Les professionnels de l'intelligence économique face aux nouvelles technologies de falsification

Par Christophe Deschamps, Laboratoire CEREGE, Université de Poitiers.

Résumé :

Depuis 2017, année où les premiers *deepfakes*, ces vidéos truquées grâce à l'intelligence artificielle, ont commencé d'émerger, les risques qu'ils représentent deviennent de plus en plus concrets. Or, ces technologies reposant sur l'apprentissage profond (*deep learning*) se multiplient et s'appliquent dorénavant à tous types d'images. La recréation via intelligence artificielle s'applique également à la voix qui, via clonage numérique, peut être reproduite à l'identique de plus en plus aisément. Enfin, plus inquiétant, des écrits parfaitement crédibles peuvent dorénavant être aisément générés en quantité et degrés de variation infinis par des modèles de langage adaptés. Cela ouvre la voie à d'innombrables possibilités utiles mais également à des usages nuisibles tels que la création de campagne de désinformation sur les réseaux sociaux ou l'usurpation d'identité par téléphone.

A partir d'une étude de la littérature nous dressons un panorama de ces technologies et de leurs impacts potentiels sur les organisations lorsqu'elles sont mises au services d'attaques informationnelles les visant. Nous en déduisons plusieurs implications pour les professionnels de l'intelligence économique : la nécessité d'intégrer ces menaces à la compréhension qu'ils ont de leur environnement afin de mieux les anticiper et d'y réagir plus efficacement ; l'importance d'entraîner et d'entretenir leur esprit critique afin de ne pas tomber dans les pièges de la désinformation et la nécessité de connaître et maîtriser les méthodes et outils susceptibles de les aider à détecter ces attaques informationnelles.

Mots-clés :

« intelligence économique », « veille stratégique », « désinformation », « technologies du faux », « intelligence artificielle »

Introduction

"Le mensonge vole et la vérité la suit en boitant, de sorte que lorsque les hommes ne sont plus trompés, il est trop tard, la plaisanterie est terminée et le conte a eu son effet...". Jonathan Swift

Depuis 2017, année où les premiers *deepfakes*, ces vidéos truquées grâce à l'intelligence artificielle, ont commencé d'émerger, les risques qu'ils représentent deviennent de plus en plus concrets. Or, ces technologies reposant sur l'apprentissage profond (*deep learning*) se multiplient et s'appliquent dorénavant à tous types d'images. La recréation via intelligence artificielle s'applique également à la voix qui, via clonage numérique, peut être reproduite à l'identique de plus en plus aisément. Enfin, plus inquiétant, des écrits parfaitement crédibles peuvent dorénavant être aisément générés en quantité et degrés de variation infinis par des plateformes comme GPT-3 ou MT-NLG (Deschamps, 2021). Cela ouvre la voie à d'innombrables possibilités utiles mais également à des usages nuisibles tels que la création de campagnes de désinformation sur les réseaux sociaux ou l'usurpation d'identité par téléphone.

Nous proposons d'étudier dans cet article les différents usages potentiels ou avérés des technologies du faux lorsqu'elles sont mises en œuvre pour désinformer, influencer ou dénigrer des organisations. Après avoir précisé qui sont les professionnels de l'intelligence économique et ce que nous entendons par « désinformation », nous décrirons les différents types de technologies du faux existantes. Dans un troisième temps, nous verrons quels risques et impacts elles sont susceptibles d'entraîner lorsqu'elles ciblent les organisations et présenterons un modèle global d'offensive les utilisant. Enfin, nous évoquerons les méthodes de prévention et de détection existantes et ce qu'elles impliquent en terme d'évolution des métiers de l'intelligence économique.

Qui sont les praticiens de l'IE

Si, dans ses acceptions anglo-saxonne de "*business intelligence*" puis de "*competitive intelligence*" (Goria, 2006), l'on peut faire remonter l'histoire de l'intelligence économique à la fin des années 50, l'expression n'est proposée en France qu'en 1992 par Christian Harbulot pour désigner « *toutes les opérations de surveillance de l'environnement concurrentiel : veille, protection, manipulation de l'information (leurre, contre-information, ...), influence* » (Harbulot, 1992). Elle sera popularisée deux ans plus tard par le Rapport Martre qui indique que « *L'intelligence économique peut être définie comme l'ensemble des actions coordonnées de recherche, de traitement et de distribution, en vue de son exploitation, de l'information utile aux acteurs économiques.*" (Martre et al., 1994). Il

ressort de ces définitions que l'information est au cœur du dispositif d'intelligence économique, elle est "*la différence qui engendre des différences*" (Huyghe, 2010). Il faut donc être en mesure de repérer celle qui est pertinente par la mise en place d'un dispositif de veille et de l'analyser. L'intelligence économique regroupe plusieurs métiers avec cette même préoccupation, faire remonter et exploiter de l'information de qualité pour les décideurs (Anonyme, 2019). Les veilleurs et les analystes sont la cheville ouvrière de ce dispositif. Les premiers collectent l'information potentiellement utile à leur organisation et sont chargés de son primo-traitement : ils doivent à la fois effectuer un tri parmi les nouveaux éléments informationnels détectés, les diffuser aux experts concernés et s'assurer de leur authenticité ou, a minima, de leur adéquation avec les besoins internes. Les analystes, souvent nommés "experts" dans les organisations, sont en charge d'exploiter l'information collectée. De fait, hormis le cas spécifique des services de renseignement dont est issue cette distinction, le rôle de veilleurs et analystes dans les organisations publiques et privées est moins délimité, les premiers ayant régulièrement à donner leur avis sur l'information collectée et à l'enrichir de leurs connaissances, même s'ils n'en sont pas "experts", les seconds à mener des recherches ponctuelles d'informations et à faire de la veille. Ces deux fonctions se trouvent donc en première ligne face à d'éventuelles attaques informationnelles. En effet, si les organisations publiques ou privées tirent parti de l'information elles peuvent également en être victimes. Il peut s'agir de mésinformation, c'est à dire d'informations considérées comme véridiques par l'émetteur mais, en réalité, fausse ou erronées, de désinformation globale qui ne leur est pas directement adressée mais peut les amener à une mauvaise appréhension de leur environnement ou, bien entendu, d'actions de désinformation ciblées visant à perturber leur fonctionnement en les intoxiquant, en les discréditant ou en les diffamant. François-Bernard Huyghe explique que « *Si chacun peut devenir émetteur à son tour et non simple récepteur des mass médias(…) il peut informer donc désinformer. »* Il évoque corollairement la nécessité pour les réseaux sociaux de capter l'attention des « cibles », ce qui passe par la possibilité de pouvoir modifier les contenus numériques *« à très faible coût, avec des exigences de plus en plus faible en termes de compétence techniques (logiciels plus simples et accessibles)* » et précise que « *les ressources documentaires, banques d'images, bases d'information en ligne, immédiatement, gratuitement… permettent de piocher dans des réserves de données qui permettent de forger des trucages vraisemblables. Le travail du faussaire est donc facilité pour ne pas dire banalisé.* » Et de définir trois conditions permettant d'établir l'action de désinformation et de la distinguer d'une simple rumeur :

- « *l'intention stratégique qui se traduit par une fabrication (faux documents, fausses scènes, faux témoignages, fausses images) ;*
- *que cette intention soit médiatisée, c'est-à-dire relayée par des médias ou par des groupes humains (associations, communautés...) qui amplifient le message, l'authentifient, en dissimulent la source partisane ou intéressée ;*
- *que le processus serve aux intérêts de son initiateur au détriment de la cible. (...) il faut d'abord provoquer un effet de croyance en un danger imaginaire, en un crime supposé, en une conspiration, en une manœuvre occulte... »* (F-B Huyghe, 2016)

L'émergence d'une nouvelle génération de technologies de falsification

Cette fabrication du faux, ou falsification, est ici mise au service d'actions de désinformation, un terme qui apparaît initialement dans un dictionnaire soviétique de 1953[1]. Si le terme latin de *falsificatio* n'est attesté lui que par le latin médiéval, la pratique est évidemment plus ancienne. L'exemple le plus célèbre étant celui de la donation de Constantin 1er (Scheid, 2016) par lequel cet empereur romain du 3ème siècle attribue l' « imperium », ou commandement, de l'Occident au pape Sylvestre[2].

