Enforcing Against Manipulated Media ». *About Facebook* (blog). 7 janvier 2020. https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

Bordron, Jean-François. 2007. « Image et vérité ». *AS - Actes Sémiotiques*. https://www.unilim.fr/actes-semiotiques/3355.

———. 2009. « Expérience d'objet et expérience d'image. » *Visible : Images et dispositifs de visualisation scientifiques*, nº 5 : 111-22.

Claveau, Vincent. 2020. « Deepfake : peut-on s'en prémunir ? » *Techniques de l'Ingénieur*, mars. https://www-techniques-ingenieur-fr.ezproxy.unilim.fr/actualite/articles/deepfake-peut-on-sen-premunir-76522/.

Courtés, Joseph, et Algirdas Julien Greimas. 1979. *Sémiotique : dictionnaire raisonné de la théorie du langage*. Hachette université. Paris : Hachette.

Fœssel, Michaël, et Antoine Garapon. 2006. « Biométrie : les nouvelles formes de l'identité ». *Esprit* Août/septembre (8) : 165-72. https://doi.org/10.3917/espri.0608.0165.

Giles, Martin. 2018. « The GANfather : The Man Who's given Machines the Gift of Imagination ». MIT Technology Review. 21 février 2018. https://www.technologyreview.com/2018/02/21/145289/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination/.

Greimas, Algirdas Julien, et Joseph Courtés, éd. 1986. *Sémiotique : Dictionnaire raisonné de la théorie du langage. Tome 2*. Paris : Hachette Éducation.

Groupe µ, Francis Edeline, Jean-Marie Klinkenberg, et Philippe Minguet. 1992. *Traité du signe visuel : Pour une rhétorique de l'image*. La couleur des idées. Paris : Du Seuil.

Koering, Jérémie. 2021. *Les iconophages : une histoire de l'ingestion des images*. Arles, France : Actes Sud.

Kumar, Prabhat, Mayank Vatsa, et Richa Singh. 2020. « Detecting Face2Face Facial Reenactment in Videos ». *2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Applications of Computer Vision (WACV), 2020 IEEE Winter Conference on*, mars, 2578-86. https://doi.org/10.1109/WACV45572.2020.9093628.

Lévi-Strauss, Claude. 1962. *La pensée sauvage*. Paris : Plon.

Lloveria, Vivien. 2014. « (D)énonciation visuelle et coopération de l'image ». *Lexia*, nº 17-18 (venir).

Maresca, Sylvain. 2011. « Spécimens ou individus ?. Les usages incertains du portrait photographique ». *L'Homme. Revue française d'anthropologie*, nº 198-199 (juillet) : 67-87. https://doi.org/10.4000/lhomme.22690.

Marty, François, et Hélène Marie-Grimaldi. 2004. *L'aventure de la recherche en psychologie clinique et psychopathologie*. Presses universitaires de Rouen et du Havre.

Mercier, Louis-Sébastien. 1801. *Néologie ou Vocabulaire de mots nouveaux, à renouveler, ou pris dans des acceptions nouvelles*. Moussard.

Morgagni, Simone. 2011. « Repenser la notion d'affordance dans ses dynamiques sémiotiques ». *Intellectica*, nº 55 : 241-67.

Westerlund, Mika. 2019. « The Emergence of Deepfake Technology : A Review ». *Technology Innovation Management Review* 9 (11) : 40-53. https://doi.org/10.22215/timreview/1282.