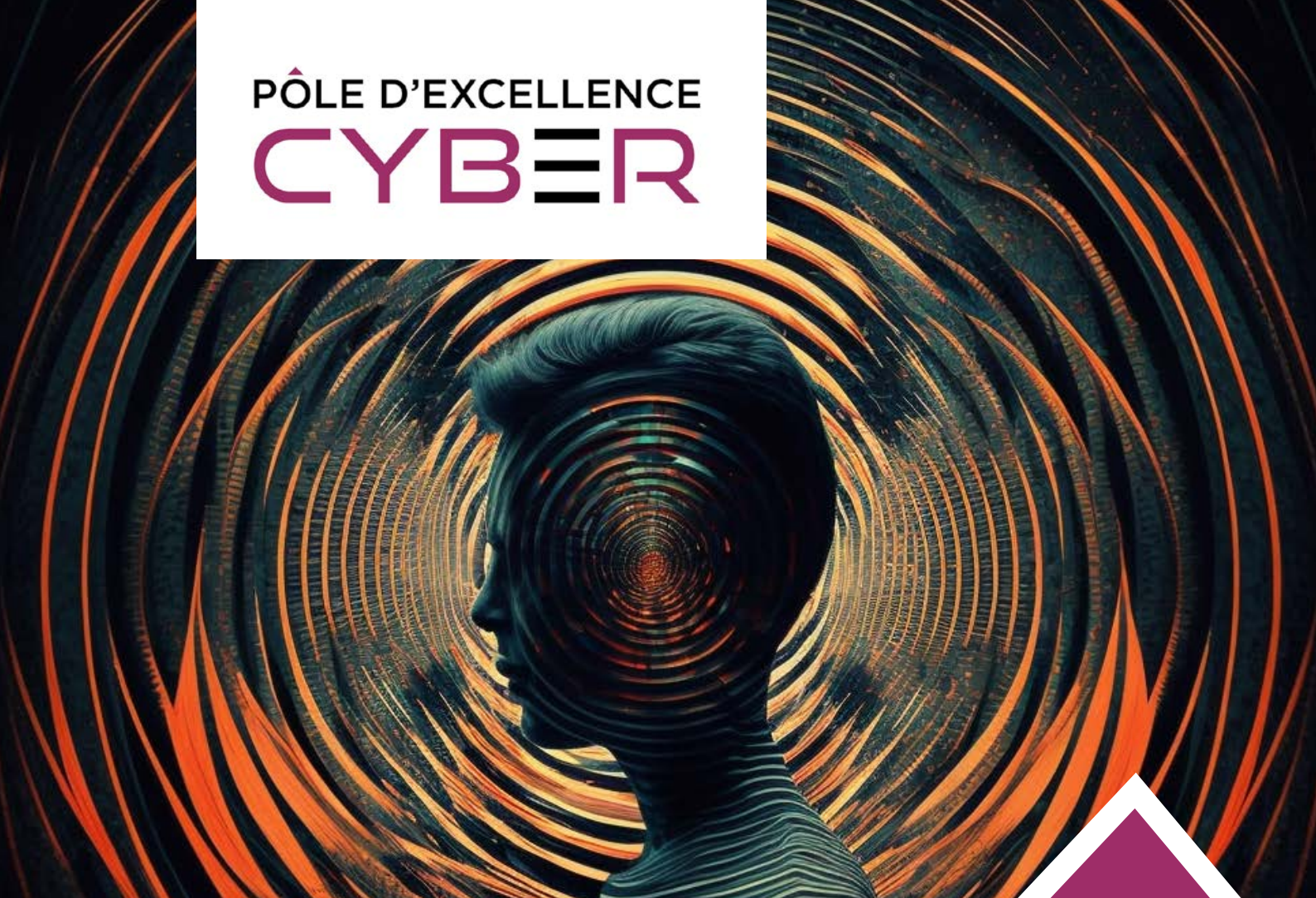


PÔLE D'EXCELLENCE
CYBER



Lutte contre les manipulations de l'information

Regards croisés de spécialistes et d'acteurs du domaine

.....
Mai 2023





Sommaire

Préface

Jean-Yves LE DRIAN 5

Introduction

Jean-Luc GIBERNON / Jean-Philippe Riant 7

PARTIE I : Contexte défense et géopolitique 8

Désinformation : quels enjeux ? quels effets systémiques ? 10

Lutte contre la manipulation de l'information : définitions et concepts
clefs 12

La centralité des manipulations de l'information dans les rapports
géopolitiques 14

La stratégie de prédation russe s'appuie sur la désinformation en
Afrique 16

Exploitation du désir de confort : la victoire de l'entre-soi et mise en
danger de la démocratie 18

Enseignement supérieur et désinformation : quel bilan dans les écoles
d'ingénieur ? 20

PARTIE II : Aspects opérationnels et techniques 22

Moyens technologiques actuels pour lutter contre les manipulations
d'informations 24

Lutte contre les manipulations de l'information : la technologie et sa
nécessaire régulation 26

Authenticité de l'information : le rôle essentiel des opérateurs 28

Comment déceler le manipulateur derrière l'avatar 30

Les bots sociaux et la manipulation de l'information 32

La lutte informatique d'influence 34

Renseigner, défendre et agir face aux infox
Retours d'expérience et interdisciplinarité au service des luttes
informatiques 36

Usurpation d'identité par le biais des Deep Fakes dans le Metaverse 38

PARTIE III : Un environnement juridique en construction.	40
Lutte contre la désinformation : quelle régulation juridique ?	42
L'Union européenne et la lutte contre les campagnes de manipulation de l'information dans le contexte du conflit armé en Ukraine.	44
Les enjeux pour les droits fondamentaux de la lutte contre la manipulation de l'information	46
Dispositifs étatiques pour lutter contre la manipulation d'information. . . .	48
Lexique	52
Notes et références	58
Conclusion.	64



Copyright Pôle d'excellence cyber©. Édition de avril 2023.

Cette œuvre est mise à disposition sous licence Creative Commons,

Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 France.

Pour voir une copie de cette licence, visitez <http://creativecommons.org/licenses/by-nc-nd/3.0/fr/> ou écrivez à Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

PRÉFACE



L'espace informationnel est l'un des champs d'action privilégiés des puissances qui, dans le dessein de pousser leur avantage et de faire valoir leur propre modèle, entendent saper les fondements de nos sociétés démocratiques, entraver notre influence dans le monde et détruire les conditions mêmes de l'action collective internationale, pourtant indispensable face aux grands bouleversements écologiques, technologiques et humains du XXI^e siècle. Afin d'y parvenir, les plus désinhibées d'entre elles n'hésitent ni à tenter de fausser les règles du débat public, ni à travestir la réalité de nos engagements auprès de nos partenaires, ni à relativiser les faits pour imposer le relativisme des valeurs.

De telles offensives, et les batailles qu'elles nous imposent de mener pour défendre nos intérêts, sont d'autant plus redoutables qu'elles mettent en jeu, dans l'immense majorité des cas, ces nouveaux repères quotidiens que sont smartphones, médias numériques et réseaux sociaux. Dans un monde en voie de brutalisation, force est désormais de reconnaître qu'il peut s'agir d'autant d'armes par destination – et pas uniquement, d'ailleurs, entre les mains de nos compétiteurs.

Pour répondre à ce défi informationnel d'un genre nouveau, il est donc essentiel de mobiliser l'expertise et les capacités d'action de notre écosystème de cyberdéfense. C'est pourquoi je me félicite de voir le Pôle d'excellence cyber s'emparer de ce sujet majeur, fort de la position stratégique qui est la sienne à l'intersection de nos Armées, du secteur privé et du monde académique.

Croiser les regards sur les manipulations de l'information est la meilleure manière d'en saisir les ressorts, d'en mesurer les enjeux et d'y apporter des parades efficaces. C'est ce qui rend les pages qu'on s'apprête à lire aussi stimulantes, et les perspectives qu'elles ouvrent au Pôle d'excellence cyber aussi prometteuses.

Jean-Yves Le Drian

Ancien ministre de l'Europe et des Affaires étrangères

Ancien ministre de la Défense

INTRODUCTION



Jean-Luc GIBERNON

Vice-président
du Pôle d'excellence cyber



Dr. Jean-Philippe Riant

Directeur conseil
en intelligence artificielle
et communication digitale



Réseaux sociaux, fake news, deep fakes... les fausses nouvelles et les tentatives de désinformation prolifèrent de plus en plus, année après année. Le phénomène est tel que, aujourd'hui, l'ensemble des opérations de manipulation de l'information menacent très sérieusement nos sociétés démocratiques.

En effet, si la propagande, les luttes d'influence et les manipulations de l'information sont un concept vieux comme le monde, leur importance dans le monde ultra connecté dans lequel nous vivons ne fait que se renforcer. De nos jours, les réseaux sociaux sont en effet largement utilisés pour relayer des opérations d'influence, dans lesquelles les utilisateurs des réseaux sociaux deviennent à la fois les victimes des manipulations et bien souvent les complices involontaires, dès lors qu'ils relayent ces fausses informations. Dans le flot de données et messages qui circulent sur les réseaux sociaux, il est excessivement difficile de détecter automatiquement une tentative de manipulation ou d'ingérence.

Comme les solutions techniques qui permettraient d'appuyer cette lutte sont aujourd'hui très peu développées, la capacité de réaction et de réponse reste aujourd'hui encore très limitée. Par ailleurs, les instruments juridiques de la lutte contre les manipulations de l'information, qu'ils se placent au niveau français ou européen, sont aujourd'hui largement insuffisants. Lutter contre les manipulations de l'information requiert pourtant des outils juridiques adaptés aux besoins de cette lutte, qui, aujourd'hui, restent encore largement à initier ou consolider.



Face à ce constat, il est indispensable de chercher à y voir plus clair et préparer un plan d'action ambitieux national voire européen. C'est l'objectif que s'est fixé le Pôle d'excellence cyber (PEC), qui a conduit à la création du groupe de travail « Lutte contre les manipulations de l'information », dans lequel les membres du PEC ont pu se concerter, partager une vision claire de la situation et rédiger collectivement cet ouvrage.

Face à la complexité et aux enjeux que représente la lutte contre les manipulations de l'information, le PEC est particulièrement bien placé pour piloter une réflexion, faire avancer le débat public et proposer un plan de lutte. Association fondée en 2014 par le ministère des Armées et la région Bretagne, le PEC regroupe en effet plus de 100 membres issus des entreprises, du monde universitaire, du monde de la recherche et des services publics, qui sont autant d'acteurs de cette lutte (civils et militaires, spécialistes techniques et juridiques, etc.).

Les travaux de notre groupe de travail ont confirmé à quel point il est essentiel de faire prendre conscience de la situation et des enjeux. L'analyse par points de vue successifs des acteurs les plus pertinents en France sur le sujet permet de lever le brouillard, d'y voir clair et d'envisager un plan d'action pour la lutte. Telle est l'ambition de cet ouvrage.

Jean-Luc GIBERNON / Jean-Philippe RIAN
Coordinateurs de cet ouvrage

Contexte défense & géopolitique

PARTIE

Contributeurs



Pr. Brunessen BERTRAND



Juliette MOREAU



Dr. Julien NOCETTI



Commandement de la Cyberdéfense



Dr. Jean-Philippe RIANT



Marie PUREN



Désinformation : quels enjeux ? quels effets systémiques ?



Pr. Brunessen BERTRAND

Professeure de droit
Université de Rennes
Chaire Jean Monnet



Effets systématiques de la désinformation

Si la désinformation et la manipulation ont toujours existé, le vecteur qu'elles empruntent aujourd'hui, les plateformes numériques, leur donne des effets systémiques. De la publicité politique aux tentatives de désinformation et de manipulations, les plateformes numériques constituent un vecteur évident de perturbation des processus démocratiques : cela est lié aux effets systémiques de leur modèle économique, au fait que ce modèle repose sur la collecte massive de données qui permettent d'établir un profil très précis des utilisateurs, donc de cibler les contenus qui leur sont adressés. Les algorithmes de recommandation sont alimentés par ces données, ciblant ainsi les préférences individuelles des individus. Ce profilage est propice à la diffusion virale des informations, surtout celles qui sont fausses. Si les très grandes plateformes ont pu alimenter ces phénomènes, y compris parfois un peu à leur insu, les nouveaux usages numériques ont amplifié les risques : la bulle informationnelle dans laquelle s'enferment des citoyens qui ne s'informent que par des réseaux sociaux a aussi contribué à entretenir ce phénomène qui tourne largement autour de ces plateformes numériques. Les enjeux de l'activité des plateformes, notamment de réseaux sociaux, sur les démocraties sont considérables, car ils révèlent de nouvelles vulnérabilités, en particulier la question de la manipulation des opinions publiques et de la désinformation véhiculée par ces plateformes numériques. La captation du débat public par les plateformes

numériques devient désormais un enjeu démocratique fondamental. Elles sont le lieu de tous les excès : de la haine en ligne aux multiples formes de désinformations, en passant par la censure d'un président en exercice, les enjeux sont considérables.

Un enjeu démocratique fondamental

Ces enjeux démocratiques ne sont cependant pas dissociés des enjeux économiques dès lors que la fabrique de la désinformation alimente toute une économie. Ici comme ailleurs, l'Union européenne cherche à défendre ses valeurs : démocratie, protection des droits fondamentaux, sécurité et confiance numériques. Mais cette défense est biaisée par la dépendance européenne à des plateformes qui ne le sont pas. Cette absence de plateforme européenne est forcément un sujet de préoccupation, au regard de l'instrumentalisation politique dont elles peuvent être l'objet ou qu'elles peuvent contribuer à alimenter, même parfois à leur insu. Les démocraties sont fragiles face aux campagnes de désinformation et de manipulation de l'opinion dont l'effet est démultiplié par des algorithmes alimentés par les données personnelles des citoyens : c'est ce que l'affaire Cambridge Analytica a révélé de façon spectaculaire. Une société de conseil en communication, dont la mission était de changer les comportements par le traitement des données (« Data-driven behavior change »), avait utilisé les données personnelles de dizaines de millions d'utilisateurs de Facebook

pour cibler des publicités et messages politiques afin d'orienter les résultats de l'élection américaine. Ces formes de désinformation sont d'autant plus fortes que le numérique est devenu la principale source d'information des citoyens : une étude conduite par l'Université d'Oxford montre que 66 % des Français s'informent via des contenus numériques et près de 40 % uniquement par les réseaux sociaux (dont 60 % sur smartphones). La régulation des contenus devient un enjeu démocratique fondamental, mais sa concrétisation, à l'échelle européenne, n'a été que progressive.