Si depuis une vingtaine d'années déjà, les réseaux sociaux permettent de relayer et d'amplifier un message, la fabrication de faux connait pour sa part un véritable essor dû à l'intelligence artificielle et à sa composante d'apprentissage profond (*deep learning*). Les premiers hypertrucages (*deepfakes*) sont apparus à l'automne 2017. Il s'agissait de vidéos pornographiques dans lesquelles les visages d'actrices connues remplaçaient ceux des actrices originales (Cole, 2017). L'expression "*deepfake* " a été forgée en ajoutant au terme "*fake*" le "*deep*" de "*deep learning*" qui désigne un ensemble de techniques reposant sur le modèle des réseaux de neurones dans le but d'apprendre à des machines à reconnaître et reproduire des formes, structures, objets, visages, etc en tirant parti de bases existantes. C'est donc une sous-branche du *machine learning* qui n'avait évidemment pas pour vocation initiale la falsification mais qui, du fait de cet usage possible, a transformé le faussaire en un faussaire augmenté. Techniquement, deux innovations ont permis l'émergence des *deepfakes*, le système de reconnaissance facial développé par Yann LeCun pour Facebook, baptisé

[1] Il sera ajouté au dictionnaire de l'Académie française en 1980.

[2] Il est intéressant de noter que le droit romain, via la loi des XII tables (5ème siècle avant J-C), prévoyait qu'une atteinte portée par voie de falsification, de destruction ou de substitution au patrimoine public ou privé, mettait en cause la cité, était un délit grave.

Deepface (LeCun et al., 2015), et les GAN[3] développés par Ian Goodfellow (Goodfellow et al., 2014). Un GAN repose sur la mise en compétition de deux réseaux neuronaux, le générateur et le discriminateur, le premier produisant des faux de plus en plus crédibles a mesure que le second les détecte, le "dialogue" étant entretenu jusqu'à obtenir des faux plausibles pour l'œil humain. Étendant leur champ d'application, les GAN sont maintenant utilisés pour générer toutes sortes de reproductions du réel :

- Des vidéos substituant les visages de personnes connues ou non à d'autres. Le corps "porteur" pouvant être celui d'un acteur ou être extrait d'une vidéo existante (*deepfake*).
- De faux visages pouvant être utilisés, par exemple, pour illustrer des profils sur les réseaux sociaux[4] ;
- Des représentations de corps humains synthétiques complets et "animables", conçus dans le but de jouer le rôle de personnes dans des simulations. On parle de « marionnettisme » (*puppeteering*)[5].
- des voix que l'on clonera à partir d'enregistrements de volontaires ou de personnes "cibles" (*voicecloning*)[6] ;
- des images satellitaires conçus initialement en tant que sets de données permettant à une intelligence artificielle de détecter des objets ou anomalies mais pouvant servir à tromper un adversaire (Sun et al., 2021; Zhao et al., 2021)[7] ;
- des images médicales également conçues comme sets de données d'entraînement à la détection de tumeurs ou autres mais susceptibles d'être détournées de cet usage (Skandarani et al., 2021) ;
- des jeux de données divers (*synthetic datas*) comme ceux que la startup Mostly.ai fournit à des sociétés financières, d'assurance, etc. Répartis différemment des originaux mais sur le même volume d'affaire, ils leurs permettent, par exemple, de partager leurs bases de

[3] *Generative Adversarial Network*. En français on parle de Réseau antagoniste génératif
[4] On pourra tester par exemple les services Thispersondoesnotexist.com (https://thispersondoesnotexist.com/) et Generated.photos (https://generated.photos/face-generator/). Accédé le 25/03/2022.
[5] Voir par exemple cette vidéo de la société Ariel.ai : https://www.youtube.com/watch?v=aQ4shIsQabo . Accédé le 25/03/2022.
[6] Cf. cette vidéo de la société Synthesia.ai : https://www.youtube.com/watch?v=CF_e0kMCW2o . Accédé le 25/03/2022.
[7] On peut tester le service Thiscitydoesnotexist : http://thiscitydoesnotexist.com/. Accédé le 25/03/2022.

données clients avec des fournisseurs extérieurs en restant en conformité avec les lois relatives au respect de la vie privée[8].

La génération automatique de texte au cœur de la production de faux

Si elle n'est pas la plus spectaculaire, la génération automatique de texte par l'intelligence artificielle est la plus susceptible d'être mise en œuvre dans la production de faux en masse et peut, comme nous le verrons, en devenir le moteur. En effet, avec l'avènement de nouveaux modèles de traitement du langage reposant sur des réseaux de neurones spécifiques, les possibilités de générer du texte automatiquement se sont démultipliées. Le modèle de langage le plus prometteur est GPT-3[9], de la société OpenAI[10]. Il est propulsé par 175 milliards de paramètres[11], soit de valeurs qu'un réseau de neurones tente d'optimiser lors de son entraînement[12]. Le modèle est conçu pour générer du texte à l'aide d'algorithmes pré-entraînés sur des corpus de référence collectés sur le Web[13]. Cet entraînement lui permet, via une analyse sémantique, de « comprendre » la mécanique d'une langue. Une fois passée cette étape, lorsqu'on fournit un extrait de texte au système, par exemple une phrase d'introduction, celui-ci tente de le compléter en prédisant les mots qui pourraient faire sens pour l'utilisateur[14]. Les possibilités sont alors infinies et limitées par la seule imagination. Il pourra s'agir d'écrire des textes « à la manière de », de créer des robots conversationnels (*chat bots*), de répondre à une question d'ordre médicale, de créer des interfaces utilisateurs à partir d'une simple description textuelle,…[15]. Mais si la génération de textes synthétiques constitue la pierre angulaire d'un déploiement à grande échelle des technologies du faux dans le cadre d'actions de désinformation c'est avant tout parce que ces modèles, notamment GPT-3, ont aussi la capacité d'écrire des programmes informatiques sur simple injonction vocale ou textuelle. La fonctionnalité a été lancée en avril 2021 avec des résultats remarquables[16]. L'impression laissée par la vidéo de démonstration est celle d'une magie en acte, d'une parole qui devient créatrice par l'intermédiaire

