Images, mensonges et algorithmes. La sémiotique au défi du *Deepfake*

Images, falsehoods and algorithms. Semiotics faced with Deepfake

< Ludovic Châtenet [1] >

1. MICA, Université Bordeaux Montaigne
CeReS, Université de Limoges
Ludovic.chatenet@u-bordeaux-montaigne.fr

DOI : 10.25965/interfaces-numeriques.4826

Remerciements à Anne Beyaert-Geslin et Andréa Alexander

1. Présentation du numéro

Depuis quelques années les nouveaux instruments de communication, largement formatés par les algorithmes issus du développement de l'IA, ont donné lieu à de nouvelles formes d'inscription et circulation du sens. Les numéros 1/2020, puis les 2 numéros de 2021 d'*Interfaces Numériques* respectivement intitulés *Le design de l'« intelligence artificielle » à l'épreuve du vivant* et *Représentation(s) et Numérique(s) (Parties I et II)* ont contribué à circonscrire un cadre épistémologique autour des données et des pratiques communicationnelles. Le présent numéro entend poursuivre ces réflexions en abordant les problèmes posés par les représentations algorithmiques et numériques, plus particulièrement les *deepfakes,* et leur influence sur la construction du sens dans nos sociétés.

À ses origines dans les années 1950, l'IA s'est définie comme un système de simulation des fonctions cognitives humaines par le calcul

(perception, organisation des connaissances, apprentissage, raisonnement, communication, décision), permettant une représentation symbolique de ces données (McCarthy, 1955). La sémiotique s'est intéressée au sujet dans les années 1980 (*Actes Sémiotiques* n°36, 1985 ; *Actes Sémiotiques* n°40, 1986) en questionnant d'abord ces modèles formels et l'organisation de la connaissance appliquée aux textes avant d'identifier un point de convergence au niveau de l'image. Les modèles sémiotiques ont ainsi servi de support à l'IA (Thérien, 1989 ; Arnold, 1989) désormais considérée comme une « machine voyante », assimilée à un organisme capable de reconnaître des images (d'objets, de visages) mais aussi d'en produire. L'IA est alors apparue comme un modèle de médiation (Rialle, 1989 ; Bachimont, 1993) capable d'interpréter mais aussi de représenter, autrement dit d'énoncer. Toutefois, cette capacité énonciative se trouve accrue par le *Deep Learning* (Le Cun, 1985, 2019) qui donne une autonomie aux réseaux de calculs et leur permet entre autres de reproduire les gestes et les émotions.

Ces dernières années, ces développements ont contribué à l'émergence, du phénomène des *deepfakes.* Le terme désigne des images (ou vidéos) fausses fabriquées par synthèse (substitution, fusion) d'images existantes (Caldera, 2019 ; Burkell et Gosse, 2019) aux moyens de l'*IA*, du *Deep Learning* et des *GAN* (*generative adversarial networks*, cf. Goodfellow 2014). Les *deepfakes* ont émergé grâce au stockage massif de données, notamment d'images, sur internet depuis les années 1990. Dans ce cadre, les réseaux sociaux ont offert à la fois la « matière première » nécessaire à leur composition (images, vidéos, sons) et les moyens de leur diffusion massive dans l'espace médiatique. Si bien que les *deepfakes* peuvent être considéré aujourd'hui comme la manifestation, ou la production langagière, la plus aboutie du domaine numérique considéré comme espace du « faux absolu » (Leone 2021).

D'abord exploité à la marge, dans le domaine de la pornographie pour donner corps aux imaginaires et aux fantasmes, les *hypertrucages* se sont progressivement installés dans le paysage numérique quotidien où ils circulent sur les plateformes de partage social (Youtube, Facebook, Twitter, etc.). Parmi ces usages, on relèvera d'un côté les pratiques ludiques, qui permettent par exemple à un usager de se projeter à la place d'un acteur dans un extrait de film (application *Face Swap*), mais aussi la

production de mèmes internet, de pastiches ou de caricatures (Nicolas Canteloup dans *C'est Canteloup ! La grande métamorphose* sur TF1) ou des visées artistiques pour recréer ou simuler des styles picturaux (« *votre portrait dans le style de Van Gogh* »). De l'autre côté, on assiste à des usages problématiques susceptibles de tromper intentionnellement un observateur ; des exemples récents d'*hypertrucages* mettant en scène des personnalités publiques, notamment politiques, comme Barack Obama ou Donald Trump, introduisent en effet de l'ambiguïté dans le discours médiatique et alimentent la « crise de la vérité ». À cet endroit, les *deepfakes* questionnent de manière directe la vulnérabilité des cultures, de la communication, jusqu'à la *vraisemblance* des images, reposant sur un *croire* ou *faire croire* (Greimas, 1983).