Stratégies d'ingérences étrangères et guerre hybride

La lutte contre la désinformation passe par d'abord l'identification des formes de désinformation : des stratégies d'influence aux cyberattaques, les formes sont multiples et n'appellent pas toutes la même réponse juridique. Elle passe également par la compréhension des vecteurs de la désinformation pour adopter une régulation pertinente : on songe en premier lieu aux plateformes numériques, mais il ne faut pas oublier la part des usages et des comportements individuels et collectifs. Les campagnes de désinformation font souvent partie d'une guerre hybride comprenant des cyberattaques et le piratage de réseaux. La protection de l'intégrité des élections et des processus démocratiques passe donc aussi par la question des cyberattaques qui peuvent cibler des infrastructures électorales critiques. Les cyberattaques contre les infrastructures électorales critiques contribuent à la manipulation de l'opinion. Il peut s'agir d'intrusions ciblées pour collecter des informations sensibles et organiser des fuites : ce sont d'abord les pratiques de « hack and leak » (HLO), c'est-à-dire de piratage et de divulgation, qui peuvent avoir lieu avec ou sans falsification de ces informations.

Ce sont ensuite des perturbations de systèmes informatiques notamment d'entreprises de radiotélédiffusion ou de commissions électorales. Les élections démocratiques sont devenues la cible de ces cyberstratégies malveillantes, comme la campagne électorale de l'élection présidentielle américaine de 2016 qui a fait l'objet de hack and leak operation par les services de renseignement russes. Ces différentes formes de désinformation n'appellent pas toutes la même réponse juridique et politique.

Lutte contre la manipulation de l'information définitions et concepts clefs



Juliette MOREAU

Cadettes de la cyber
Etudiante en master 2
Sciences Po Rennes



De quoi parle-t-on ?

Les sociétés modernes sont de plus en plus confrontées à des problématiques de manipulation de l'information. Représentant une menace réelle pour les démocraties occidentales, ce phénomène n'est pourtant pas nouveau. Dès l'Antiquité, les philosophes grecs ont compris que la parole était un enjeu de pouvoir. Platon, dans les Dialogues, et Aristote, dans la Rhétorique, se sont intéressés au traitement de l'information et à sa manipulation. Dans l'Illiade, le héros grec n'hésite pas à recourir à un stratagème, le cheval de Troie, pour s'introduire sans coup férir dans la cité troyenne. N'étant pas nouvelle, la manipulation de l'information a toutefois changé d'échelle au XXIème siècle avec le développement des médias et l'essor des réseaux sociaux. Ce phénomène est une arme efficace qui représente une menace réelle pour les démocraties et un défi en termes de sécurité. Depuis les attaques menées contre l'État estonien en 2007, le cyberspace est considéré comme un nouveau champ de conflictualité au même titre que la conquête de la terre, de la mer, de l'air et de l'espace. Ainsi, l'élection présidentielle américaine de 2016⁽¹⁾ a pu faire l'objet d'importantes manipulations de l'information et de cyberattaques, identifiées pour certaines comme étant des ingérences russes dans le fonctionnement démocratique des États-Unis. En cause, la publication d'échanges confidentiels de la candidate Clinton en se servant de sa messagerie mél personnelle, campagne médiatique orchestrée par Facebook, utilisation d'usines à trolls pour influencer l'opinion publique. Les fausses informations peuvent être utilisées avec de nombreux objectifs, parmi lesquels l'étouffement d'une affaire ou la délégitimation d'un scrutin. Derrière cette idée d'ingérence étrangère, il ne s'agit pas nécessairement de favoriser un candidat par connivence politique mais plutôt de déstabiliser l'État américain. Derrière ces questions,

l'enjeu est aussi celui de la manipulation des cerveaux. C'est dans cette perspective que se justifie l'usage abondant de « nudges ». De manière générale, la menace provoquée par une manipulation de l'information peut être soit endogène, c'est-à-dire à l'intérieur de la communauté visée, soit exogène, comme dans l'exemple évoqué plus tôt. Cependant, les États n'en sont pas nécessairement à l'origine. Les manipulations de l'information peuvent être opérées également par des entreprises, des organisations ou bien encore des groupes informels. Le plus souvent, ces manipulations se recoupent en trois grands domaines⁽²⁾. Il y a bien évidemment la sphère politique, militaire, mais également financière et industrielle.

Fake news

ou manipulation de l'information ?

La différence entre les deux notions peut paraître ténue mais mérite une clarification. Le terme anglo-saxon prend en compte les informations erronées à proprement parler, tandis que cet ouvrage s'intéresse au traitement de l'information et à ses manipulations, c'est-à-dire celles qui sont sciemment orientées. Pour rappel, le but est de manipuler l'opinion publique, pour - in fine - perturber la décision démocratique. Du latin *manipulare*, l'action de manipuler renvoie au fait de « conduire par la main ». L'idée sous-jacente est ici celle d'une paralysie du jugement pour en forcer l'adhésion. Pour compléter cette définition, il faudrait rappeler que toute information publiée en tant que telle comporte une part importante de subjectivité. Cela rend la distinction entre l'erreur humaine et la manipulation préméditée loin d'être évidente. L'étymologie de « information » est à cet égard significative : le terme vient du latin *informare* qui signifie littéralement « donner une forme à ».

Un enjeu majeur ?

La question de la lutte contre les manipulations de l'information est devenue un tel enjeu sécuritaire et politique qu'il est désormais admis de parler de « guerre de l'information »⁽³⁾. Cette expression était utilisée dès 1976 par un ingénieur de l'entreprise Boeing dans l'étude « Systems and Information War ». Il est de plus en plus difficile de lutter contre ces manipulations en raison de la complexité du traitement de l'information : pluralité d'acteurs, des canaux de transmission informationnels et surtout des modes opératoires utilisés. Tout cela est à mettre en perspective avec l'émergence du web, de ses plateformes numériques et de ses réseaux sociaux. Raphael Chauvancy, officier supérieur des Troupes de marine, va aussi dans le même sens lorsqu'il déclare que « l'âge de l'information est aussi celui où, paradoxalement, les masses ont de plus en plus de difficultés, réelles ou perçues, à accéder à une information fiable »⁽⁴⁾. Manipuler l'information reste relativement peu technique et surtout peu risqué, notamment grâce à l'usage d'un pseudonyme.

L'avancée des technologies paraît très nettement devancer les outils de lutte contre les manipulations de l'information, comme en témoigne la réalisation de « deep fake ». En réponse à ces menaces, le 13 juillet 2021, l'État a créé un service du nom de « Viginum » pour lutter contre les ingérences numériques étrangères (INE). Ainsi, sur la période électorale 2022, le service technique et opérationnel a détecté cinq phénomènes suspects qui réunissaient tous les critères d'une ingérence en ligne⁽⁵⁾. Mais globalement, il faut le reconnaître, les technologies de lutte contre les manipulations de l'information sont actuellement peu développées, qu'il s'agisse de la data visualisation, de l'analyse sémantique ou bien encore de l'Intelligence Artificielle. De surcroît, la législation encadrant la manipulation de l'information se trouve confrontée à des questions nouvelles et non résolues. C'est notamment le cas de l'encadrement juridique du mensonge et des fausses rumeurs qui n'ont donné lieu qu'à très peu d'études alors qu'il apparaît nécessaire de réguler ces pratiques.



La centralité des manipulations de l'information dans les rapports géopolitiques



Dr. Julien NOCETTI

Enseignant-chercheur
Responsable de chaire
SAINT-CYR / IFRI / RSB



“ Diffuser la désinformation, c’est jouer avec la vie humaine. La désinformation tue”

Par ces mots émis au printemps 2020, le haut représentant de l'Union européenne pour les affaires étrangères et la politique de sécurité, Josep Borrell, mettait en garde contre les conséquences d'une désinformation ciblée en lien avec le coronavirus. Deux ans plus tard, soit quelques semaines après le début de l'invasion russe de l'Ukraine, la directrice du média d'Etat russe RT, Margarita Simonyan, comparait les plateformes américaines du numérique à « des armées étrangères qui [nous] tirent dessus ». L'actualité internationale récente, phagocytée par la pandémie de Covid-19 et la guerre d'Ukraine, a rappelé l'acuité des enjeux et menaces informationnels. Chaque crise - nationale, régionale ou internationale - est désormais assortie de son lot de suspicions de manipulations de l'information. Les huit dernières années l'ont montré à l'envi, avec une forte visibilité donnée aux opérations d'influence, surtout numériques, de la Russie et de sa nébuleuse informationnelle. De l'annexion de la Crimée (2014) aux élections présidentielles américaines de 2016, en passant par le conflit syrien (depuis 2011) et l'affaire Skripal (2018), les autorités russes ont anticipé le fait que l'âge du tout-numérique s'est mué en ère de la désinformation, retournant à leur profit le discours forgé par la diplomatie américaine sur le potentiel émancipateur de l'Internet.

L'espace numérique, théâtre privilégié des manipulations de l'information

Armé d'un smartphone, chacun devient un « producteur de contenus » qui se diffusent au-delà du strict cadre de la zone géographique concernée. Cette dilatation de l'espace s'accompagne d'un phénomène inverse pour le temps. Là où il fallait parfois plusieurs jours à une dépêche pour rejoindre une capitale européenne, les réseaux sociaux alimentent, en continu, un public avide et émotionnellement réceptif. Les contenus numériques ne sont plus des informations au sens classique du terme, ils ne relèvent pas toujours de l'action journalistique et ne sont, la plupart du temps, soumis à aucune forme de vérification. Sans contrôle des contenus ni commentaires, il revient à chacun de mettre en perspective pour relativiser le poids croissant de l'émotion. Cette faiblesse ouvre opportunément le champ à des manipulations grossières, ou plus subtiles, qui favorisent et entretiennent une forme de doute permanent au cœur des opinions publiques en démocratie. Les ingérences réelles ou supposées de puissances étrangères dans les processus électoraux en sont une illustration.

Les démocraties fragilisées

Plus largement, les tentatives d'ingérences informationnelles – parfois couplées avec des cyberattaques – accentuent une brutalisation numérique de la vie politique internationale. L'information est devenue une arme dans une véritable guerre d'influence et de déstabilisation. De la mésinformation à la désinformation, du truquage de vidéo et au partage de données, en passant par la création de faux comptes orchestrant des raids numériques, diverses stratégies sont mises en œuvre pour fragiliser la démocratie en ciblant deux de ses piliers : le vote et la confiance. L'avantage va nettement aux pourvoyeurs de désinformation puisque les fausses informations se propagent nettement plus vite sur Twitter que les vraies informations. Une autre difficulté majeure tient à l'alignement de facto des intérêts des principaux acteurs de l'économie de l'attention (les GAFAM – Google, Apple, Facebook, Amazon et Microsoft) et ceux produisant de la désinformation politique. Des algorithmes d'apprentissage automatique sont déjà intégrés dans les plateformes de publicité ciblée et l'analyse des données complexes. Le scandale Cambridge Analytica, révélé en mars 2018, avait exposé la frontière ténue entre publicité politique et propagande. Au-delà de la campagne présidentielle américaine en 2016, durant laquelle les données personnelles de près de 90 millions d'utilisateurs américains de Facebook avaient été siphonnées par la société de conseil, la campagne pour le Brexit avait décelé l'urgence de démêler la convergence des intérêts entre les grandes plateformes du numérique, dont les ressorts demeurent commerciaux, et les entrepreneurs du marketing politique recourant à la désinformation numérique.

Cas d'étude #1 :

L'«étrange intermède » de l'élection présidentielle américaine (2020)

En pleine pandémie, la campagne électorale aux Etats-Unis, puis le scrutin, en novembre 2020, n'ont pas donné lieu à une répétition du scénario de 2016. Les tentatives de manipulations informationnelles sont venues moins de l'étranger que de l'intérieur du pays. L'apparente indigénisation des opérations informationnelles contribue à brouiller les repères en la matière et à polariser davantage la société américaine. L'élection américaine a, surtout, donné lieu à une mise en avant sans précédent des réseaux et médias sociaux les plus populaires (Facebook, YouTube, Twitter). La neutralité revendiquée par les grandes plateformes a été mise à l'épreuve par Donald Trump et le camp républicain, qui se sont affranchis de toutes les règles habituelles du débat politique. Le consensus sur la vérité s'est effrité – et, avec lui, la position immuable des grands réseaux sociaux, qui répètent à l'envi depuis leur création qu'ils ne sont aucunement des médias mais de simples intermédiaires techniques.

Cas d'étude #2

L'affirmation chinoise dans le champ informationnel

La Chine inscrit sa démarche informationnelle dans l'affichage décomplexé de sa puissance nationale. Sa stratégie consiste à positionner son modèle de gouvernance comme une alternative attractive à la démocratie libérale des Occidentaux. Cette approche, défendue par Xi Jinping depuis 2013, se veut une stratégie de politique étrangère de long terme. La stratégie informationnelle de Pékin se caractérise de façon croissante par un ton désinhibé, en rupture avec le soft power de ces dernières années projetant l'image d'une Chine « bienveillante » au monde. À cette fin, la Chine semble s'appuyer sur les tactiques éprouvées développées par la Russie, qu'elle oriente vers le même objectif : semer le doute chez les opinions publiques comme chez les décideurs occidentaux. Comme la Russie, la Chine intègre dans sa stratégie différentes briques qui se complètent : dissémination de multiples récits afin de répandre la confusion à l'étranger, alternance entre rhétorique consensuelle et propos véhéments, attaques informatiques. Ces tactiques comprennent l'utilisation de chaînes officielles pour diffuser des théories conspirationnistes variées et parfois contradictoires, y compris en amplifiant les propos complotistes de médias marginaux en recourant à un appareil médiatique d'État tentaculaire afin de les faire gagner en visibilité voire en viralité.

La stratégie de prédation russe s'appuie sur la désinformation en Afrique



Commandement
de la Cyberdéfense



Compétition, contestation et affrontement sont désormais les trois états qui structurent les relations internationales et la géopolitique contemporaine. L'un des outils mis en œuvre dans ce triptyque est la manipulation de l'information qui trouve un point d'application particulièrement actuel dans la notion de guerre hybride. C'est dans ce contexte que la France est ciblée par des démarches agressives de manipulations de l'information, coordonnées et pilotées notamment par la Russie.