[8] Voir cette présentation des sets de données synthétiques par la société Mostly.ai (https://www.youtube.com/watch?v=NjpkeiUcc5c). Accédé le 25/03/2022.
[9] Pour Generative Pre-Trained Transformer 3
[10] Cette entreprise à but lucratif plafonné a notamment été fondée par Elon Musk en 2015. https://openai.com/. Accédé le 25/03/2022.
[11] Le modèle précédent, GPT-2 n'en utilisait "que" 1,5 milliard. GPT-4 est annoncé pour le début de l'année 2023. Il devrait comporter 100 000 milliards de paramètres, soit plus de 500 fois la taille de GPT-3.
[12] Il en existe d'autres comme BERT (Google), MT-NLG (NVIDIA et Microsoft), Plato-XL (Baidu),
[13] Par exemple la totalité des contenus de la Wikipedia
[14] Il est possible de tester une instance de GPT-2 sur le site Write with transformer : https://transformer.huggingface.co/doc/distil-gpt2. Accédé le 25/03/2022.
[15] Ces exemples sont tirés du site GPT3demo qui en propose plus de 300. https://gpt3demo.com/
[16] Présentation de l'outil de génération de code via IA d'OpenAI (https://www.youtube.com/watch?v=SGUCcjHTmGY). Accédé le 25/03/2022.

d'une IA invisible ou qui le deviendra bientôt. Ou quand le logiciel laisse la place au Logos (Malik, 2022)…

Bien entendu les programmes ainsi créés sont loin d'être exempts de défauts et des chercheurs ont constaté que pour certaines tâches où la sécurité est cruciale le code contenait des failles de sécurité environ 40% du temps (Knight, 2021). Mais l'avantage du *deep learning* est sa capacité à s'améliorer par itérations constantes. De tels outils annoncent donc à la fois un changement dans la manière dont les développeurs écriront le code, mais également la possibilité pour des non-informaticiens de développer des programmes simples. Microsoft propose déjà via PowerAutomate (Cunningham, 2021), son outil *nocode*[17] disponible dans la suite pour entreprise Microsoft 365, la possibilité d'utiliser une version bridée de GPT-3 et plusieurs startups ont lancé récemment des « briques » de traitement par IA s'appuyant sur des interfaces visuelles qui permettent de construire des modèles d'apprentissage automatique[18]. Ces technologies de falsification reposent donc sur une capacité croissante à entraîner des algorithmes de *deep learning* à partir de jeux de données, réels ou virtuels, c'est à dire à créer du faux vraisemblable à partir du réel ou de faux crédible.

Les risques induits par les technologies de falsification

Pour les professionnels de l'intelligence économique, le risque s'accroît en conséquence. En effet, les technologies de falsification outilleront de plus en plus les campagnes d'attaques informationnelles et leur démocratisation en marche en font déjà une menace potentielle crédible pour les organisations. Par ailleurs, le contexte global dans lequel elles se développent étend leur capacité de nuisance. En effet, le développement inédit des technologies de l'information qui accompagnent et nourrissent la mondialisation, ainsi que les prétextes d'attaques informationnelles fournis par les thèmes désormais centraux des risques environnementaux ou éthiques, du principe de précaution ou des questions de genre, ont initié des risques nouveaux pour les organisations. Publiques ou privées, elles présenteront inévitablement un ou plusieurs « angles morts » sur ces sujets, qui, une fois détectés et documentés, seront autant de leviers à de potentielles actions de déstabilisation.

Il existe peu de travaux sur le coût de la désinformation pour les entreprises mais une étude menée en 2019 par l'économiste Roberto Cavazos estimait celui de la désinformation financière à 17

[17] Le cabinet Gartner estime que d'ici 2024, le développement d'applications *Low code/No code* sera responsable de plus de 65 % de l'activité de développement d'applications informatiques. Bedi (2021).
[18] C'est le cas par exemple de C3AI avec Ex Machina, Dataiku avec DataRobot ou encore Databricks d'AutoML

milliards de dollars par an et celui des attaques à la réputation à 9,5 milliards (Cavazos, 2020). Chiffres qu'il faut comparer aux faibles investissements nécessaires pour mener à bien des actions de désinformation. Ainsi l'IRA (Internet Research Agency), cette « ferme à trolls » étatique russe très active durant les élections américaines de 2016, dont le fonctionnement est maintenant bien documenté (Nimmo et al., 2020), proposait en 2013 cette offre d'emploi : « *Opérateurs Internet recherchés ! Travaillez dans un bureau luxueux à Olgino. Le salaire est de 25960 roubles par mois*[19]*. La tâche : placer des commentaires sur des sites Internet spécifiques, rédiger des billets thématiques, des blogs sur les médias sociaux....NOURRITURE GRATUITE.* » (cité dans Buchanan et al., 2021). Un rapport de la société Trend Micro de 2017 indiquait pour sa part que créer un compte Twitter de 300000 followers en un mois coûtait en moyenne 2600 dollars (Gu et al.). Et ces coûts peuvent être encore plus faibles. Ainsi au Venezuela, des documents divulgués en 2018 décrivent comment des brigades de désinformation recrutent des personnes qui « *s'inscrivaient à des comptes Twitter et Instagram dans des kiosques approuvés par le gouvernement* » et recevaient des coupons pour de la nourriture et des marchandises et d'autres avantages gouvernementaux (Bradshaw et Howard, 2017).

Risques potentiels et risques avérés

Dans une note de mars 2021 le FBI indiquait qu': « *Il est presque certain que les acteurs malveillants tireront parti du contenu synthétique (y compris les deepfakes) pour des opérations de cyberinfluence et d'influence étrangère dans les 12 à 18 prochains mois* » (Anonyme, 2021) . Un panorama global des risques liés au mauvais usages de l'intelligence artificielle a par ailleurs été dressé en 2020 par une équipe de chercheurs qui a interrogé des spécialistes du domaine afin de recueillir leurs avis (Caldwell et al., 2020).

[19] Soit environ 600 euros au taux moyen de 2013.