Peut-on encore croire les images ? comment identifier les faux ? Pour apporter des éléments de réponse, les textes recueillis dans ce numéro entendent interroger la manière dont les *deepfakes* produisent du sens (énonciation) mais aussi comment ils révèlent ou jouent de la falsification propre à leur modèle de simulation. L'examen de la nature sémiotique de ces objets numériques et de leur rapport à la vérité s'articulera principalement autour de la notion de *vraisemblance*.

Les *deepfakes* sont en effet des trucages si crédibles qu'on ne peut distinguer l'authentique du contrefait, au point que ces images peuvent avoir un impact important sur la société. Cette « transparence » de l'énonciation de la machine pose à la sémiotique le problème de l'instauration du plan de l'expression lorsque la limite entre *être* et *paraître* est brouillée. Elle situe plus précisément nos préoccupations au point d'articulation entre, d'un côté, la production par la machine, approchée comme un *faire paraître vrai*, c'est-à-dire une manipulation (persuasion) lors de la mise en forme, et de l'autre côté, un *croire vrai* relatif à la compétence interprétative, à l'expertise de l'observateur (cf. Greimas et Courtés, 1979 : 422-423). Du point de vue de ce dernier, il s'agira de saisir comment on peut distinguer le vrai et le faux ou identifier les indices ou les traces d'une manipulation (Bessy et Chateauraynaud 1995, Eco 1987, Leone 2021). Une question analogue peut se poser au chercheur confronté aux d*eepfakes* : quels critères (indices, traces) mobiliser pour constituer un corpus lorsque l'on n'est pas capable de distinguer une image « authentique » d'une image « truquée » ?

Ces questions sollicitent tout particulièrement la sémiotique et l'inviter à réinterroger ses outils pour décrire et circonscrire les mécanismes de production du sens des nouveaux objets numériques. L'enjeu principal semble résider dans l'examen de leur statut « vague » (vrai/faux) qui, en opérant un croisement entre l'*hyperréalité* (Eco 1985) et l'*imaginaire*, réinterroge notre conception même de la signification toujours fluctuante entre les faits et les fictions, et sa dimension éthique. Les difficultés posées par les *deepfakes* sont d'autant plus prégnantes qu'elles concernent l'ontologie des faits, et des corps (visage) constitués ; l'inversion du rapport entre la réalité et l'image s'opère au risque de la naturalisation de ces objets qui semblent déjà dicter leur esthétique.

Dans une optique de dialogue pluridisciplinaire, les contributions présentées proposeront tour à tour des réflexions générales sur le faux et l'authenticité. Elles discuteront ensuite le *Deepfake* appliqué aux visages numériques. Enfin, le *Deepfake* sera également approché du point de vue de l'art et interrogera sa capacité générative et créative, notamment sa capacité à devenir co-énonciatrice d'œuvres.

2. Contributions

2.1. La question du fake à l'ère du numérique

La question générale du faux (*fake*) et de la « croyance » du public qui y est exposé est abordée par **JULIEN NELSON** qui fournit un cadre général aux réflexions de la sémiotique sur les *fakes news* et les *deepfakes,* en s'appuyant sur les connaissances actuelles en psychologie sociale. À travers des exemples d'études sur la crédulité face à des contenus trompeurs, Nelson montre que la « vraisemblance perçue » de l'information est influencée d'une part, par la diffusion massive des items (répétition), d'autre part par les croyances et convictions politiques d'un individu. La crédibilité de l'information et la vulnérabilité aux *fakes* reposent sur des mécanismes de biais sur lesquels les médias (et réseaux sociaux) peuvent agir. En outre, les *hypertrucages* suscitent une crise de confiance reposant sur l'incertitude de l'individu exposé à un faux.