Un cas emblématique mais dramatique : la désinformation en Afrique

Après être parvenue à prendre l'ascendant dans l'espace informationnel malien la manœuvre de désinformation massive russe se déploie désormais dans d'autres pays d'Afrique de l'Ouest, notamment au Burkina Faso. Selon un schéma identique à celui qui a pu être observé en République centrafricaine et au Mali, on observe la mobilisation des acteurs panafricanistes, l'implantation dans des communautés locales, la viralisation inauthentique des contenus par des fermes à trolls. Les narratifs utilisés sont similaires : stratégie de confusion entre la présence de la France et les enjeux sociaux et sécuritaires, ingérence supposée de la France contre les nouvelles autorités, appel à réduire la présence militaire française et dénonciation des médias français. Cette désinformation massive organisée par les officines de la sphère de l'oligarchie russe à la tête de la société militaire privée Wagner, Evgueni Prigojine, vise à provoquer ou au moins à nourrir un sentiment anti-français.

- Certaines de ces campagnes sont menées par des sociétés proches d'Evgueni Prigojine sous couvert de consulting politique. Opérant comme un lobby agressif dont le but est de faire accéder au pouvoir des candidats pro-russes ces sociétés ciblent les armées françaises qui sont quotidiennement attaquées. La palme de l'attaque cynique revient à l'accusation sous de multiples formes d'un soutien de la France aux groupes terroristes d'Afrique de l'Ouest, alors même qu'elle les combat depuis plus d'une décennie et a perdu 59 de ses soldats au Sahel.
- Depuis la fin de l'année 2022, une vidéo qui circule sur les réseaux sociaux est supposée montrer l'ambassadeur de France au Congo pris dans une foule hostile à Kinshasa. Dans les faits, si la vidéo a bien lieu dans la ville de Kinshasa, elle ne montre pas l'ambassadeur de France mais un attaché de l'ambassade de Turquie et un homme d'affaires libanais.



Exemple d'images identiques utilisées dans RIAFAN et un media local au profit des actions russes en république centrafricaine. Ici, une manifestation contre MINUSCA.

Derrière ces attaques informationnelles une stratégie globale de prédation

Les attaques informationnelles menées par des acteurs non-étatiques proches de la Russie s'inscrivent dans le cadre d'une stratégie globale d'influence de Moscou qui met en œuvre des stratégies hybrides pour restaurer son influence ou servir des intérêts privés à des fins d'enrichissement en prenant le contrôle des ressources naturelles des Etats, telles que les diamants en Centrafrique, l'or au Soudan ou encore le pétrole en Libye. Ces attaques révèlent une stratégie de prédation économique réalisée à l'échelle du continent africain et plusieurs articles ont démontré que de nombreuses filiales exploitant les mines d'or du Soudan ou les mines de diamants en Centrafrique sont en majorité liées à la sphère Prigojine et ne rétribuent pas les Etats. Ces narratifs diffusés par les influenceurs panafricanistes pro-Russie alimentent confusion et colère au sein des populations pour qui la France devient un bouc émissaire et permet de se détourner des réels problèmes que traversent ces pays (corruption, insécurité, faillite économique...). Ils conviennent à ceux qui ont pris le pouvoir par la force (Mali...) afin masquer leurs échecs et leur absence de légitimité.

Comment faire face à ces attaques ?

Dans un premier temps, dénoncer les contradictions de ces attaques et rappeler l'absence de légitimité des relais de l'influence russe

Les panafricanistes dénoncent « le colonialisme occidental » mais ferment les yeux sur l'impérialisme russe : l'invasion de l'Ukraine par la Russie est l'expression la plus brutale et la plus violente de l'impérialisme russe.

Le véritable panafricanisme est un mouvement légitime et respecté prônant l'indépendance totale de l'Afrique, il ne réside pas dans le rejet de l'Occident mais dans la constitution d'une Afrique souveraine, aux partenariats diversifiés. La France soutient pleinement la volonté des Etats africains de diversifier leurs partenariats. Le partenariat avec la France n'est pas exclusif et ne l'a jamais été. Les auditoires de ces « panafricanistes » sont majoritairement issus des diasporas et non pas des populations africaines. Leurs discours superficiels ne sont pas connectés aux préoccupations des populations locales qu'ils fréquentent peu. Enfin, ces « panafricanistes » n'ont aucune légitimité dans leur soi-disant combat pour la défense des intérêts des pays africains. Ils ne sont ni élus, ni reconnus mais relèvent davantage de la catégorie des entrepreneurs politiques dont les discours sont opportunément relayés et soutenus par des intérêts de pays cherchant à affaiblir la France dans cette zone pour leur propre profit, sans aucune considération pour les populations locales.

Dans un second temps, s'organiser au niveau national, faire valoir nos valeurs et nos engagements

Faire face à la lutte informationnelle est l'un des objectifs stratégiques de la dernière revue nationale stratégique présentée le 9 novembre 2022 par le Président de la République. Il s'agit de faire évoluer l'organisation et les méthodes de travail au sein du ministère des Armées. Il s'agit également de disposer d'un large éventail d'options de réponses, qu'elles soient d'ordre normatif, de sanctions nationales ou européennes, voire d'actions militaires ou de poursuites judiciaires.

Exploitation du désir de confort La victoire de l'entre-soi et mise en danger de la démocratie



Dr. Jean-Philippe Riant

Directeur conseil
en intelligence artificielle
et communication digitale



« Si tu diffères de moi, mon frère, loin de me léser, tu m'enrichis. »

Antoine de Saint-Exupéry

Aurions-nous un penchant naturel pour le meilleur confort et pour une certaine prévisibilité de notre environnement ? Pour les entrepreneurs des réseaux sociaux, à coup sûr, c'est une évidence, et c'est un levier qu'ils développent à l'envi. Ils cherchent à nous construire des environnements favorables à leur modèle d'affaire : augmenter pour chaque utilisateur la quantité de temps de cerveau disponible afin de pouvoir cliquer sur toujours plus de contenus d'annonceurs payants. Les concepteurs d'univers digitaux à dynamique sociale ont bien compris comment utiliser ce penchant afin de créer des environnements qui nous sont agréables, mais qu'est-ce qu'un environnement numérique que nous considérons comme confortable ? Partons d'un univers de consommation bien réel : Starbucks a dès sa création mis des canapés dans ses points de vente pour que les clients aient envie de passer du temps à consommer des boissons à forte marge, avec autour d'eux des gens qui nous ressemblent. Facebook a créé le même univers de protection de façade, mais dans un environnement en ligne. La proximité de profil et de comportement de ceux qui nous entourent rend l'environnement prévisible et nous met en confiance. Nous ne risquons pas d'être déstabilisé ou de voir nos valeurs remises en question.

Le repli sur soi

Une soirée avec de vieux amis rejoue des schémas que nous connaissons bien et nous savons y prendre du plaisir et consommer du temps social. Un nouvel invité qui n'aurait pas nos codes déstabiliserait ce système humain et il est un risque. Alors pour créer du confort, les réseaux sociaux ont mécaniquement éliminé cette option de surprise : pas ou peu de risque de nouvel entrant qui pourrait nous surprendre ou nous poser des questions qui nous obligeraient à penser autrement, notamment avec des expressions d'opinions qui dérangent. Mais qu'est-ce qui a changé ? Qu'est-ce qui nous surprend dans cette situation qui crée les conditions de vagues de désinformation et diffusion de fake news. A quoi pouvons-nous donc nous raccrocher pour lutter efficacement ?

L'information tombée du ciel, quelles réactions ?

Comme en laboratoire, imaginons une situation en temps de guerre : une zone assiégée voyant tomber du ciel des milliers de tracts de propagande véhiculant des informations partiales ou fausses. Comment réagiraient les populations assiégées à la lecture de ces informations : un sentiment mêlé de crainte, mais surtout de doute. Chacun exercerait sûrement une forme d'esprit critique pour se demander si ce qu'il lit est vrai, conscient du danger et du caractère manipulateur de la méthode de communication

utilisée. Alors pourquoi nous semble-t-il que cet esprit critique a disparu ou n'est plus exercé par ces internautes sensibles à ces caisses de résonance de messages en ligne relayés en masse, formulons deux hypothèses. Première hypothèse : l'absence de prise de conscience du caractère international et menaçant de la situation nationale dans laquelle nous nous trouvons. Nos pays sont en guerre d'influence comme ils l'ont toujours été, mais avec une intensité inédite due à la mondialisation et aux nouveaux moyens de communication digitaux transnationaux (voire apatrides en apparence) et aux plateformes sociales mondiales. Deuxième hypothèse : les publics se sentent effectivement en guerre d'opinion, mais à des niveaux d'échelle plus réduits. Ce n'est plus la nation qui est le groupe d'appartenance majeure, mais la communauté, la tribu, le collectif de pensée qui se sent attaqué par d'autres, et surtout par la nation vue comme plus contraignante que structurante. Ce changement d'échelle ferait perdre la notion de collectif national et des champs de contraintes qui pèsent sur un pays et sur la déstabilisation de son fonctionnement pas des pays ennemis.

Esprit critique, résistance, curiosité, histoire commune ?

Alors que pouvons-nous croire ? Au-delà, de la technologie qui pourrait nous permettre d'atténuer la menace ou de

filtrer les contenus auxquels nous sommes exposés, il semble essentiel de prendre conscience au niveau national de la menace qui pèse sur notre équilibre social si nous nous replions sur des groupes trop petits qui ne voient que leurs contraintes propres. Sans être conservateur ou réactionnaire, il serait intéressant de s'interroger sur notre capacité individuelle à entrer en résistance face à la facilité de consommer ou de retransmettre des messages ou des informations qui ne vont que dans le sens de nos convictions. Mobilisons les politiques dans leur devoir d'exemplarité et de rassemblement, et motivons chacun à réagir face à un message social comme s'il recevait un tract tombé du ciel dans une ville assiégée. Si tout le monde avait toujours cru ce qui tombe du ciel, le monde libre le serait sûrement un peu moins. Alors il y a probablement un chantier de mobilisation générale de la population comme elle le ferait si elle se sentait attaquée comme nation. C'est un champ de communication délicat à adresser, nous l'avons vu quand le mot "guerre" a été utilisé pour décrire la lutte contre le covid au début de la pandémie et la nécessaire responsabilisation de tous, mais il semble qu'une prise de conscience que la défense de notre démocratie et de notre liberté de pensée passe par cette dynamique collective d'esprit critique et surtout de curiosité, même sur les réseaux sociaux.

Pour aller plus loin, accédez au contenu complémentaire en fin d'ouvrage ⁽¹⁾



Tracts largués par l'armée américaine au Vietnam en 1966. L'appareil est équipé de haut-parleurs pour diffuser des messages à l'attention des troupes ennemies

Enseignement supérieur et désinformation

quel bilan dans les écoles d'ingénieur ?



Marie Puren

Responsable Équipe MNSHS du
Laboratoire de Recherche (LRE)
de l'EPITA



La désinformation fait des ravages parmi les utilisateurs de réseaux sociaux. Mais touche-t-elle aussi les étudiants qui développent ces applications ? Nous avons cherché à le savoir en interrogeant des étudiants en école d'ingénieurs informatique

19 millions de bots ont twitté en faveur d'Hillary Clinton ou Donald Trump la semaine précédant l'élection américaine de 2016 ⁽¹⁾. Durant l'épidémie de COVID, 800.000 faux comptes comptabilisant 18 millions de vues ont diffusé des informations non vérifiées sur TikTok ⁽²⁾. Les plateformes ont alors misé sur l'intelligence artificielle et le deep learning pour développer des outils de détection automatisée permettant d'étiqueter, voire même supprimer, certains types d'informations ^{(3),(4),(5)}. Les concepteurs de ces solutions ont donc un rôle essentiel à jouer, car ce sont leurs choix qui détermineront les informations mises à l'index ou promues. Ce qui n'est pas sans danger, puisqu'on ferait en grande partie reposer la régulation de l'information sur leurs jugements de valeur ⁽⁶⁾. Il est donc crucial de former les ingénieurs de demain sur ces questions, car ils pèseront fortement sur les solutions adoptées par les médias sociaux, et plus généralement sur les politiques publiques de lutte contre les "désordres de l'information".

Les étudiants en informatique et les "désordres de l'information"

Nous avons interrogé 207 étudiants en informatique âgés de 17 à 24 ans afin d'évaluer leur attitude vis-à-vis des "désordres de l'information" - comprenant la désinformation, la mésinformation et l'information malveillante ⁽⁵⁾ -, et plus particulièrement leur volonté de

réguler ou non ces informations. Nous leur avons proposé huit affirmations provenant de l'actualité et représentant une de ces trois catégories. Les répondants devaient indiquer leur niveau de confiance dans l'affirmation proposée, puis s'ils auraient ou non autorisé sa diffusion en ligne.

55,4% des répondants préfèrent diffuser des informations qu'ils considèrent comme fausses plutôt que de les censurer ; mais ce n'est pas pour autant qu'ils rejettent toute régulation sur les réseaux sociaux.

Notre enquête fait apparaître trois groupes principaux :

- Un groupe **'libertaire'**, opposé à la régulation des réseaux sociaux, qui représente 42 % des étudiants. Il éprouve une forte défiance vis-à-vis des institutions, et témoigne d'une plus grande perméabilité aux théories complotistes. De manière très paradoxale, c'est le groupe qui s'informe le plus, et qui est le mieux capable d'identifier les fakes news.
- Un groupe **'conformiste'**, qui comprend 31% des étudiants ayant répondu à l'enquête. C'est celui avec la confiance la plus marquée dans les institutions, un souhait marqué de régulation des réseaux, et la plus grande volonté d'adapter ses opinions aux faits nouveaux. C'est aussi le groupe des étudiants les plus intuitifs dans l'analyse des informations auquel ils sont exposés.

- Un groupe, **'analytique'**, qui regroupe 27% des étudiants. C'est celui qui est le plus rationnel dans l'analyse des questions de logique, et le moins perméable aux thèses complotistes. C'est le groupe le plus favorable à la régulation des réseaux sociaux, mais aussi celui qui est le moins capable de reconnaître des fake news - peut-être parce qu'ils savent s'en protéger.