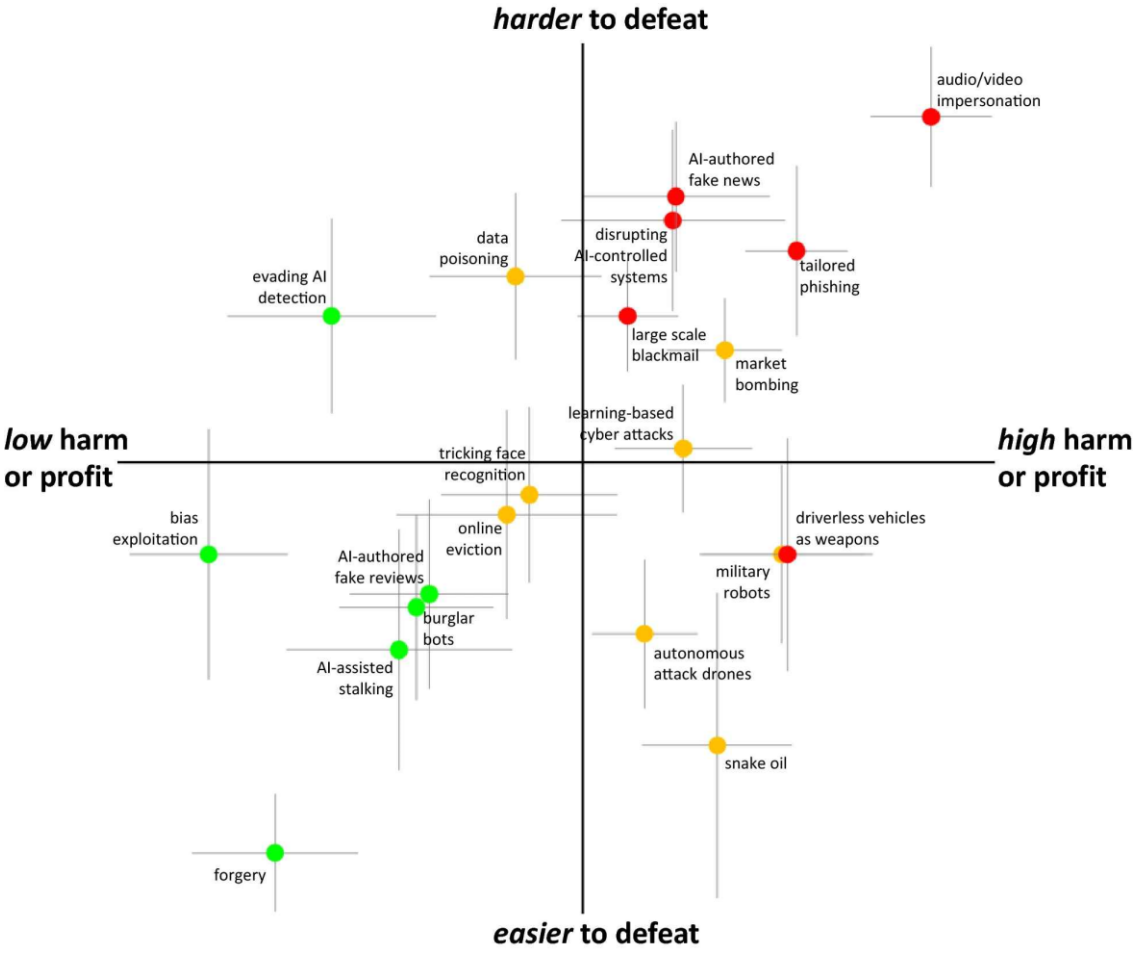

Fig. 1 Tableau récapitulatif des risques liés à l'intelligence artificielle[20]

Il ressort de leur classement que l'usurpation d'identité par audio/vidéo (*deepfake*, *voicecloning*) est le type de crime considéré comme le plus préoccupant de tous, sa détection étant jugée difficile et les bénéfices qu'elle entraîne (profits ou préjudices) élevés. Non loin, viennent les actions de *phishing* personnalisés via intelligence artificielle et les *fakes news* générées par des modèles de langage de type GPT-3.

[20] Tous les risques présentés dans ce graphique ne sont pas imputables au seul *deep learning*.

Nous avons synthétisé ci-dessous les types d'attaques informationnelles potentielles dans lesquelles les technologies de falsification pourraient être utilisées contre les organisations :

Types de faux contenus utilisés	**Types d'actions menées (indicatif)**	**Opérateurs potentiels de désinformation**	**Vecteurs de diffusion**
Image (faux visages) Vidéo (deep fakes) Audio (voix synthétique, *voice cloning*) Texte (texte et code informatique)	Dénigrement de produits et services Attaque sur l'éthique de l'organisation Attaque sur la politique environnementale et/ou le greenwashing Attaque sur les questions de genre et de diversité Déstabilisation de la communauté de clients Chantage à l'image sur les collaborateurs Usurpation d'identité (PDG et membres de la direction) Création de faux souvenirs en ligne pour activation ultérieure	Concurrents (via une société spécialisée) Société spécialisée dans les campagnes d'influence (RP, publicité, marketing,…) Groupes mafieux Hackers Services de renseignement Groupes d'activistes Clients mécontents	Réseaux sociaux Applications de messagerie Plateformes de vidéos Plateformes de jeux Presse traditionnelle Site ou blogs créés pour servir de caisses de résonnance Forums de discussion Sites de presse à *clickbai*t et actualités automatisées

De nombreuses campagnes de désinformation et de déstabilisation sont détectées régulièrement. Ainsi, l'équipe *Computational Propaganda Research Project* de l'Université d'Oxford indique que 81 pays utilisent les médias sociaux dans un but de propagande informatique (Bradshaw et Howard, 2017). De leur côté, les équipes de Facebook annonçaient avoir démantelé entre 2017 et mi-2021 plus de 150 opérations d'influence violant leur politique contre les « comportements

inauthentiques coordonnés » (CIB)[21] en provenance de plus de cinquante pays et visant le débat public national ou d'autres pays (Gleicher et al., 2021).

Faux visages

L'usage de technologies du faux est avéré dans plusieurs de ces campagnes et plus spécifiquement l'utilisation de faux visages. Ainsi en juin 2019, Katie Jones, une personne non-réelle dotée d'un avatar généré par GAN, a été détectée sur Linkedin. Elle était liée à de nombreux universitaires et personnalités clés de la communauté de la sécurité nationale américaine dans ce qui peut être perçu comme une éventuelle opération d'espionnage (Satter, 2019). En décembre 2019, Facebook a démantelé un réseau de plusieurs centaines de comptes utilisant des avatars générés par GAN lié au média conservateur Epoch Media détenu par la secte Falun Gong et dont l'objectif était le soutien à Donald Trump et le dénigrement du gouvernement chinois (Alba, 2019). A l'inverse, en août 2020, l'opération Spamouflage Dragon a déployé un réseau d'avatars aux profils générés par GAN pour soutenir le Parti communiste chinois sur Twitter et YouTube, et a ciblé le public américain avec des messages critiquant la réponse américaine à la pandémie de COVID-19 ainsi que ses politiques envers la Chine (Sedova et al., 2021). Si pendant plusieurs années le risque engendré par les faux-visages était considéré comme limité car ceux-ci comportaient des anomalies les rendant aisément détectables, une récente étude a montré que non seulement ce n'était plus le cas mais que les personnes interrogées leur faisaient plus facilement confiance qu'aux visages réels (Nightingale et Farid, 2022).

Voix clonées

Hormis les images, des usages frauduleux de voix clonées ont également été détectés. Ainsi en 2019 le directeur général d'une entreprise britannique du secteur de l'énergie, croyant que son PDG était au téléphone, a exécuté son ordre de virer plus de 240 000 dollars sur un compte en Hongrie. Il explique que s'il a trouvé la demande étrange, il n'a pas douté un instant de l'authenticité de la voix de son patron (Stupp, 2019). Dans un autre cas rendu public, une succursale d'une banque des Etats Arabes Unis a transféré une somme de 35 millions de dollars sur un faux compte suite à un appel frauduleux utilisant la voix cloné d'un des directeurs de la compagnie (Brewster, 2021).

[21] Facebook définit les CIB comme : « *tout réseau coordonné de comptes, de pages et de groupes (…) qui s'appuie de manière centralisée sur de faux comptes pour tromper Facebook et les personnes utilisant nos services sur l'identité de ceux qui sont derrière l'opération et sur ce qu'ils font* ».