Les contributions de Bruno Bachimont et Massimo Leone ont en commun de mettre en évidence les problèmes liés au « faux numérique » et de proposer des pistes de recherche pour les circonscrire. Dans un premier temps, **BRUNO BACHIMONT** met en évidence le problème de l'authenticité des documents numériques. En les définissant comme des enregistrements, il montre que ces derniers transforment leurs objets et « coupent » le lien indiciel qui devrait permettre de garantir leur fidélité à ces derniers. Le problème repose sur la nature algorithmique des documents numériques qui, comme le souligne Bachimont, ne permet pas de fonder son authenticité sur l'interprétation du support, comme c'était le cas avant. Pour y parvenir, il propose d'étudier les diverses manifestations du code, en développant notamment différents domaines d'expertises dédiés : codicologie, *forensics*, philologie du numérique. Dans le sillage de ces réflexions, **MASSIMO LEONE** argumente la nécessité d'une sémiotique du faux numérique, situé dans le cadre plus large d'une crise de la véridiction, en prenant appui sur ses travaux récents sur les « faux visages numériques ». Les nouveaux dispositifs numériques, algorithmiques (GAN), contribuent à une « spirale du faux » : une circulation croissante de produit dont la fausseté est de plus en plus difficile à identifier par l'observateur humain. En examinant le modèle des GAN de Goodfellow, Massimo Leone dégage les propriétés du faux numériques, reposant principalement sur la reproductibilité et diffusibilité massive d'items susceptibles de représenter des objets non-existants. L'omniprésence des faux numériques et le danger qu'ils apportent à la société (notamment les médias) conduit Leone à souligner la nécessite de la sémiotique comme outil d'expertise pour la « détection des faux ».

2.2. Deepfake et visages numériques

Le visage apparait comme un enjeu majeur de la production de faux numériques, notamment lorsque ces derniers cherchent à tromper l'observateur. La représentation du visage par les *deepfakes* conduit (i) à un examen sémiotique des mécanismes, ou stratégies, au cœur de leur efficace ; qui permet, ensuite, (ii) de mettre ne évidence les éléments qui, de manière implicite ou explicite, révèlent le trucage. Partant d'une réflexion sur le portrait, **MARIA GIULIA DONDERO** met en relation la

« photographie composite » de Galton et le *Deepfake*. Elle considère que ce dernier a perfectionné la standardisation de l'individu au moyen d'algorithmes analysant de larges bases de données, puis réalisant les mêmes opérations de superposition et fusion des identités. Dondero porte une attention particulière à la notion de « flou » qui, émergeant durant le processus d'extraction d'une image du visage, permet le passage d'une identité à une autre tout en trahissant le trucage. La crédibilité des *deepfakes* est donc limitée par ces effets de flou qui donnent aux « visages numériques » le statut de « masque ». La remise en question du statut du visage par les *hypertrucages* conduit **MAXIME FABRE** à étudier comment ils remettent en cause le statut de « la vérité » dans l'image numérique. Ses réflexions montrent que les *deepfakes* produisent des « interprétations de visages » (des visagéités purement médiatiques), jouant de l'évidence même des visages pensés comme « naturels ». En adoptant le point de vue de la sémio-pragmatique peircienne (théorie du vague), Fabre souligne qu'ils exploitent les représentations des visages sur les réseaux pour produire des visages « médiagéniques », des « faces » qui semblent s'imposer aux visages naturels pour mieux leur ressembler. L'article de **VIVIEN LLOVERIA** met à contribution les outils sémiotiques pour l'identification des faux numériques. Comme l'ont également noté d'autres auteurs du numéro, il part du principe que les *deepfakes* révèlent qu'ils sont faux à cause des effets rhétoriques visuels produits par leur nature d'agencement ou synthèse d'autres images. A partir de la notion de *vraisemblance* en sémiotique, Lloveria propose d'étudier les méta-discours, de « dénonciation visuelle », des médias pour rendre visible les trucages des *deepfakes*. En questionnant la conformité des images à leur objet, leur mode de représentation ou de production, il propose une typologie convaincante, étayée par des exemples, des « manières de montrer le faux » des *deepfakes* (involontaire ou volontaire-implicite et volontaire-explicite).

2.3. De l'art de faire sens : ludicité du fake et énonciation

Si le *Deepfake* a d'abord été considéré du point de vue du « faux », au sein d'une crise de la vérité, il peut aussi être approché comme lieu de « jeu » et questionner la production même du « sens » (énonciation). Le domaine de l'Art, notamment l'esthétique du faux, offre un terrain