Les *Digital natives* sont perméables à la désinformation

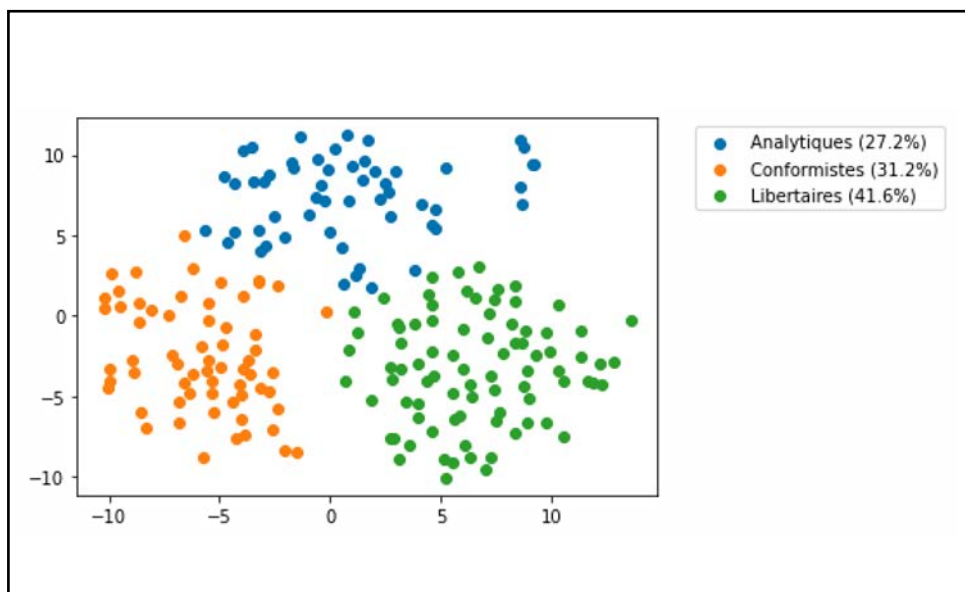
Fin 2022, une enquête IFOP menée auprès de jeunes de 11 à 24 ans a montré la perméabilité des *digital natives* à la désinformation et aux croyances irrationnelles ⁽⁷⁾. De plus en plus défiants vis-à-vis des médias traditionnels, les jeunes s'informent principalement via Internet et des réseaux sociaux gangrenés par des "fake news". Ainsi, parmi les étudiants que nous avons interrogés, seuls 11% d'entre eux font confiance aux médias. Cette défiance est un appel d'air pour les sources d'informations alternatives.

Mieux former les ingénieurs de demain

Nous pensons que l'éducation aux médias et à l'information⁽⁸⁾ et une bonne compréhension de la liberté d'expression⁽⁹⁾ sont des éléments clé pour développer l'esprit critique des étudiants. Il est donc essentiel que les enseignants dans les écoles d'informatique les accompagnent dans un travail de réflexion et de mise à distance critique des informations. La cartographie des controverses⁽¹⁰⁾ nous paraît être une approche fructueuse, car elle place les étudiants dans une position à la fois compréhensive et critique⁽¹¹⁾.

« Le développement des Soft Skills est l'une des clés de la formation des étudiants dans les écoles d'ingénieurs »

Ainsi les écoles d'ingénieurs intègrent de plus en plus les « Soft Skills » dans la formation des étudiants. En plus d'être des qualités appréciées en entreprise, apprendre à écouter l'autre, accepter le point de vue opposé et savoir se remettre en question sont également des compétences indispensables pour exercer pleinement son esprit critique. Le développement des Soft Skills est l'une des clés pour aider les étudiants à se mettre dans une posture compréhensive, et il doit accompagner l'éducation aux médias et l'analyse des sources d'information.



Article rédigé en collaboration avec :
[Younes BENREGUIEG](#),
étudiant à EPITA Paris

[PhD. Pierre PARREND](#),
Directeur adjoint du LRE
de l'EPITA, Responsable
de l'équipe Sécurité et
Systèmes du LRE

Aspects opérationnels et techniques

PARTIE



Contributeurs

-  Aurélie LAIZÉ
-  Stéphane PAQUELET
-  Eric LABOURÉ
-  Pierre BISCHETTI
-  Claire MABILLE
-  Général (2S) Bruno COURTOIS
-  Thomas DELAVALLADE
-  Dr. Marc-Oliver PAHL



Moyens technologiques actuels pour lutter contre les manipulations d'informations



Aurélie LAIZÉ

Head of Airbus Cyber
Programmes site Rennes



“La maîtrise du champ informationnel numérique requiert le maintien de connaissances au meilleur état de l’art dans les domaines techniques associés. La capacité à veiller des réseaux, à détecter des contenus et à analyser un environnement est liée à des outils spécifiques en constante évolution, utilisant les technologies de traitement des informations en masse (big data) et d’intelligence artificielle.”

ÉLÉMENTS PUBLICS DE DOCTRINE MILITAIRE DE LUTTE INFORMATIQUE D’INFLUENCE (L2I)

L’espace informationnel est devenu un champ de bataille, visant tant des objectifs tactiques tels que l’obtention de renseignements ou la restriction de liberté de manœuvre, que des objectifs stratégiques tels que la légitimité d’un déploiement militaire, voire même la légitimité de la démocratie. Les manœuvres informationnelles adverses sont aujourd’hui largement facilitées par l’essor des plateformes de diffusion de contenus (réseaux sociaux) et par le développement phénoménal des technologies à base d’Intelligence Artificielle, notamment les Deep Fakes. L’analyse humaine, malgré sa grande pertinence, ne suffit donc plus à lutter contre la prolifération massive d’informations manipulées. Le recours à des moyens technologiques en support des analyses humaines semble dès lors indispensable à la détection et l’identification des manœuvres informationnelles adverses.

Captation et analyse des faits

Une première réponse technologique mise en place consiste à analyser les faits et leur véracité. La captation de rumeurs dès leur émission requiert l’orientation de capteurs à tous les niveaux, dans la presse mais aussi sur les forums et réseaux sociaux. Les instituts journalistiques se sont pour la plupart dotés d’équipes

de fact-checking (tel AFP factuel⁽¹⁾). Ce travail manuel, minutieux, s’appuie notamment sur des technologies de recherche d’information (OSINT)⁽²⁾, d’indexation de faits⁽³⁾, de requête auprès de bases de données statistiques mais aussi de vérification de l’existence préalable d’une image⁽⁴⁾. Sur l’aspect purement factuel, des projets de recherche s’appuient sur l’extraction d’informations et la comparaison automatique avec des connaissances établies pour alerter en “temps réel” d’une in vraisemblance avérée par rapport à un référentiel.

Analyses et identification du faux

Une seconde réponse est l’étude automatisée d’artefacts exhibés par les contenus : la sophistication des faux contenus, largement popularisés par la production de mêmes humoristiques impliquant des personnalités publiques, appelle un recours renforcé aux technologies de traitement d’images (reconnaissance d’objets, analyse de similarité, détection de faux⁽⁵⁾), d’analyse vocale ou de traitement du langage naturel (détection d’émotions, identification de narratifs, traduction). Ces moyens s’appuient largement sur les réseaux de neurones profonds⁽⁶⁾ dont l’usage nécessite d’importantes capacités de calcul, ce qui en limite l’adoption à grande échelle.

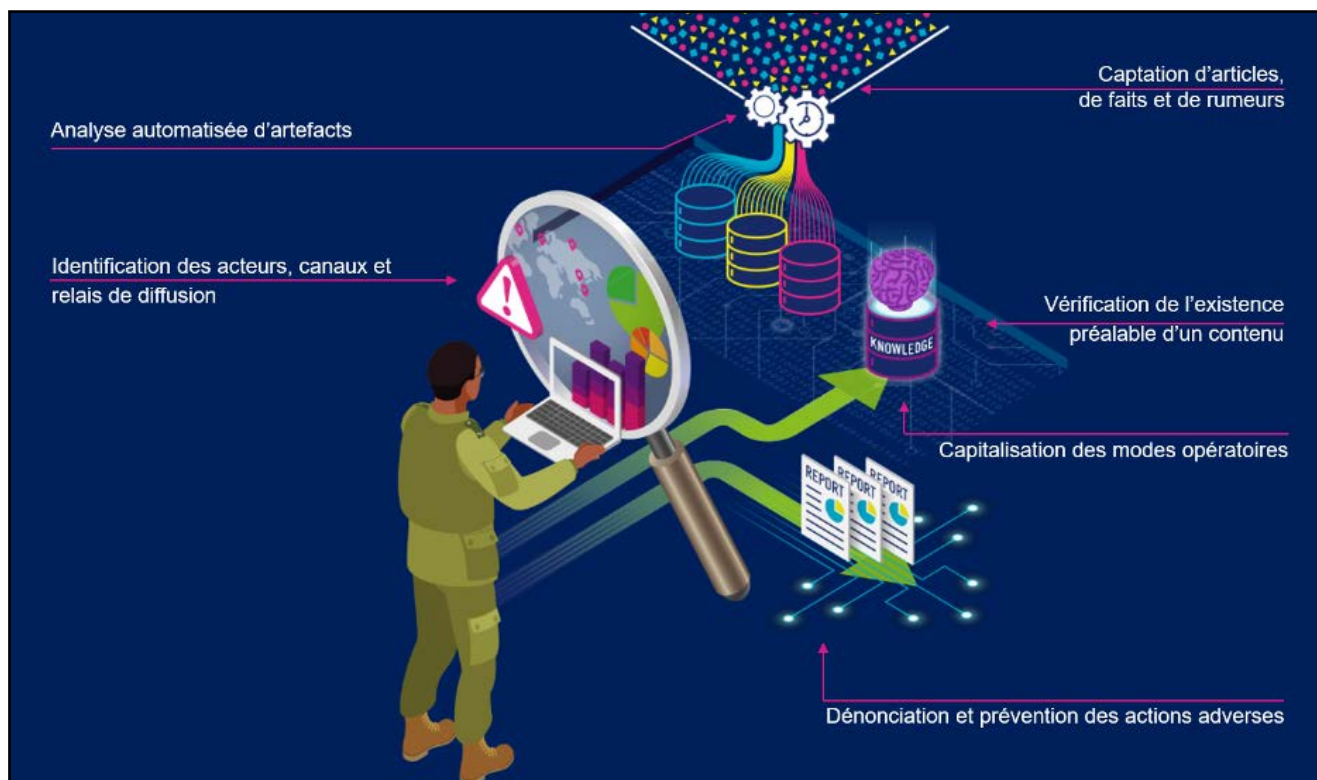


Figure : Les moyens de lutte contre les manipulations d'informations

Identification des acteurs

Outre les faits et les contenus truqués, l'identification des acteurs, canaux et relais de diffusion est un enjeu majeur qui repose entre autres sur la caractérisation des sources (typosquatting de domaines, détection d'avatars/bots) et des narratifs (vérité et origine des contenus, détection de chambres d'écho). En particulier, l'exploitation des liens et leur historisation, la reconnaissance des auteurs, l'étiquetage de sites comme étant "de confiance" ou "douteux" permettent de se faire une idée a priori de la crédibilité de chaque source, et de remettre celle-ci en question au cours du temps.

Capitalisation des modes opératoires

Enfin, la capitalisation et l'agrégation de ces éléments et des modes opératoires associés constituent le socle indispensable pour relever le défi de la désinformation et permettre la réponse coordonnée à des campagnes informationnelles de plus en plus organisées⁽⁷⁾. La création de faux comptes, leur niveau de robotisation et de sophistication, le recours à des narratifs pré-existants sont autant de points d'accroche pour extraire du renseignement élaboré, améliorer les détecteurs existants et accompagner des analystes dans leur mission.

Prévenir et réagir

En allant plus loin, la lutte contre la manipulation d'information se poursuit activement par une action adéquate contre l'adversaire dont la dénonciation et le bannissement de comptes, notamment par le renforcement de la coopération avec les équipes de modération. Plus largement, la prévention des actions adverses nécessite de nombreuses actions d'amélioration de la résilience collective dont l'éducation aux médias et à l'esprit critique d'une myriade d'acteurs au travers de sensibilisations et à l'aide de plateformes de formation et d'entraînement dédiées.

Article rédigé en collaboration avec :
[Guillaume Gadek](#)
 et
[Benjamin Coste](#)
 chercheurs en désinformation et cybersécurité chez Airbus Defence and Space

Manipulation de l'information

La technologie et sa nécessaire régulation



Stéphane PAQUELET
Responsable du laboratoire
Intelligence Artificielle
b<>com



Les progrès accomplis par l'Intelligence Artificielle (IA) sont désormais tels qu'il est possible de produire des contenus audio-visuels et textuels d'un réalisme stupéfiant à l'échelle industrielle. Appuyés par des moyens de diffusion planétaires, ils nous font entrer dans un monde de l'information où l'art de duper et celui de confondre promettent de régner en maître. Dès lors, nous proposons une analyse prospective des enjeux technologiques induits par ces progrès, ainsi que des relations qu'ils entretiennent avec les problématiques juridiques et éthiques.

L'IA aux deux visages : génération et analyse

Pour comprendre les évolutions en cours, il est nécessaire de distinguer deux modes de fonctionnement de l'IA dans le cadre du traitement de données. Commençons par l'IA *générative*, qui a connu d'immenses avancées lors de la dernière décennie pour des données de nature variée. Avec l'avènement de la technologie Generative Adversarial Networks (GAN), l'IA fournit des modèles réalistes pour produire des contenus audio-visuels (StyleGAN). Parallèlement les dernières méthodes de représentation de données symboliques, les réseaux Transformers, ont bénéficié au traitement du langage avec la production textuelle automatique comme les chatbots (technologie GPT3, chatGPT). Les deux modalités peuvent être désormais combinées pour créer des contenus artistiques à partir d'une description sommaire (Dall-e). L'une des applications dérivée est la manipulation de contenu, aussi appelée hypertrucage (Deepfake), dont l'objectif est de travestir les attitudes ou propos d'une personne. Poursuivons par l'IA *analyse* dont l'utilisation est antérieure,

notamment par sa capacité à détecter ou classifier des formes de nature arbitraire mais également des contenus textuels. La détection de contenus malveillants (pornographie, propos haineux) ou la modération sur les réseaux sociaux sont rendus possible grâce à ces méthodes. Nous disposons donc de technologies aux propriétés ambivalentes, qui montrent que la lutte contre les manipulations de l'information est une question fondée. Pour poser l'ensemble des problèmes sous-jacents d'une manière adéquate, examinons maintenant les limites de ces technologies. Le point de vue considéré sera celui de la détection de données manipulées et de contenus indésirables.