Deepfakes

Hormis les cas, déjà nombreux, dans lesquels un maître chanteur utilise les *deepfake* pour incruster le visage de la personne ciblée dans des vidéos pornographiques (Burgess, 2021), cette technologie a notamment été utilisée en 2020 pour tenter de déstabiliser la position de la France au Cameroun via la diffusion d'un discours offensant pour les camerounais de l'avatar de l'Ambassadeur de France Christophe Guilhou (Anonyme, 2020). Si la tentative a rapidement été déjouée, il est facile d'imaginer les conséquences potentielles de telles actions sur les expatriés français en cas de vérification trop lente ou de multiplication de fausses vidéos du même type. Plus récemment c'est un *deepfake* du président ukrainien Volodymyr Zelensky capitulant qui a été diffusé dans le but de démobiliser ses troupes et son peuple (Capron, 2022). A l'aune de ses exemples, il est aisé d'anticiper les réactions d'investisseurs et de clients à une fausse vidéo qui mettrait en scène les doutes d'un PDG sur la qualité d'un produit avant son lancement...

Enfin, il existe un risque plus global non évoqué dans le graphique ci-dessus, c'est la possibilité qu'à long terme les *deepfakes* affaiblissent progressivement l'idée même de l'historicité d'un évènement. Finalement ce que l'on croit avoir vu a-t-il réellement existé ? A l'inverse, des vidéos pourraient implanter de faux souvenirs dans les mémoire, même après qu'elles aient été repérées, la théorie du double-processus validant le fait que nous nous souvenions d'une vidéo ou d'un article, mais pas nécessairement de leur démystification (Pennycook et Rand, 2021).

Le texte synthétique et la propagande computationnelle

Si les cas d'usage de *deepfakes* lors d'actions de désinformation sont pour l'instant peu nombreux, c'est probablement parce qu'ils sont encore assez faciles à détecter et pourraient donc compromettre les attaques informationnelles. Si leur côté spectaculaire a vite donné lieu à la recherche de contre-mesures (Hatmaker, 2018)(Luebke, 2021), il n'en va pas de même pour le texte synthétique déjà présent sur le web et bien plus inquiétant en terme d'usage et d'impact. Qu'il s'agisse de publications journalistiques semi-automatisées (Danzon-Chambaud, 2022), de génération de *fake news* (Vandeginste, 2021) ou de publications automatisées sur les médias sociaux via *chat bots* dans le cadre de campagnes de désinformation (Shao et al., 2018), le texte synthétique est beaucoup plus difficile à repérer que les vidéos et les images. Ainsi a t-on vu les participants d'une étude récente incapables de détecter de manière fiable la poésie générée par un algorithme de celle de poètes réels qu'elle était sensée copier (Köbis et Mossink, 2021). Dans une autre étude, 72 % des répondants jugeaient les articles générés par le modèle GPT2 crédibles et

pratiquement impossibles à distinguer de ceux rédigés par des journalistes, en particulier lorsque l'angle partisan du contenu augmentait (Kreps et al., 2020). Plus inquiétant, une étude menée en 2021 indiquait que des rapports générés avec GPT-2 sur la cybersécurité avaient réussi à tromper des experts de ce secteur (Joshi et al., 2021).

La connaissance détaillée de certaines campagnes de désinformation menées ces dernières années, comme celle de l'IRA, permet de comprendre que l'ensemble du travail d'écriture et de publication de contenus était réalisé jusqu'à maintenant par des opérateurs humains. Avec les possibilités grandissantes des modèles de langage, ce volet sera de plus en plus dévolu à l'IA, l'humain étant seulement là pour l'orienter et le piloter. Et cette possibilité n'est déjà plus une fiction : les chercheurs de la société FireEye ont par exemple entraîné GPT-2 sur plusieurs millions de tweets et de publications diffusées sur le site Reddit d'opérateurs de l'IRA (Sajidur et al., 2019). A l'issue du traitement le système était capable de produire des tweets crédibles sur les mêmes thématiques[22].

Présentation d'un modèle global de désinformation par les technologies de falsification

Les possibilités d'industrialisation des campagnes de désinformation en ligne sont donc parfaitement envisageables à très courts terme[23] et, grâce aux méthodes d'apprentissage par "*transfer learning*"[24] qui limitent les besoins en entraînement et en puissance machine, accessibles à tout acteur doté des bonnes compétences en interne. Acteurs qui existent déjà puisque, outre l'IRA déjà citée, l'équipe d'Oxford indique dans son rapport 2020 avoir identifié 65 firmes (agences de relations publiques, de communication, de marketing, …) actives dans quarante-huit pays. Ces entreprises offrent une variété de services allant de la création de personnages inauthentiques (*sockpuppet*) au microciblage en passant par l'amplification de messages via faux likes et followers (Bradshaw et al., 2021). Elles permettent aux opérateurs de désinformation d'externaliser leurs actions et d'en rendre l'attribution difficile. C'est l'émergence d'un marché de l'influence « *as a service* » , que des *chat bots* nourris aux contenus élaborés par des modèles de langage de plus en

[22] Une grande partie de ce travail a été réalisée par un seul étudiant de l'Université de Floride, au cours d'un stage de trois mois. (https://www.wired.com/story/to-see-the-future-of-disinformation-you-build-robo-trolls). Accédé le 25/03/2022.
[23] Et probablement déjà actives.
[24] Dans l'apprentissage par transfert, l'on part d'un modèle générique qui a été pré-entraîné pour une tâche initiale où de nombreuses données sont disponibles et l'on tire ensuite parti des connaissances acquises par le modèle pour l'entraîner sur un ensemble de données différent et plus limité.

plus performants, viendront appuyer sans limites.

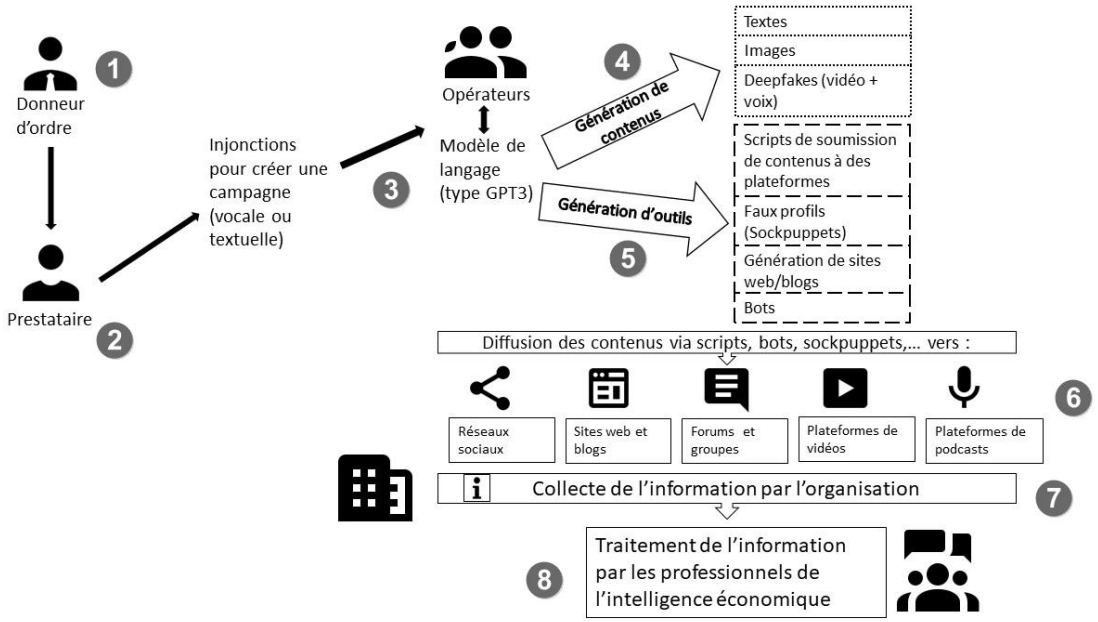

Fig. 2 Modèle de désinformation assisté par les technologies du faux

Nous proposons ci-dessus un modèle d'industrialisation de campagnes de désinformation assistée par intelligence artificielle dans lequel :