critique et expérimental pour questionner à la fois le rôle du faux et du contrefait, mais aussi la capacité des algorithmes à produire du sens. Dans son texte très documenté sur les pratiques artistiques, **STEFANIA CALIANDRO** montre comment le *fake* est exploité pour le plaisir ludique et offre à l'art un terrain de jeu favorable à l'imaginaire esthétique. Dans le cadre spécifique de l'Art, Caliandro souligne que les *deepfakes* permettent la création de *faux-originaux*. Ils mettent en effet le faux au service d'une démarche créative, de la concrétisation d'un potentiel de signification et des mondes possibles. Finalement, les dispositifs fondés sur l'IA permettent aujourd'hui à l'art, au *fake art*, d'exercer le rôle critique de la démarche artistique en montrant et questionnant la tromperie des techniques de l'image. La contribution de **NICOLE PIGNIER** déplace la question de l'usage ludique du *Deepfake* vers sa dimension éthique. La notion de *vérité* est abordée à nouveaux frais dans le cadre de l'énonciation pour interroger la nature communicationnelle des *deepfakes*, notamment ses limites. Un énoncé *deepfake* ne ment ni ne dit la vérité mais dans la mesure où ils sont produits par une *praxis* humaine ils expriment la vérité des manipulations (*faire-faire*, *faire-croire*) opérées sur les images. Si l'IA, non-vivante, ne semble pouvoir co-énoncer, re-présenter une expérience sensible, elle produit donc une énonciation feinte. Pignier souligne qu'elle peut s'avérer problématique lorsque les trucages s'affranchissent de la réciprocité de la communication – garante de l'éthique sociale – pour et font dire à l'autre ce qu'on veut. L'entretien de l'artiste-chercheuse **ALIZÉE ARMET**, présenté dans ce numéro, s'inscrit dans le même cadre. En effet, en utilisant les algorithmes pour ses installations artistiques, Alizée Armet remet en question le processus de création tout en rediscutant l'autonomie des algorithmes. La question d'une co-énonciation de l'IA est posée en pratique puisqu'Armet l'expérimente en laissant les algorithmes produire eux-mêmes des objets non-préconçus, sans trop de programmation. L'IA, considérée comme forme « vivante » apparait comme un moyen d'explorer des mondes possibles ; elle conduit également des nouvelles formes de production, collectives, décentrées et participatives.

Bibliographie

Arnold Madeleine (1989). « La sémiotique : un instrument pour la représentation des connaissances en intelligence artificielle ». *Études littéraires*, *21* (3), pp. 81–90.

Bachimont Bruno (1993). « Nature, Culture et Artefacture : la place de l 'intelligence artificielle dans les sciences cognitives ». *Intellectica. Revue de l'Association pour la Recherche Cognitive*, n°17, 2, pp. 213 - 238.

Bessy Christian et Chateauraynaud Francis (1995). *Experts et faussaires : pour une sociologie de la perception*, Métailié, Paris.

Burkell Jacquie et Gosse Chandell (2019). « Nothing new here: Emphasizing the social and cultural context of deepfakes". *First Monday*, *24*(12).

Caldera Elizabeth (2019). « Reject the Evidence of Your Eyes and Ears: Deepfakes and the Law of Virtual Replicants », *Seton Hall Law Review*: Vol. 50: Iss. 1, Article 5.

Eco Umberto (1985). *La Guerre du faux*, Paris, Grasset & Fasquelle.

Eco Umberto (1987). "Fake, identity and the real thing", *Special issue of Versus*, 46, Bompiani, Milan.

Goodfellow Ian J. *et al.* (2014). « Generative Adversarial Networks » ; [en ligne] https://arxiv.org/abs/1406.2661

Greimas Algirdas J. (1983). *Du Sens II*, Seuil, Paris.

Le Cun Yann (1985). « Une procédure d'apprentissage pour réseau a seuil asymmetrique (a Learning Scheme for Asymmetric Threshold Networks) », *Proceedings of Cognitiva 85*, 599-604, Paris, France.

Le Cun Yann (2019). *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob, Paris.

Leone Massimo (2021). Prefazione / Preface. Dans Leone, Massimo, dir. 2021. *Volti artificiali / Artificial Faces*, numéro spécial de *Lexia* : *International Journal of Semiotics*, vols 37-8, Rome, Aracne, pp. 9-25.

Rialle Vincent (1989). « IA et sujet humain : entre physis et sémiosis ». *Intellectica*, 1996/2, 23, pp.121-153.

Stockinger Peter (Dir.), (1985). « Intelligence artificielle et théorie sémio-linguistique », *Actes Sémiotiques*, n°36.

Stockinger Peter (Dir.), (1986). « Intelligence artificielle, tome II : approches cognitives du texte », *Actes Sémiotiques*, n°40.

Thérien Gilles (1989). « Sémiotique et intelligence artificielle ». *Études littéraires*, *21* (3), pp. 67–80.