Les défauts de la cuirasse

Les créations audio-visuelles de l'IA *générative* présentent encore aujourd'hui des artefacts, liés par exemple à l'absence d'optique de mesure ou aux retouches sur un visage. Invisibles à l'œil nu, ces artefacts peuvent néanmoins passer au crible de l'IA *analyse* qui propose des méthodes de représentations abstraites capables

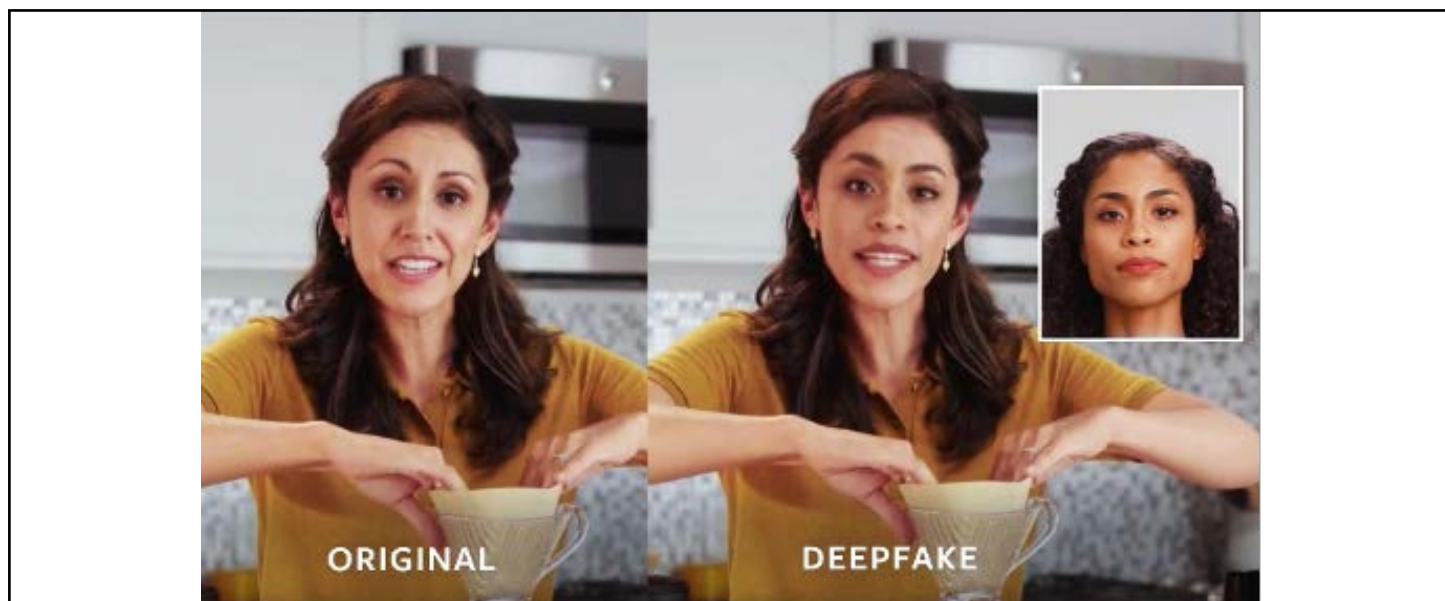


Figure 1. Exemple de Deepfake. World Economic Forum Annual Meeting

d'amplifier ces défauts et de les détecter. Cependant, cette détection s'accompagne toujours d'un taux d'erreur statistique résiduel, qui intègre les faux positifs et négatifs. Ainsi, la conception des détecteurs de Deepfake repose sur une compréhension approfondie des générateurs associés. Rétroactivement, l'amélioration des premiers proposera de nouveaux défis aux seconds, structurant le domaine selon un paradigme adversaire générateur-détecteur. L'horizon ultime de cette course-poursuite sera l'indiscernabilité entre des contenus authentiques et hypertruqués. Dans cette nouvelle configuration, la vraisemblance d'une source ne pourra plus être appréciée à partir de son examen brut, quel que soit le degré de sophistication de l'outil d'analyse. Il sera nécessaire d'imaginer des moyens additionnels susceptibles de certifier le mode de production de la source diffusée lorsqu'il s'agit d'une captation.

Les productions textuelles automatiques ne présentent pas, quant à elles, d'artefacts de conception, si bien que le problème de manipulation se situe d'emblée au stade ultime exposé pour l'hypertrucage. En toute généralité, la détection de rumeurs, ou plus largement de contenus mensongers, nécessite de confronter l'élément suspecté à un contexte, en général vaste et dispersé. De tels procédés d'investigation en sont aujourd'hui à un stade encore balbutiant. Ils soulèvent des problématiques fondamentales en traitement automatique du langage, en voici deux : la reconnaissance d'entités nommées relativement à une base de connaissance structurée (ontologie) et la production automatique de telles ontologies. Supposons maintenant résolu le problème de la modération automatique. Il n'est pas restrictif pour cela de considérer les robots actuellement utilisés sur les

plateformes numériques. Le principe de modération recèle deux arbitrages : d'une part les critères retenus pour la modération (le contenu de l'ontologie de référence) et d'autre part l'erreur inhérente à l'algorithme qui pourrait abusivement censurer un contenu pourtant licite (les faux positifs). Ces questions connectent naturellement la technologie à des enjeux juridiques et éthiques.

En conclusion

La course technologique entre les manipulateurs d'information et leur censeur est engagée. Nul doute qu'elle annonce des percées scientifiques majeures. Mais en faire une lecture strictement technologique est une impasse car cette problématique cristallise des enjeux sociétaux de premier ordre comme l'avènement d'une société que George Orwell lui-même n'aurait imaginé. La lutte contre les manipulations de l'information est promise à remodeler en profondeur le domaine de la cyber-sécurité. A un stade naissant, elle est donc une opportunité pour structurer une nouvelle filière industrielle en y associant les débats réglementaires et sociétaux nécessaires.

Authenticité de l'information

le rôle essentiel des opérateurs



Eric LABOURÉ

Head of Business Operations
Managed Security Services
NOKIA



Au cours des trente dernières années, les techniques de partage de l'information ont drastiquement évolué. Non seulement vers le digital, mais aussi en passant de fournisseurs centralisés (agences de presse, journaux, télévision..) à une multitude d'acteurs individuels via les réseaux sociaux. Cette révolution est largement due à l'évolution du web et des applications mobiles, mais rien n'eût été possible sans les opérateurs de télécommunication grâce au déploiement et à l'amélioration constante des réseaux fixes et en particulier de la téléphonie mobile.

Une chaîne d'information complexe

Dorénavant, l'interrogation porte sur le sérieux et l'honnêteté des fournisseurs de contenus, mais entre le fournisseur et le lecteur existe toute une chaîne de transmission de l'information. Qu'il s'agisse de texte, de son, d'image ou de vidéo, le contenu doit passer par un grand nombre de relais et transformations intermédiaires. Les opérateurs doivent s'assurer non seulement de la disponibilité des informations, mais aussi de l'intégrité et de la confidentialité, tout en sécurisant l'identification des individus sources d'information. A notre époque, loin d'être de simples « tuyaux » transférant des contenus, les opérateurs et les équipementiers télécom sont au contraire des acteurs clefs de la confiance numérique. Dans le cas des agrégateurs d'informations (journaux en ligne, blogs...), l'objectif est de s'assurer que le lecteur obtient ses informations depuis la source escomptée sans être modifiées jusqu'au téléphone du lecteur. De ce point de vue, plusieurs mécanismes informatiques se sont très répandus, tels que les protocoles SSL/HTTPS, et garantissent l'identité auprès de l'équipement de l'utilisateur ainsi que l'intégrité des données de la source jusqu'au lecteur à travers le réseau

de l'opérateur. Toutefois, avoir confiance en un flux d'informations nécessite que l'identité du rédacteur ait été sécurisée dès le début de la création de contenu.

Une sécurité renforcée

Beaucoup de rédacteurs se reposent désormais sur les réseaux mobiles pour s'authentifier dans leur application favorite. Actuellement, chaque équipement d'un réseau mobile doit s'authentifier au sein du réseau pour instaurer une confiance entre « voisins » et éviter les attaques de type « man-in-the-middle » (MitM) ou d'usurpation. A l'époque du GSM, en raison de faiblesses de design dans les premières versions du standard 3GPP de téléphonie mobile, les réseaux 2G (GSM) et leur protocole SS7 étaient insuffisamment sécurisés et sujets à des attaques d'usurpation sur les éléments de réseau, pouvant entraîner des attaques de type MitM avec interception de SMS et même la redirection des SMS vers l'appareil de l'attaquant. D'où un risque sur l'authentification à double facteur permettant à l'attaquant un accès aux applications d'information et donc de potentiellement modifier les contenus. Les technologies 4G LTE et 5G ont donc introduit une authentification très poussée.

Soulignons que tant que la 2G reste active sur les réseaux, elle impose certains mécanismes moins sécurisés aux téléphones utilisant la 2G. De même, le protocole Diameter, toujours actif en 4G LTE et 5G NSA, peut ouvrir certaines failles. Heureusement, des solutions existent pour se protéger, tels que les pare-feux SS7 ou les infrastructures PKI : il incombe aux opérateurs leur mise en place ainsi qu'une configuration adéquate de leur réseau.

Authentifier voire localiser l'utilisateur

Côté usager, grâce au chiffrement et aux clefs des cartes SIM, les téléphones doivent s'authentifier avec le réseau mobile, instaurant la confiance envers une information provenant directement d'une personne du même opérateur, comme lors d'un appel téléphonique. Depuis la 3G, l'authentification est même réciproque, obligeant le réseau de l'opérateur à prouver son identité auprès du téléphone. La carte SIM, et dorénavant la eSIM intégrée au hardware, reste la pierre angulaire de l'authentification des usagers. Bien sûr, ce mécanisme ne permet pas de certifier que la bonne personne utilise cet appareil en cas de vol du téléphone ou de la carte SIM. Certaines attaques comme le SIMjacking ou le SIMcloning existent, mais les opérateurs occidentaux mènent des contrôles sur les cartes SIM, sur les comportements frauduleux et sur les adhésions. Cette sécurité est clairement cruciale pour le partage d'information et les réseaux sociaux qui doivent maintenir cette confiance dans l'origine des informations. Au-delà d'identifier une source, la localiser est également très utile, pour accroître la confiance dans la source ou déterminer l'origine du contenu. Les réseaux mobiles peuvent jouer un rôle important dans cette perspective en fournissant une preuve de localisation. Même si celle-ci n'est pas accessible aux abonnés, les autorités peuvent requérir des opérateurs les détails de géolocalisation. A l'heure de « fake news », des images et vidéos factices, et de l'altération des données GPS d'un fichier, une telle possibilité s'intègre dans le processus de validation des sources d'information.

Maîtriser les configurations

Les opérateurs deviennent donc une pierre angulaire de la transmission d'information et les standards 3GPP de téléphonie mobile aident à identifier, formaliser et standardiser les mesures de sécurité, améliorant sans cesse le degré de sécurité de la 2G à la 5G. Les opérateurs doivent cependant être très vigilants sur les configurations des équipements puisque les standards émettent souvent des recommandations et non des obligations. De mauvais paramétrages mettent en péril la sécurité d'un réseau, avec de potentielles fuites ou modifications de données. Les équipementiers télécom ont bien compris cette menace et ont d'ailleurs développé

des services de sécurisation adaptés aux environnements télécom pour aider opérateurs et entreprises utilisatrices de tels réseaux à définir leur stratégie de cybersécurité et les configurations dans un souci d'équilibre entre sécurité et performance. Par ailleurs, les opérateurs vont héberger de plus en plus d'applications B2B et B2B2C sur leurs réseaux 5G : si ces applications portent sur la création ou le partage d'information, les opérateurs seront de facto partie prenante dans la sécurisation de ces applications et non plus seulement dans la transmission sécurisée de données à travers leur réseau.

D'autres challenges à venir

Une nouvelle menace plane sur la sécurité des données : les ordinateurs quantiques pourraient compromettre l'authentification et le chiffrement par les algorithmes à clefs actuels. La maturité est encore insuffisante pour être une menace imminente, mais les investissements augmentent et certains états semblent s'y lancer massivement. Heureusement, quelques algorithmes sont toujours considérés comme résistants au calcul quantique. Bien que non finalisés et standardisés, le NIST a ainsi publié en 2022 un algorithme « quantum resistant » pour le chiffrement de données et trois autres pour les signatures digitales. Equipementiers télécom et opérateurs devront s'atteler à de grands chantiers : assurer la disponibilité de cryptographie post-quantique dans les produits, la compatibilité des architectures, des procédures et des nouveaux paramétrages qui seront à nouveau une source de faiblesses et des plans de migration à l'échelle des réseaux. Tout cela pour conserver nos informations à l'abri de l'espionnage ou de l'altération.



Comment déceler le manipulateur derrière l'avatar



Pierre BISCHETTI

Analyste OSINT
OWLINT



Dans une logique de guerre de l'information, les faux comptes s'en prennent aujourd'hui aux entreprises comme aux États. La détection de ces réseaux d'influence et des acteurs qui les utilisent passe — entre autres — par leur maillage relationnel, l'identification des éléments de langage et la recherche de failles de sécurité opérationnelle (SecOps). Comme le montre le conflit ukrainien, un certain nombre d'outils open sources sont à la disposition de tout un chacun et renforcent les capacités des analystes, amateurs ou non.

Éléments de langage distincts

Aujourd'hui, des solutions variées se positionnent sur la construction automatique d'éléments de langage ou sur la détection d'actions d'influence par l'étude du narratif. Partant du postulat qu'un individu est cohérent, il est probable qu'il utilise le même langage sur le web en général. La détection d'un avatar — créé à fin d'influence — passe alors en premier lieu par la compréhension de ses habitudes syntaxiques. Elles peuvent permettre de trouver des similitudes avec d'autres comptes ou d'identifier les traces d'opérations passées. Elles montreraient alors un « recyclage » de l'avatar : p.ex. tantôt actif sur la politique africaine, tantôt engagé auprès d'une entreprise européenne. Prenons un exemple : le maire d'une ville est dénigré sur Twitter. L'avatar, au-delà des insultes à caractère xénophobe, utilise plusieurs formules telles que : « la maire qui se croit la mère (...) » ou « la reine maire et sa cour ». On peut ainsi supposer que ces tournures de phrases sont utilisées ailleurs, notamment sur les

comptes personnels de l'individu qui utilise l'avatar. Dans ce cas précis, il s'avère qu'elles faisaient écho à un blog tenu par un habitant de la ville concernée. On obtient alors un pivot permettant de continuer l'enquête.

Dans le cadre d'analyses textuelles longues - p.ex. des déclarations d'intentions — l'utilisation d'outils de textométrie permet d'aller plus loin dans l'étude des habitudes syntaxiques d'une identité numérique. Si cette première étape peut s'avérer rébarbative, elle est indispensable dans le cadre d'opération d'influence complexe et représente ici une belle expression des compétences de l'analyste.