1. Le client définit sa demande.
2. Le prestataire (agence de relations-presse, de communication, hacker,…) reçoit la demande.
3. Il précise les axes de la campagne (textuellement ou vocalement) et demande au modèle de langage, seul ou en mode d'action hybride avec des opérateurs humains de :
4. générer les différents types de faux contenus pertinents.
5. générer les outils informatiques et vecteurs qui permettront leur diffusion.
6. Les contenus sont délivrés automatiquement, manuellement ou en mode hybride par les outils vers différents types de médias et supports, certains originaux (réseaux sociaux, plateformes de diffusion de vidéos) et d'autres créés pour l'occasion (sites web, blogs, groupes sur les réseaux sociaux,…).
7. Les contenus sont collectés par le dispositif de veille de l'organisation cible.
8. Les professionnels de l'intelligence économique traitent l'information seuls ou aidés d'outils de traitement spécifiques. Ils en vérifient l'authenticité et préviennent les acteurs concernés de l'organisation (Direction générale, Direction informatique, cellule de crise,…) en cas d'attaque informationnelle.

Notons que l'impact des modèles de langage du type de GPT-3 peut s'avérer décisif à plusieurs niveaux. Il permettra par exemple aux adversaires de tester de nombreux types de messages et de variantes de ces messages en fonction des populations ciblées[25]. En capitalisant les mesures des retombées de ces actions il sera possible d'identifier les versions ayant trouvé le meilleur écho et d'orienter ainsi les futures campagnes. Par conséquent, la puissance des modèles linguistiques ne réside pas seulement dans l'échelle à laquelle ils permettent d'amplifier une campagne mais aussi dans la façon dont, combinés à des mécanismes efficaces d'évaluation et de raffinement, ainsi qu'à des dispositifs d'analyse psychographique (Schiavon, 2019) et de sentiment, ils augmenteront l'efficacité des futures campagnes.

Prévention et détection des technologie de falsification

Les possibilités de détection existent et donnent lieu à une recherche abondante qui suit globalement quatre axes :

	Mise en œuvre	Exemple
Détection technique	Détection d'éléments spécifiques au support utilisé (métadonnées, trame de l'image, certificat de hachage...)	Microsoft Video Authentificator
Détection de l'infrastructure de communication	Analyser l'infrastructure utilisée pour propager les messages de la campagne (faux comptes, n° de téléphone utilisés pour créer les faux comptes, cartes de crédit, ...).	Algorithme de détection de comportements trompeurs CIB croisant plus de 150 critères. Par l'équipe sécurité de Facebook.(Gleicher et al., 2021)
Détection par IA	Utiliser le *machine learning* pour détecter des anomalies dans des vidéos, images, textes,...	GAN-Scanner atteint une précision de 95 % dans l'identification des images synthétiques(Luebke, 2021) GLTR : aide à détecter si un texte a été généré à l'aide du *machine learning* (GPT-2, GPT-3, ...)[26]
Détection par l'humain assisté par l'IA	Mêler esprit critique, *factchecking* et "outil" d'aide à la détection assistée par IA	Solutions commerciales : Primer.ai, Sensity.ai, Deepnews.ai, Quantum Integrity,…[27]

Il existe également des mesures axées sur la prévention où les États prennent l'initiative au moyen de la législation, comme c'est déjà le cas aux États-Unis (Clarke, 2019), en Chine (Ye, 2022) ou en Corée du Sud (So-Yeon, 2022). Les plateformes sociales jouent bien sûr un rôle essentiel dans la

[25] Sur le principe de l'A/B testing qui permet de comparer deux version d'un même contenu sur une population cible afin de déterminer celle qui convertit le mieux.
[26] GLTR peut être testé à cette adresse : http://gltr.io/dist/index.html (accédé le 25/03/2022)
[27] Un panorama plus complet de ces entreprises et disponible ici : https://impact.dealroom.co/lists/20937 (accédé le 25/03/2022)

lutte contre les fausses informations. Cependant, si elles sont généralement proactives sur la question, elles font face à certaines limites liées à l'évolution protéiforme des campagnes de désinformation. En effet, du fait de leurs capacités de détection améliorées, les opérateurs ont le plus souvent tendance à les contourner et à lancer leurs campagnes sur des réseaux sociaux secondaires où le degré de transparence est très variable. L'optimisation des mécanismes de collaboration entre plateformes permettrait d'améliorer l'identification d'actions coordonnées, mais certaines d'entre elles ont été créées spécifiquement à des fins d'activisme (par exemple Gab ou Gettr) et n'ont donc aucune raison de participer à ces échanges.

Par ailleurs, la détection technique et la prévention, quoi qu'indispensables, ne sont pas suffisantes. Ce serait oublier en effet que la désinformation ne croît que sur des esprits prêts à la recevoir, et qu'elle est d'autant plus efficace qu'elle conforte des opinions préexistantes et alimente des divisions établies. Les individus peuvent par exemple être dupés par de fausses vidéos parce qu'elles confirment leurs biais, mais aussi parce qu'ils ne savent pas qu'elles peuvent être manipulées. Cela implique donc d'améliorer l'éducation aux médias, mais l'effort principal doit porter sur le développement de l'esprit critique, seul susceptible de permettre une prise de recul efficiente et durable. Bien que des approches de bon sens soient utiles[28], elles ne peuvent remplacer le développement d'une pensée critique acquise au cours de l'enseignement primaire et secondaire. Les sciences et les humanités, en particulier l'enseignement de la philosophie, doivent fournir le substrat nécessaires à l'application de cet esprit critique.

Conséquences pour les professionnels de l'intelligence économique

Pour les professionnels de l'intelligence économique les solutions ne peuvent donc être simplistes. La désinformation peut être utilisée contre leurs organisations selon différentes modalités. Les campagnes peuvent cibler la réputation globale de l'organisation (produits, service, valeurs,…), mais aussi leurs employés à titre individuel, ou encore leurs clients et fournisseurs. Elles peuvent être menées par des acteurs humains ou des bots et, pour les plus sophistiquées, lors d'attaques hybrides intégrant les deux composantes. Les attaques automatisées menées par les bots ont notamment pour objectif de saturer l'espace informationnel dans le but de contraindre les professionnels de l'information à un travail de vérification important, les détournant ainsi de tâches

[28] Voir par exemple les méthodes SIFT (https://libguides.colorado.edu/c.php?g=645411&p=7347477 – Accédé le 26/03/2022) ou SHEEP (https://firstdraftnews.org/articles/think-sheep-before-you-share-to-avoid-getting-tricked-by-online-misinformation/. Accédé le 26/03/2022)

d'analyse à plus forte valeur ajoutée. En utilisant, la boucle OODA comme cadre d'analyse (Moinet, 2019), l'on comprend qu'il s'agit pour l'attaquant de limiter la capacité d'action de l'adversaire en le saturant d'information fausses ou dépassées qui l'amèneront à perdre du temps et à retarder d'autant les prises de décision nécessaires. Quoique déjà au fait des risques de désinformation susceptibles de toucher leurs organisations, les praticiens de l'intelligence économique, doivent prendre conscience de la mutation de la menace liée aux nouvelles technologies de falsification, qui entraîne de fait des évolutions dans leurs métiers. Les solutions passeront donc par :

- Leur capacité à anticiper les risques potentiels de désinformation corrélatifs à leurs sujets d'intérêt (axes de veille) mais aussi, plus globalement, aux sujets sensibles relatifs à leur domaine d'activité et à l'organisation qui les emploie. Cela passera notamment par un travail de cartographie de ces risques.
- Le suivi de l'évolution des plateformes de veille et des outils informatiques susceptibles de les aider dans cette tâche de détection en automatisant ce qui peut l'être (veille métier) et la formation à ces solutions. En corollaire, cela implique que les éditeurs de plateformes de veille intègrent au plus tôt des fonctionnalités d'aide à la détection, qui viendront assister les professionnels de l'IE.
- Le développement de leur sensibilité à la possibilité d'attaques informationnelles utilisant les technologies de falsification et la nécessité de faire de la distance critique une seconde nature.
- La maîtrise des méthodes d'OSINT[29] qui aident à l'investigation et à la vérification de faits (*factchecking*)
- La capacité à bien communiquer car ils auront de plus en plus souvent à convaincre des clients internes déjà « intoxiqués » par les actions de désinformation adverses.

Conclusion

Les technologies de falsification impactent directement les professionnels de l'intelligence économique qui doivent à la fois faire évoluer leur degré de prise de conscience du risque informationnel et leurs capacités à détecter et vérifier l'information. Plus généralement, dans un contexte où l'idée même de « preuves par l'image » s'estompe, elles pourraient faire émerger une société de la défiance dans laquelle les individus partiront du principe qu'une information est fausse

[29] L'OSINT (Open Source Intelligence) regroupe un ensemble de méthodes et outils permettant d'exploiter l'information accessible en ligne.

tant que l'inverse n'est pas prouvé, entretenant et accroissant ainsi l'impression de « sables mouvants » informationnelles que les réseaux sociaux et la surinformation avaient déjà initié. Car si tout peut être truqué il sera encore plus aisé de ne retenir que ce qui conforte les avis et préjugés de chacun en ne tenant pas compte de faits importuns que l'on imputera aux technologies de falsification. L'intérêt porté à l'actualité et le niveau de confiance dans les médias, qui baissent d'année en année en France, pourraient ne pas s'en remettre (Carasco, 2022). Plus largement, c'est le fondement des démocraties qui peut être ainsi miné puisque n'importe quel homme politique ou citoyen peut nier les preuves d'une action douteuse ou illicite en arguant d'une manipulation à son encontre. C'est l'idée de « dividende du menteur » (Chesney et Citron, 2018) qui exprime le fait que celui-ci n'a aucun effort à faire pour que son déni soit plausible alors qu'il en faudra beaucoup pour prouver ses mensonges[30]. Enfin, la dimension « temps réel » des médias, réseaux sociaux et autres « caisses de résonance » donne également la prime à la propagation de la mauvaise information plutôt qu'à celle de son démenti qui impliquera nécessairement le temps de la vérification, ce qui paradoxalement pourrait renforcer le sentiment de défiance. Les technologies de falsification sont donc un problème pour les professionnels de l'intelligence économique, en première ligne face aux attaques informationnelles, mais aussi un problème plus global pour les démocraties qui doivent prendre conscience de la polarisation de la société qu'elles sont à même d'exploiter et de renforcer.

Références

Alba, D. (2019, 20 décembre). Facebook Discovers Fakes That Show Evolution of Disinformation. *The New York Times*. https://www.nytimes.com/2019/12/20/business/facebook-ai-generated-profiles.html

Anonyme. (2019). *Cartographie des métiers de l'intelligence économique*. https://www.ege.fr/sites/ege.fr/files/downloads/metiersIE_AEGE2019.pdf

Anonyme (2020, 29 juin). Attention, cette vidéo de l'ambassadeur français au Cameroun est un "deepfake". *Les Observateurs - France 24*. https://observers.france24.com/fr/20200629-attention-video-ambassadeur-francais-cameroun-guilhou-biya-deepfake

Anonyme. (2021). *Malicious Actors Almost Certainly Will Leverage Synthetic Content for Cyber and Foreign Influence Operations.* FBI. https://www.ic3.gov/Media/News/2021/210310-2.pdf

Bedi, C. (2021, 13 juillet). Citizen Developers and the Democratization of Code. *CIO*. https://www.cio.com/article/188948/citizen-developers-and-the-democratization-of-code.html

Bradshaw, S., Bailey, H. et Howard, P. N. (2021). *Industrialized Disinformation : 2020 Global Inventory of Organized Social Media Manipulation.* Oxford, UK : Programme on Democracy & Technology. https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/

Bradshaw, S. et Howard, P. N. (2017). *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation.* Computational Propaganda Research Project.

[30] L'asymétrie bénéficie au menteur, ainsi que le contexte de défiance, car plus nous sommes sceptiques, plus nous sommes perméables à ses allégations, considérant qu'il faut prendre le temps de les vérifier.

https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf

Brewster, T. (2021, 14 octobre). Fraudsters Cloned Company Director's Voice In $35 Million Bank Heist, Police Find. *Forbes*. https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=6601fd7c7559

Buchanan, B., Lohn, A., Musser, M. et Sedova, K. (2021). *Truth, Lies, and Automation: How Language Models Could Change Disinformation.* Center for Security and Emerging Technology. https://doi.org/10.51593/2021CA003

Burgess, M. (2021, 15 décembre). The Biggest Deepfake Abuse Site Is Growing in Disturbing Ways. *WIRED*. https://www.wired.com/story/deepfake-nude-abuse/

Caldwell, M., Andrews, J. T. A., Tanay, T. et Griffin, L. D. (2020). *Ai-enabled future crime.* BioMed Central. https://crimesciencejournal.biomedcentral.com/articles/10.1186/s40163-020-00123-8

Capron, A. (2022, 17 mars). Des vidéos "deepfake" de Zelensky et Poutine émergent en marge de la guerre en Ukraine. *Les Observateurs - France 24*. https://observers.france24.com/fr/europe/20220317-deepfake-zelensky-poutine-ukraine-russie-video-fake

Carasco, A. (2022, 20 janvier). Baromètre des médias 2022 : la confiance des Français au plus bas. *La Croix*. https://www.la-croix.com/Economie/Barometre-medias-2022-confiance-Francais-bas-2022-01-20-1201195923

Cavazos, R. (2020). *The Massive Indirect Cost of Digital Ad Fraud*. https://cheq.ai/the-massive-indirect-cost-of-digital-ad-fraud/

Chesney, R. et Citron, D. K. (2018). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.* https://doi.org/10.2139/ssrn.3213954

Clarke, Y. D. (2019). *H.R.3230 - 116th Congress (2019-2020): Deep FAKES Accountability Act*. https://www.congress.gov/bill/116th-congress/house-bill/3230

Cole, S. (2017, 12 novembre). AI-Assisted Fake Porn Is Here and We're All Fucked. *VICE*. https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn

Cunningham, R. (2021). *Introducing Power Apps Ideas: AI-powered assistance now helps anyone create apps using natural language | Microsoft Power Apps*. https://powerapps.microsoft.com/en-us/blog/introducing-power-apps-ideas-ai-powered-assistance-now-helps-anyone-create-apps-using-natural-language/