Canaux de diffusion et failles de SECOPS

Au-delà de l'analyse du discours, lorsqu'un contenu préjudiciable est détecté, il est primordial d'étudier qui en sont les diffuseurs, véritables caisses de résonance. Selon la viralité du contenu, il peut être difficile d'en connaître la source !

La chronologie se trouve être ici la meilleure alliée de l'analyste. Les dates de publications permettent de rebondir vers l'origine d'une rumeur, mais également d'analyser l'évolution des comptes en question (confère supra). Aujourd'hui, dépendant des réseaux sociaux étudiés, de nombreux outils permettent de réaliser une première étude, comparée ou non, du contenu de plusieurs comptes. C'est — encore — particulièrement vrai sur Twitter. Dès les premiers relais identifiés, il s'agira d'étudier chacun d'eux afin d'isoler ceux qui sont générateurs de rebonds. Il est tout à fait possible que l'un des relais ait effectué une faute de sécurité opérationnelle dans ses démarches (e-mail d'enregistrement visible, comptes publics sur les réseaux sociaux...) ou que l'on retrouve des récurrences dans les heures d'activités, le type de support utilisé, la date de création... On peut alors espérer retrouver des localisations, des numéros de téléphone ou une liste publique d'amis.

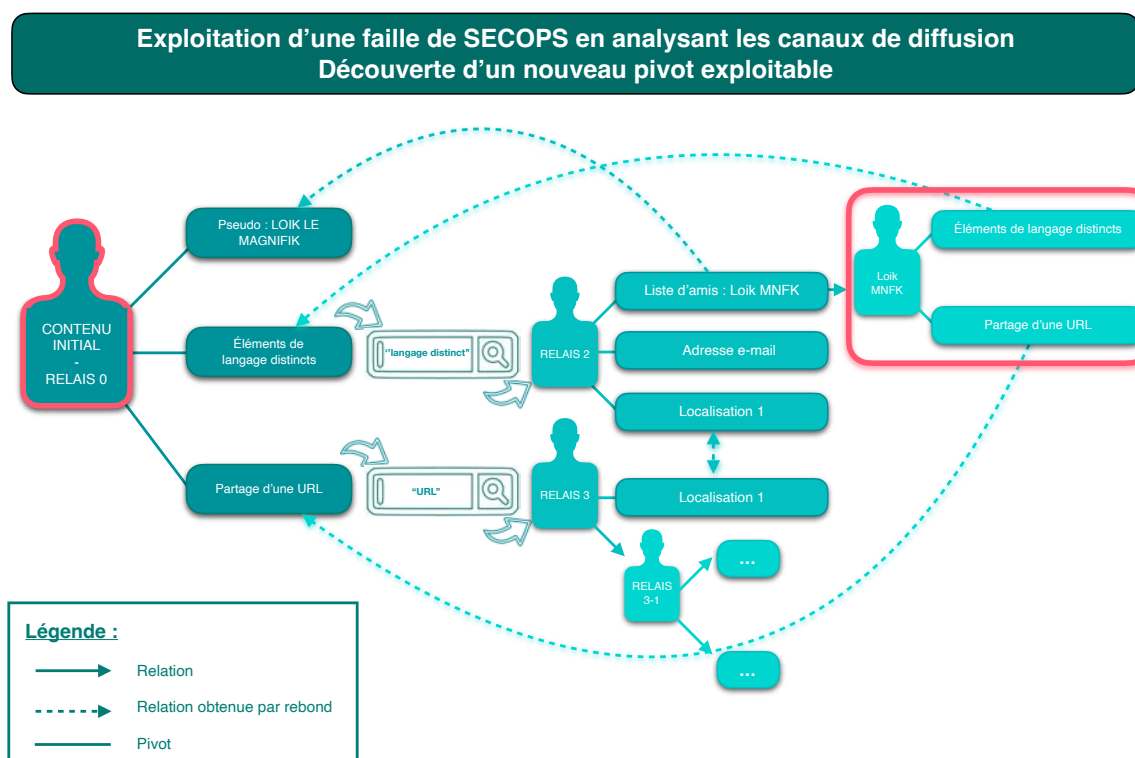
Outils et enjeux

Les outils OSINT disponibles évoluent, certains surfent sur l'accessibilité d'une information, la rapidité d'un script, puis disparaissent. Cependant, il existe certains modules qui font aujourd'hui office de standard dans la communauté par la qualité de leurs développements et leurs constantes mises à jour. La majorité d'entre eux automatisent des tâches rébarbatives comme la recherche de pseudonymes

similaires en forgeant à la volée des centaines d'URL. D'autres s'intéressent à la captation d'informations, possible via des sélecteurs techniques précis comme les éléments d'une adresse en Gmail, dans l'objectif de générer du pivot. Si ces outils sont souvent le résultat d'initiative individuelle, sur certain aspect, le monde universitaire n'est pas en reste. On notera par exemple l'outil open source Hoaxy, originaire d'Inde, qui permet de mettre en lumière la probabilité qu'un utilisateur Twitter soit un avatar.

Cependant l'acculturation du monde universitaire à ces sujets est particulièrement inégale et l'OSINT reste cantonnée aux formations Cyber et à leurs CTF, là où le besoin d'appropriation des enjeux du numérique touche, au-delà des avatars, l'ensemble des professions. L'évolution des réseaux sociaux pousse les analystes à une constante réinvention. Au regard de la dynamique générale et du besoin opérationnel, il est probable que nous assistions à une transformation du secteur avec des avatars de meilleure qualité et, par conséquent, de plus grandes difficultés à les déceler uniquement manuellement. L'OSINT et sa boîte à outils ont donc encore de beaux jours devant eux !

Article rédigé en collaboration avec le Pôle services d'Owlint



Les Bots sociaux et la manipulation de l'information



Claire MABILLE

Analyste chez Sahar
Étudiante en Master 2
Cyber-stratégie et terrain numérique



Alors que les campagnes de manipulation de l'information en ligne se multiplient, la détection des bots sociaux, qui jouent un rôle central au sein de ces campagnes, devient un enjeu d'une importance croissante.

Le terme de « bot », dérivé du mot « robot », désigne les entités partiellement ou entièrement automatisées qui fonctionnent sur des plateformes en ligne. Cette définition générale recouvre une très grande diversité de technologies, qui ont été conçues différemment (par exemple, les bots peuvent être entièrement ou partiellement automatisés), fonctionnent différemment (certains recueillent de la donnée, d'autres interagissent avec les utilisateurs) et servent différents usages (privé, commercial ou politique). On inclut sous le terme de bot aussi bien le chabot que le gaming bot, les web robots (crawlers et scrapers), les spam bots, les bots sociaux (social bots), les cyborgs, les sock puppets et les trolls⁽¹⁾. Les bots sociaux, qui ont vocation à imiter des utilisateurs authentiques afin de manipuler l'opinion publique sur les plateformes et les réseaux, sont l'un des outils mobilisés par les acteurs politiques dans le cadre d'opérations d'influence en ligne. Le cas le plus connu est leur utilisation par la Russie en 2016, dans le cadre d'une campagne d'influence sur les élections américaines. Poussés par les enjeux démocratiques sous-jacents, tant les chercheurs en sciences sociales qu'en science de l'informatique se sont intéressés aux méthodes de détection des bots sociaux. Twitter, la plateforme sur laquelle ce phénomène est le plus flagrant, concentre les efforts de recherche. Cependant, en fonction des acteurs, des

campagnes et des objectifs visés, le comportement des bots sociaux varie. Aussi, des stratégies diverses, reposant sur différents types de critères, ont été élaborées.

Les approches de détection des bots sociaux

Les chercheurs en sciences sociales et en OSINT spécialisés dans la manipulation de l'information s'appuient généralement sur un ou plusieurs critères pour identifier les bots. Certains catégorisent comme des bots les utilisateurs qui produisent plus de 50 publications par jour⁽²⁾. D'autres s'appuient sur la similarité du contenu des publications et des noms de comptes pour affirmer qu'un groupe de comptes est constitué de bots⁽³⁾. Le caractère inauthentique des photos de profils des comptes (copiées ou générées par des GAN - generative adversarial networks) est également mobilisé. Enfin, les plateformes spécialisées dans la détection de bots, comme Botometer, sont employées pour tester le « score de bot » des comptes identifiés comme suspicieux. Ces approches souffrent néanmoins d'importantes limites : lorsque les analyses sont réalisées manuellement, elles gagnent en fiabilité mais perdent en exhaustivité ; inversement, lorsqu'un critère unique est mobilisé, elles gagnent en rapidité mais deviennent insensibles à la diversité des bots sociaux. Du côté des chercheurs en science de l'informatique,

les méthodes de détection automatisée des bots sociaux s'appuient généralement sur une ou plusieurs dimensions d'un compte⁽²⁾, parmi lesquelles : ses attributs (métadonnées, fréquence de publication, etc.), le contenu textuel qu'il produit, et son graphe relationnel (le réseau de following et followers dans lequel il s'inscrit). Ces dimensions relèvent de différentes approches, et font appel à différentes techniques. Tandis que les premières dimensions relèvent d'une approche individualisée (« account-based ») de la détection de bots, la dernière s'inscrit dans une approche relationnelle (« group-based »). Par ailleurs, si toutes ces dimensions requièrent de faire appel à des algorithmes de classification, l'identification des bots grâce aux attributs des comptes passe souvent par des algorithmes de machine learning comme Random Forest, tandis que l'identification des bots grâce aux publications des comptes passe souvent par des modèles de deep learning comme BERT (un modèle de NLP - Natural Language Processing)

Avantages et inconvénients des méthodes automatisées

Les méthodes automatisées sont très performantes en matière de détection des bots. L'évaluation réalisée sur le plus grand dataset de détection de bots issus de Twitter⁽²⁾ fait ressortir certaines d'entre elles en particulier. Parmi les méthodes qui s'appuient sur les attributs des comptes, SGBot⁽³⁾, qui utilise les métadonnées numériques et binaires des comptes (fréquence de publication, etc.), et DeeProBot⁽⁴⁾, qui exploite la description textuelle des utilisateurs en plus de ces métadonnées, atteignent une précision qui avoisine 75 %. Les méthodes qui s'appuient uniquement sur le contenu textuel des comptes, en capitalisant sur RoBERTa⁽⁵⁾ ou T5⁽⁶⁾, mobilisent des techniques de NLP pour analyser les publications et les descriptions (« bios ») des utilisateurs, et atteignent des scores de précision supérieurs à 70 %. L'étude de Kouvela et al.⁽⁷⁾ mêle ces deux dimensions en utilisant à la fois trente attributs des comptes et leurs vingt derniers tweets pour entraîner un algorithme de classification Random Forest. BotRGCN⁽⁸⁾, la méthode de détection de bots la plus performante (avec un score de précision de 79,7 %), combine les deux approches en y ajoutant une approche basée sur des graphes : elle mobilise les attributs des comptes, le contenu textuel, et génère un graphe relationnel des comptes regroupant les informations précédentes pour chaque utilisateur. Ces approches présentent néanmoins des inconvénients. Les méthodes qui s'appuient sur des graphes requièrent des ressources computationnelles élevées, et ont un temps d'exécution plus long que les modèles de détection fondés sur les attributs des comptes, ce qui rend leur utilisation coûteuse sur de très grands datasets. Les modèles qui s'appuient uniquement sur les attributs ou le contenu textuel des comptes manquent une partie importante des indicateurs disponibles, et sont très dépendants des

données d'entraînement. Enfin, la complexité de ces modèles de détection et la faible explicabilité de leurs résultats (en particulier lorsqu'ils s'appuient sur des algorithmes de deep learning) les rendent peu accessibles à des chercheurs non-formés à la science de l'informatique.

Conclusion

Il est nécessaire de construire des outils permettant aux chercheurs spécialistes du « terrain numérique » d'identifier des bots sociaux au sein de bases de données massives. Du côté des analystes, un travail d'adaptation aux outils automatisés doit être réalisé ; en parallèle, un travail d'explicabilité et d'accessibilité mérite d'être conduit par les concepteurs des outils afin d'assurer une synergie entre les deux domaines.

Article rédigé en
collaboration avec :
[Clément Apavou](#)
Data Scientist, SAHAR

[Gauthier Schweitzer](#)
Directeur Général, SAHAR

La lutte informatique d'influence



**Général (2S)
Bruno COURTOIS**

Conseiller Défense & Cyber
SOPRA STERIA



la lutte informatique d'influence représente un atout majeur pour faire face aux menaces croissantes dans le champ informationnel via les réseaux sociaux, dont l'utilisation fait désormais partie du quotidien des populations.

Une évolution spectaculaire

Le cyberspace, source de nombreux espoirs et vecteur de progrès pour l'humanité, constitue aussi un des talons d'Achille des organisations internationales comme des États, notamment ceux du monde démocratique occidental. En effet, inexistant au début du 21^e siècle, les réseaux sociaux ont bouleversé depuis les structures du pouvoir médiatique, en réduisant considérablement les barrières d'accès à la médiatisation. Ils sont ainsi devenus à la fois outils de campagne d'information, de recueil de renseignement, de recrutement, mais aussi de désinformation, de propagande, voire de levée de fonds dans certains cas. Leur rôle est considérable et, en dépit des efforts de protection accomplis récemment, permet à des acteurs mal intentionnés de manipuler les populations qui n'ont plus confiance dans les médias officiels et s'informent via les réseaux sociaux. Ce phénomène est accru par la transformation numérique, en cours partout dans le monde occidental avec ses effets bénéfiques spectaculaires, y compris dans le domaine de la Défense, où elle représente un multiplicateur d'efficacité militaire. Elle offre, entre autres, de nouvelles capacités de traitement automatique et de diffusion rapide de l'information, qui permettent de détecter et traiter instantanément des cibles, à l'image des affrontements en Ukraine depuis le 24 février 2022. A l'inverse, cette

numérisation rend aussi vulnérable aux attaques informatiques ou à la manipulation de l'information, deux composantes d'une nouvelle forme de guerre, dite « guerre hybride ». L'acquisition de la supériorité informationnelle demeure en conséquence un objectif majeur de tout type de conflit, y compris de haute intensité.

Un mode d'action en croissance

La France a saisi l'importance de la maîtrise de ce champ informationnel, notamment avec la loi promulguée en 2018 contre les manipulations de l'information en période électorale puis en 2021, avec la doctrine interarmées sur la lutte informatique d'influence (L2I), présentée par la ministre des Armées de l'époque.

La L2I désigne l'ensemble des opérations militaires conduites dans la couche informationnelle du cyberspace, pour détecter, caractériser et contrer les attaques, appuyer la communication stratégique, renseigner ou conduire une manœuvre de déception, de façon autonome ou en combinaison avec d'autres opérations.

En France, le général chef d'état-major des armées exerce le commandement des opérations militaires. Pour réaliser les opérations de lutte informatique d'influence, il s'appuie sur l'officier général commandant

de la cyberdéfense et sur des unités spécialisées. En effet, les attaques informationnelles sont fréquentes en opérations. La France y a été confrontée en zone sahélienne, avec pour objectif d'accroître le sentiment anti-français et de précipiter leur départ ; ainsi au Mali, avec des photos truquées accusant la France de piller les richesses aurifères de ce pays ou au Burkina Faso fin novembre 2021, à l'encontre d'un convoi logistique de Barkhane bloqué par la population locale. Elle avait été convaincue, via les réseaux sociaux, que ce convoi transportait des armes destinées aux djihadistes. Face à ces actes hostiles, l'anticipation, associée à une capacité de détection sont indispensables. Elles permettent une réaction dans des délais adaptés, idéalement en s'appuyant sur des images-preuve, comme celles des drones français filmant en avril 2022 l'enterrement à la hâte de cadavres à proximité de la base de GOSSI, récemment rendue par la Force Barkhane, pour faire croire à un massacre mené par les Français avant leur départ.

L'adaptation au contexte représente également une contrainte majeure. Au Moyen-Orient et en dépit de moyens très importants, les Etats-Unis ont obtenu des résultats mitigés de leurs campagnes informationnelles d'Irak ou d'Afghanistan jusqu'en 2020. En effet, les populations visées, peu cultivées et peu numérisées, se montrèrent peu réceptives aux messages diffusés. A l'inverse, « l'insurgé innovant » adverse fit preuve de réactivité, en diffusant régulièrement des images d'embuscades ou d'accrochages victorieux sous très courts préavis, suscitant inquiétude ou fierté.

De son côté, Israël fut confronté en 2006 au Sud-Liban à un Hezbollah résolu en matière d'attaque informationnelle, qui sut notamment exploiter habilement sa frappe par un missile de la corvette israélienne « Hanit » au large du Liban.

L'appui industriel et partenarial

Plusieurs industriels mobilisent leurs ressources pour appuyer les armées dans ce domaine, qui impose la maîtrise des savoir-faire techniques suivants :

- en liaison avec le monde du renseignement, trier et traiter automatiquement les données, afin d'assurer une première mise en forme intelligible
- détecter en permanence les signaux faibles d'attaque grâce à des algorithmes de détection de mots-clé ou de liens entre les personnes
- caractériser puis juguler les attaques informationnelles, avec des moyens automatisés d'aide à la réaction et des outils ajustés à la cible, à ses habitudes de connexion et aux enjeux locaux.

Par ailleurs, la réflexion reste déterminante pour analyser les critères de gravité d'un événement, sa viralité possible sur l'espace médiatique disponible à ce moment-là. En outre, le processus d'automatisation doit permettre aux opérateurs et à leurs autorités de se concentrer sur le discernement, la rapidité de décision et de réaction face à des situations très fugaces et évolutives. La phase de réaction demeure d'ailleurs un moment très sensible, car l'attribution d'une attaque informationnelle est à la fois compliquée techniquement et délicate politiquement. L'OTAN a identifié en 2016, le cyberspace comme un domaine d'opérations à part entière. En effet, la coopération internationale, parfois complexe, demeure un atout dans ce combat. Enfin, l'évolution technologique, notamment l'emploi de l'IA par les attaquants, rendra à l'avenir la tâche des « veilleurs du cyberspace » encore plus indispensable mais ardue.

Renseigner, défendre et agir face aux infox

Retours d'expérience et interdisciplinarité au service des luttes informatiques



Thomas DELAVALLADE

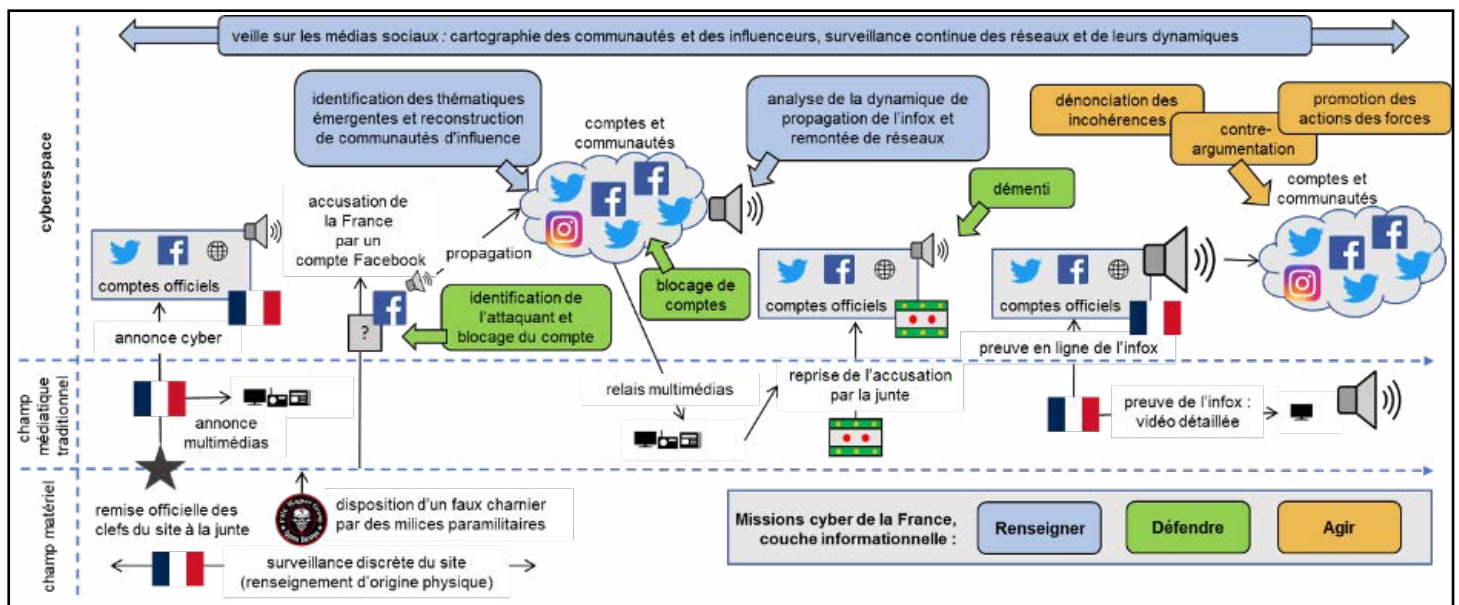
Head of Discipline
Data Valorization & AI, Thales



Sahel, années 2020. La France, qui avait déployé des forces armées dans plusieurs pays pour lutter contre des milices djihadistes, se retire progressivement de l'un d'entre eux. Or, le lendemain de la restitution d'un camp français à la junte au pouvoir, un compte Facebook inconnu accuse, photos et vidéos à l'appui, la France d'avoir laissé un charnier sur cette zone. Cette grave accusation, abondamment reprise sur des plateformes de médias sociaux, instille aussitôt le doute voire la colère, y compris sur le territoire national. Mais la France révèle alors ce qui s'est réellement passé. Ses drones avaient continué de surveiller le site et les images sont éloquentes : le charnier a été créé après le départ des Français par des mercenaires étrangers. Devant ces preuves, les opinions exprimées sur les médias sociaux s'en trouvent retournées.

Ce bref scénario, inspiré d'un cas réel, fait intervenir au moins trois notions : la « fake news » ou infox, les médias sociaux comme terrain propice au phénomène, et les changements d'opinions résultant d'une action immatérielle d'influence. Depuis le milieu des années 2010, le terme infox est devenu omniprésent dans la presse. L'irruption du terme dans le champ lexical de la manipulation de l'information, aux contours flous, ne relève pas uniquement d'un effet de mode : elle traduit de véritables nouveautés de l'époque, dans laquelle les médias sociaux constituent un terreau idoine aux créations et diffusions d'infox, publiquement et vers un auditoire potentiellement mondial.

Face à de telles nouvelles menaces ayant émergé dans le cyberspace, la France s'est progressivement dotée de doctrines de lutte informatique qui se complètent : défensive (LID, 2018), offensive (LIO, 2019), et enfin d'influence (L2I, 2021). Ce corpus doctrinal de cyberdéfense s'appuie sur une décomposition du cyberspace en trois couches interdépendantes : physique (matériels), logique (données numériques, logiciels...) et enfin cognitive, sémantique et informationnelle. Celle-ci regroupe les identités numériques des humains ainsi que la nature et le contenu des relations virtuelles qu'ils entretiennent. Les médias sociaux ont pour composantes principales de troisième couche les entités reliées en communautés et les informations émises, relayées et reçues par ces acteurs. L'infox et les opérations d'influence trouvent leur place dans cette couche. Dans le scénario, l'utilisation d'infox visait à influencer les opinions pour porter atteinte à la réputation de la France. Mais la détection de l'infox, de ses sources et vecteurs de propagation (LID) puis leur blocage (LIO), qui sont des actions dans les couches physique et logique, permettent à la France d'agir dans la couche informationnelle : elle retourne la situation à son avantage par le rétablissement de la vérité et une argumentation publiques (L2I), et conforte ainsi sa légitimité. Dans le corpus doctrinal de cyberdéfense, la lutte informatique est menée pour trois familles de missions : renseigner, défendre, agir.



Luttes informatiques défensive, offensive et d'influence : un exemple

La figure ci-dessus reprend l'ensemble des notions évoquées, éclairées par les étapes de scénario.

Le spectre du cyberspace implique une batterie de méthodes et outils pour mener à bien ces missions. Plusieurs disciplines des sciences humaines et sociales (SHS) sont mobilisées : la psychologie sociale pour les mécanismes de mouvements d'opinion, la sociologie pour les phénomènes de groupe, l'émergence de communautés et de meneurs, la rhétorique pour l'argumentation et la contre-argumentation, et même la philosophie du langage et de l'esprit pour l'analyse du concept d'infox (fiabilité des sources, véracité du contenu, croyances et intentions des acteurs...). Pour passer de l'analyse formelle et conceptuelle offerte par les SHS à une instrumentation outillée de la lutte informatique dans le cyberspace, les différents champs algorithmiques de l'intelligence artificielle (IA) peuvent être mis à contribution : les réseaux de neurones profonds pour l'analyse des contenus multi-modaux (texte, image, audio, vidéo) et de leur véracité ; la segmentation (clustering) des acteurs en fonction des profils de publication pour la fiabilité des sources ; la théorie des graphes pour identifier les communautés d'acteurs, les influenceurs ; l'analyse temporelle pour détecter les thématiques émergentes et quantifier les dynamiques de propagation de ces thématiques ; enfin l'IA symbolique, le Web sémantique et plus généralement les logiques formelles pour l'élaboration de contre-argumentaires.

Au-delà du cadre militaire, la production et la diffusion d'infox via les médias sociaux favorisent l'émergence de rumeurs infondées, de tendances complotistes,

de destruction de réputation, de radicalisation ou d'ingérence étrangère. Ces nouvelles menaces, asymétriques, protéiformes, et parfois fulgurantes, peuvent affecter aussi bien l'État et ses institutions, les entreprises, les organisations que tout individu. Les armées, par l'établissement d'un corpus doctrinal de cyberdéfense, ont ouvert la voie à une démarche structurée pour les comprendre et y faire face. Mais concrètement, les considérables défis posés par les infox via les médias sociaux nécessitent des actions de recherche & développement interdisciplinaires. En amont, un effort de sensibilisation aux risques et de pédagogie quant aux précautions à respecter sera de plus en plus requis dans les milieux tant éducatifs que journalistiques ou d'entreprises. La structuration de ces impératifs est un chantier ouvert.

Article rédigé en collaboration avec : [Philippe Capet](#), ingénieur de recherche, maîtrise de l'information, Ektimo

[Régis Quentin](#), responsable projets cyberdéfense, Thales

Usurpation d'identité par le biais des Deep Fakes dans le Metaverse



Dr. Marc-Oliver PAHL

Titulaire Chaire Cybersécurité
des Infrastructures Critiques
CyberCNI
Directeur de recherches
IMT Atlantique



Les interfaces de réalité virtuelle (RV) ont le potentiel de révolutionner toute une série de domaines, des télécommunications au contrôle à distance en passant par l'analyse immersive. Les "deep fakes", médias synthétiques créés à l'aide de l'intelligence artificielle, constituent une menace importante pour l'intégrité et la fiabilité des applications basées sur la RV. La sophistication croissante des deep fakes les rend difficiles à détecter et peut conduire à la manipulation des signaux d'image et de son. Cela a des conséquences pour les interfaces de commande à distance basées sur la RV, où l'authenticité et la fiabilité des informations et des utilisateurs sont essentielles. Cet article explore les risques potentiels des deep fakes pour les interfaces de commande à distance basées sur la RV et souligne la nécessité de disposer de méthodes de détection robustes.

Les interfaces de réalité virtuelle (RV) offrent une expérience immersive et interactive qui a le potentiel de révolutionner un certain nombre de domaines. Des jeux et divertissements à l'éducation, des soins de santé à la cybersécurité, les interfaces de réalité virtuelle permettent aux utilisateurs d'interagir et de collaborer dans des environnements numériques d'une manière qui semble plus naturelle et intuitive que les interfaces classiques. Les interfaces de commande à distance basées sur la RV permettent aux individus de contrôler des appareils ou de manipuler des objets à distance, créant ainsi un lien plus direct et immédiat avec le monde physique. La capacité croissante des interfaces de RV à reproduire fidèlement l'expérience sensorielle du monde réel leur confère un

potentiel énorme pour un large éventail d'applications. Outre le contrôle à distance et les télécommunications, les interfaces de RV ont également le potentiel de révolutionner le domaine de l'analyse et de la visualisation des données. En créant des environnements immersifs et interactifs pour l'exploration des données, la RV peut créer un nouveau niveau de compréhension qui permet aux individus de voir facilement des modèles et des relations dans des ensembles de données complexes : "L'analyse immersive".