Danzon-Chambaud, S. (2022, février 9). *Experimenting with automated news at the BBC*. https://www.cjr.org/tow_center/the-tow-center-newsletter-experimenting-with-automated-news-at-the-bbc.php

Deschamps, C. (2021, novembre 15). *Les technologies du faux : un état des lieux*. https://observatoire-strategique-information.fr/2021/11/15/les-technologies-du-faux-un-etat-des-lieux/

Gleicher, N., Franklin, M., Agranovitch, D., Nimmo, B., Belogolova, O., & Torrey. (2021). *Threat report - The State of Influence Operations 2017-2020*. https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, juin 10). *Generative Adversarial Networks*. https://arxiv.org/pdf/1406.2661

Goria, S. (2006). Knowledge management et intelligence économique : deux notions aux passés proches et aux futurs complémentaires. *Informations, Savoirs, Décisions et Médiations [Informations, Sciences for Decisions Making]*(27), 1–16. https://hal.archives-ouvertes.fr/inria-00110300v1

Gu, L., Kropotov, V., & Yarochkin, F. *The fake news machine: How propagandists abuse the internet and manipulate the public*. https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/fake-news-cyber-propaganda-the-abuse-of-social-media

Harbulot, C. (1992). *La machine de guerre économique : États-Unis, Japon, Europe*. Economica; Impr. Europe média duplication. https://doi.org/10.3917/econo.harbu.1992.01

Hatmaker, T. (2018, 1 mai). DARPA is funding new tech that can identify manipulated videos and 'deepfakes'. *TechCrunch*. https://techcrunch.com/2018/04/30/deepfakes-fake-videos-darpa-sri-international-media-forensics/

Huyghe, F.-B [Francois-Bernard]. (2010). *Principales notions sur la stratégie de l'information : Dictionnaire critique*. https://www.huyghe.fr/wp-content/uploads/2021/11/47289ed3f2c1e.pdf

Huyghe, F.-B [François-Bernard] (2016). Désinformation : armes du faux, lutte et chaos dans la société de l'information. *Securite globale*, *N° 6*(2), 63–72. https://www-cairn-info.ressources.univ-poitiers.fr/revue-securite-globale-2016-2-page-63.htm

Joshi, A [Anupam], Ranade, P. et Finin, T. (2021). *Study shows AI-generated fake reports fool experts*. https://theconversation.com/study-shows-ai-generated-fake-reports-fool-experts-160909

Knight, W. (2021, 20 septembre). AI Can Write Code Like Humans—Bugs and All. *WIRED*. https://www.wired.com/story/ai-write-code-like-humans-bugs/

Köbis, N. et Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, 106553. https://doi.org/10.1016/j.chb.2020.106553

Kreps, S. E., McCain, M. et Brundage, M. (2020). All the News that's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.3525002

LeCun, Y., Bengio, Y. et Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Luebke, D. (2021). *Researchers Use NVIDIA AI to Help Mitigate Misinformation | NVIDIA Blog*. https://blogs.nvidia.com/blog/2021/11/11/how-researchers-use-nvidia-ai-to-help-mitigate-misinformation/

Malik, A. (2022, 23 février). Mark Zuckerberg demos a tool for building virtual worlds using voice commands. *TechCrunch*. https://techcrunch.com/2022/02/23/mark-zuckerberg-demos-a-tool-for-building-virtual-worlds-using-voice-commands/?guccounter=1

Martre, H., Clerc, P., & Harbulot, C. (1994). *Intelligence économique et stratégie des entreprises*. https://www.vie-publique.fr/sites/default/files/rapport/pdf/074000410.pdf

Moinet, N. (2019). Le renseignement au prisme du couple agilité-paralysie. *Prospective et strategie*, *Numéro 10*(1), 13–27. https://www.cairn.info/revue-prospective-et-strategie-2019-1-page-13.htm

Nightingale, S. J. et Farid, H. (2022). Ai-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*(8). https://doi.org/10.1073/pnas.2120481119

Nimmo, B., François, C., Eib, C. S. et Ronzaud, L. (2020). *IRA again: Unlucky thirteen : Facebook takes down small, recently created network linked to internet research agency.* Graphika. https://public-assets.graphika.com/reports/graphika_report_ira_again_unlucky_thirteen.pdf

Pennycook, G. et Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Sajidur, R., Tully, P. et Foster, L. (2019). *Attention is All They Need: Combatting Social Media Information Operations With Neural Language Models | FireEye Inc*. https://www.fireeye.com/blog/threat-research/2019/11/combatting-social-media-information-operations-neural-language-models.html

Satter, R. (2019, 13 juin). Experts: Spy used AI-generated face to connect with targets. *Associated Press*. https://apnews.com/article/ap-top-news-artificial-intelligence-social-platforms-think-tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d

Scheid, J. (2016). Réflexions sur la falsification et le faux dans la Rome antique. *Comptes-rendus des séances de l année - Académie des inscriptions et belles-lettres*, *160*(1), 91–103. https://doi.org/10.3406/crai.2016.95872

Schiavon, N. (2019, 16 mars). La psychographie pour le marketing fondé sur les données. *METADOSI*. https://www.metadosi.fr/comment-utiliser-psychographie-pour-marketing-fonde-donnees/

Sedova, K., McNeill, C., Johnson, A., Joshi, A [Aditi] et Wulkan, I. (2021). *AI and the Future of Disinformation Campaigns: Part 1: The RICHDATA Framework.* Center for Security and Emerging Technology. https://doi.org/10.51593/2021CA005

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K., Flammini, A. et Menczer, F. (2018). *The spread of low-credibility content by social bots* (Vol. 9). https://www.arxiv-vanity.com/papers/1707.07592/ https://doi.org/10.1038/s41467-018-06930-7

Skandarani, Y., Jodoin, P.-M., & Lalande, A. (2021, mai 11). *GANs for Medical Image Synthesis: An Empirical Study*. https://arxiv.org/pdf/2105.05318

So-Yeon, Y. (2022, 28 mars). Drag and drop: Deepfakes create a new kind of crime. *Korea JoongAng Daily*. https://koreajoongangdaily.joins.com/2020/05/17/features/deepfake-artificial-intelligence-pornography/20200517190700189.html

Stupp, C. (2019, 30 août). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

Sun, X., Zhang, X., Xia, Z. et Bertino, E. (dir.). (2021). *Communications in Computer and Information Science*. *ADVANCES IN ARTIFICIAL INTELLIGENCE AND SECURITY : 7th international conference*. SPRINGER NATURE. https://doi.org/10.1007/978-3-030-78621-2

Vandeginste, P. (2021, 25 mars). Détecter les fausses nouvelles grâce au NLP. *Data Analytics Post*. https://dataanalyticspost.com/analyser-les-textes-grace-au-nlp/

Ye, J. (2022, 29 janvier). China targets deepfakes in proposed regulation governing deep learning AI technologies. *South China Morning Post*. https://www.scmp.com/tech/policy/article/3165244/china-targets-deepfakes-proposed-regulation-governing-deep-learning-ai

Zhao, B., Zhang, S., Xu, C., Sun, Y. et Deng, C. (2021). Deep fake geography? When geospatial data encounter Artificial Intelligence. *Cartography and Geographic Information Science*, *48*(4), 338–352. https://doi.org/10.1080/15230406.2021.1910075