Les "deep fakes" sont des médias synthétiques créés à l'aide d'algorithmes d'intelligence artificielle. Ils impliquent la manipulation de signaux visuels et audio pour créer des imitations réalistes et convaincantes de personnes, souvent à des fins malveillantes⁽¹⁾. Au fur et à mesure que la technologie de l'apprentissage profond (Deep Learning) progresse, les Deep Fakes deviennent plus sophistiqués et plus difficiles à détecter. Cela suscite de vives inquiétudes quant à leur impact potentiel sur divers secteurs tels que la politique, le divertissement et la sécurité de l'information. La capacité des Deep Fakes à manipuler les signaux audio et visuels, combinée à leur réalisme croissant, en fait une menace potentielle pour l'intégrité et la fiabilité des interfaces de commande à distance basées sur la RV et des analyses immersives. Il est donc essentiel de bien comprendre les risques posés par les "deep fakes" et de développer des méthodes robustes pour détecter et combattre cette menace.

Un cas d'utilisation possible des deep fakes dans un contexte de contrôle à distance basé sur la RV est la

manipulation des identités dans le contrôle collaboratif d'une infrastructure critique. Imaginez un scénario dans lequel un groupe d'individus est responsable du contrôle et de la surveillance à distance d'une infrastructure critique, telle qu'une centrale nucléaire ou un barrage. Si un Deep Fake est utilisé pour usurper l'identité de l'une des personnes impliquées dans la collaboration, les conséquences pourraient être catastrophiques. Par exemple, la personne malicieusement authentifiée pourrait manipuler les systèmes de contrôle, perturber le fonctionnement normal de l'infrastructure critique ou causer des dommages encore plus importants. En outre, le Deep Fake pourrait également abuser des relations de confiance pour convaincre les autres de fausses informations, obtenir l'accès à des informations sensibles ou manipuler le comportement des autres. Parce que le Deep Fake donne l'impression d'être un membre authentique et digne de confiance de la collaboration, les autres sont plus susceptibles de croire les informations ou les instructions de cette personne.

L'authenticité et la fiabilité de l'identité de la personne sont d'une importance capitale, car toute manipulation pourrait avoir de graves conséquences. Cela souligne l'importance non seulement de développer des méthodes d'identification robustes, mais aussi d'accroître la sensibilisation et la compréhension des risques potentiels des "Deep Fakes" dans les scénarios de contrôle à distance basés sur la RV. Ces mesures permettront d'atténuer la menace des "Deep Fakes" et de garantir la fiabilité et la sécurité des interfaces de commande à distance basées sur la RV⁽²⁾.

Pour atténuer ces risques, il faut mettre au point des méthodes d'identification robustes capables de distinguer avec précision les Deep Fakes des personnes authentiques⁽³⁾. Il existe actuellement un certain nombre d'approches, notamment celles qui analysent les signaux visuels et sonores, ainsi que celles qui utilisent des algorithmes d'apprentissage automatique pour détecter les anomalies dans les médias synthétiques. Parmi les méthodes les plus prometteuses figurent l'analyse des mouvements et des expressions du visage, la modulation de la voix et l'utilisation de réseaux neuronaux profonds pour détecter des schémas dans les médias synthétiques. Malgré ces progrès, les deep fakes sont de plus en plus sophistiqués et il reste beaucoup à faire pour que les méthodes de détection des deep fakes restent efficaces face à cette menace en constante évolution. Le défi de la détection des deep fakes devient encore plus grand dans le contexte des interfaces de commande à distance basées sur la RV, où la dimension supplémentaire de l'interactivité

et les environnements visuels et sonores complexes peuvent rendre la détection des deep fakes plus difficile. Dans de nombreux systèmes de RV actuels, tels que le Metaverse de Meta, les personnes sont représentées par des avatars, ce qui peut rendre difficile la détermination précise de leur identité. Contrairement à l'interaction en face à face, les interfaces de RV ne donnent pas accès à des indices physiques tels que le langage corporel et les expressions faciales qui peuvent aider à authentifier l'identité d'une personne.

Les systèmes basés sur des caméras et des sons en direct peuvent partiellement atténuer ce problème. Toutefois, les interfaces de RV peuvent également permettre aux individus de manipuler ou de déguiser plus facilement leur identité. Les informations sensorielles limitées disponibles dans la RV signifient également qu'il est plus difficile de détecter des contrefaçons profondes. Une autre approche de l'authentification des utilisateurs humains dans un contexte de collaboration en RV consiste à utiliser une approche hybride qui exploite plusieurs facteurs. En combinant plusieurs sources de données, il est plus difficile pour les acteurs malveillants d'usurper l'identité d'un utilisateur réel, car ils doivent manipuler plusieurs facteurs pour éviter d'être détectés. Par exemple, les données d'accélération et la résistance de la peau d'une personne pourraient être combinées avec les mouvements oculaires et les données de température pour créer une "empreinte digitale" unique pour chaque utilisateur.

En résumé, l'utilisation d'interfaces de réalité virtuelle (RV) pour le contrôle à distance dans les applications d'infrastructures critiques pose un défi majeur pour l'authentification des utilisateurs humains. Les "deep fakes" représentent un risque particulièrement important dans ce contexte, car ils peuvent être utilisés pour usurper l'identité d'utilisateurs réels et manipuler le comportement d'autres personnes.


Un environnement
juridique en
construction



PARTIE

Contributeurs

 Pr. Brunessen BERTRAND

 Florent FAVIERE

 Dr. Sandrine TURGIS

 Dr. Emmanuel BRESSON

Lutte contre la désinformation

Quelle **régulation juridique** ?



Pr. Brunessen BERTRAND

Professeure de droit
Université de Rennes
Chaire Jean Monnet



Réguler les algorithmes

Au profilage des individus par leurs données personnelles s'ajoute une forme de déterminisme algorithmique : les intermédiaires et prestataires de services utilisent des données personnelles, de manière qui ne semble pas toujours conforme au RGPD, pour établir des algorithmes de classement et de recommandation ciblant, par un profilage comportemental, les publicités politiques enfermant les utilisateurs dans une bulle informationnelle limitée aux contenus déterminés en fonction de ses préférences personnelles. Les algorithmes de recommandation des plateformes contribuent très largement à l'amplification des discours et à la viralité des « fake news ». Le problème est loin d'avoir été résolu depuis l'affaire Cambridge Analytica comme le montre encore l'audition, en octobre 2021, de Frances Haugen devant la Commission du commerce du Sénat américain, cette lanceuse d'alerte qui a travaillé pour Facebook, précisément dans les services chargés de la lutte contre la désinformation.

Les algorithmes des plateformes, en tout cas des plateformes américaines (la situation est différente pour les plateformes chinoises), sont conçus pour mettre en avant des contenus susceptibles d'attirer l'attention. Les ingérences étrangères peuvent ainsi manipuler ces algorithmes par la prise de contrôle de comptes de médias sociaux, l'utilisation de comptes de réseaux sociaux commandés par des robots : les algorithmes sont ainsi manipulés par une pléthore de faux comptes, d'usines à trolls, de bots, etc. C'est le phénomène par

exemple des « fermes à clics » et ses travailleurs du clic rémunérés pour donner artificiellement une impression de popularité à des sites Internet diffusant de fausses informations et qui se trouveront ainsi valorisés par les algorithmes de recommandation. Il y a ainsi une instrumentalisation des plateformes numériques pour diffuser largement des contenus inauthentiques, qui peut avoir un objectif politique, mais aussi économique.

Réguler le financement de la désinformation

La désinformation est ainsi très liée à des enjeux commerciaux : l'enjeu juridique est là celui de la démonétisation de la désinformation par la publicité. Il faut distinguer deux questions. D'une part, la nécessité d'imposer des restrictions au financement par la publicité des sites diffusant ouvertement de fausses informations. D'autre part, la sanction financière des entreprises qui participent à des opérations d'influence ou à des activités d'ingérence étrangère. Des enjeux commerciaux alimentent en effet la manipulation des médias sociaux avec l'émergence d'entreprises spécialisées dans l'acquisition de faux comptes. La régulation de l'intelligence artificielle et des trucages hyperréalistes. Les technologies numériques peuvent facilement accréditer toute sorte de théorie : il y a de multiples façons de modifier ou détourner des contenus, et la crédibilité de ces manipulations croît à mesure que l'intelligence artificielle se perfectionne. La proposition de règlement sur l'intelligence artificielle prévoit d'interdire la mise sur le marché, la mise en service ou l'utilisation d'un système d'IA qui a recours à des techniques subliminales

au-dessous du seuil de conscience d'une personne pour altérer substantiellement son comportement d'une manière qui cause ou est susceptible de causer un préjudice physique ou psychologique à cette personne ou à un tiers. Pour la désinformation, ce sont surtout les perspectives de deepfake qui sont susceptibles d'accroître notablement ce problème. On peut regretter l'approche minimaliste retenue par la Commission dans sa proposition de règlement sur l'intelligence artificielle à cet égard. Pour ces systèmes d'IA spécifiques qui utilisent des trucages vidéo ultra-réalistes, seules des obligations minimales en matière de transparence sont proposées.

Réguler les comportements

La manipulation est aussi liée à une évolution des usages et des comportements, amplifiée par les effets systémiques de la diffusion des fausses informations sur Internet. Penser un cadre juridique pertinent par une régulation des technologies doit aussi être finement apprécié à l'égard des comportements humains, qu'elle peut contribuer à accentuer. L'approche des réglementations comportementale, qui intègre les effets d'une norme sur les comportements qu'elle peut générer (comme des effets rebonds), semble indispensable. Les nouveaux usages numériques ont amplifié les risques de manipulation : les comportements des individus en ligne ont massivement alimenté en données des algorithmes et sont ainsi, par un effet de rétroaction, modelés en retour par les algorithmes de recommandation qui enferment les individus dans une bulle informationnelle, d'autant mieux acceptée par eux qu'elle est déterminée par leurs préférences individuelles. La tendance des citoyens à ne s'informer que par des réseaux sociaux parachève cet enfermement : le numérique, et pour une partie notable des citoyens les réseaux sociaux, est ainsi devenu la principale, parfois la seule, source d'information. C'est un déterminisme numérique qui se dessine alors, limitant progressivement l'autonomie et la liberté individuelle.

L'amplification des discours résulte aussi des comportements dès lors que les plateformes reposent sur une économie de l'attention. Il a été démontré que l'attention est attirée par les contenus les plus clivants : les ressorts même de l'économie de l'attention sont directement liés au développement de toutes formes de manipulation. La viralité des fake news est favorisée par des comportements humains. La vitesse de propagation de fausses nouvelles est beaucoup plus importante que celles des informations vraies, parce qu'elle suscite l'attrait de la nouveauté et suscite l'étonnement donc l'intérêt. La structuration des échanges sur les réseaux sociaux favorise la polarisation des idées, et ne laisse guère de place à la nuance. Les algorithmes favorisent les formes d'expression les plus clivantes, rendues plus simples à exprimer sous le

couvert de l'anonymat. La transparence des algorithmes et la levée de l'anonymat apparaissent ainsi comme des enjeux structurant de la réglementation juridique. De telles obligations pourraient limiter les biais comportementaux qui alimentent en retour les possibilités de manipulation. Le degré d'influence réelle de la désinformation et des tentatives de manipulation de l'opinion n'est cependant pas facile à mesurer avec précision. Certaines études sociologiques ont plutôt montré que les fausses informations touchent essentiellement un public sensible à ce type de contenu. À l'inverse, l'instauration de mesures perçues comme excessives pour lutter contre la désinformation pourrait susciter la méfiance et ainsi générer des effets inverses à ceux souhaités.

L'Union européenne et la lutte contre les campagnes de manipulation de l'information dans le contexte du conflit armé en Ukraine



Florent FAVIERE

Doctorant en droit public et en droit de l'UE
Université de Rennes

IODE

Alma Mater Studiorum Università di Bologna
CIRDE



« Les campagnes de désinformation sont de nature à remettre en cause les fondements des sociétés démocratiques et font partie intégrante de l'arsenal de guerre moderne ».

Cette formulation, employée à plusieurs reprises par le Tribunal de l'Union dans son arrêt *Russia Today France* du 27 juillet 2022, confirme l'utilisation de plus en plus systématique des campagnes de désinformation dans les conflits armés. Dès mars 2015, dans le contexte de l'annexion illégale de la Crimée et de Sébastopol par la Fédération de Russie, le Conseil européen a reconnu pour la première fois la menace que représentaient les campagnes de désinformation orchestrées par la Russie. Dans ce contexte, une task force « East StratCom » fut créée au sein de la division des communications stratégiques et de l'analyse de l'information du Service européen pour l'action extérieure afin de lutter contre les campagnes de désinformation menée par la Russie. Le projet EUvsDisinfo constitue une des actions concrètes menées par la task force en diffusant sur sa base de données des cas de désinformation provenant des médias pro-Kremlin. Toutefois, alors que les campagnes de désinformation menées par la Russie constituaient également pour l'Union un défi majeur, aussi bien démocratique lors des processus électoraux que dans le contexte de la pandémie de COVID-19, les dispositifs déployés par l'Union européenne demeuraient relativement modestes. L'agression armée de l'Ukraine par la Russie initiée le 24 février 2022 marqua le début d'un renforcement significatif de l'action de l'Union en matière de lutte contre les campagnes de désinformation par la mobilisation inédite de mesures restrictives dans le cadre de la politique étrangère et de sécurité commune.

Le renforcement de l'action de l'Union dans le contexte du conflit armé en Ukraine

Dans ses conclusions du 24 février 2022, le Conseil européen demanda à la Fédération de Russie de cesser sa campagne de désinformation et a marqué son accord sur de nouvelles mesures restrictives pour la Fédération de Russie. Cette demande fut relayée par le Parlement européen qui condamna, dans sa résolution du 1er mars 2022, « la guerre de l'information menée par les autorités russes, les médias d'État et les alliés de la Russie ». Le même jour, le Conseil de l'Union adopta des mesures restrictives visant à suspendre d'urgence les activités de diffusion des médias publics russes Sputnik et Russia Today dans l'Union ou en direction de l'Union. Ces médias, placés sous le contrôle permanent, direct ou indirect, des dirigeants de la Fédération de Russie et menant des campagnes de désinformation, constituent pour le Conseil, « une menace importante et directe pour la sécurité de l'Union ». Ces mesures visent par leur objet à interdire aux opérateurs de diffuser, d'autoriser ou de faciliter la diffusion de leurs contenus. Elles n'empêchent pas en revanche à ces médias d'exercer dans l'Union d'autres activités que la diffusion, telles que des enquêtes et des entretiens. Le 3 juin 2022, l'objet de ces mesures fut étendu à l'interdiction de faire la publicité de produits ou de services dans des contenus produits ou diffusés par les personnes morales, entités ou organismes énumérés. Par la même occasion, la liste des médias visés par ces mesures fut étendue aux

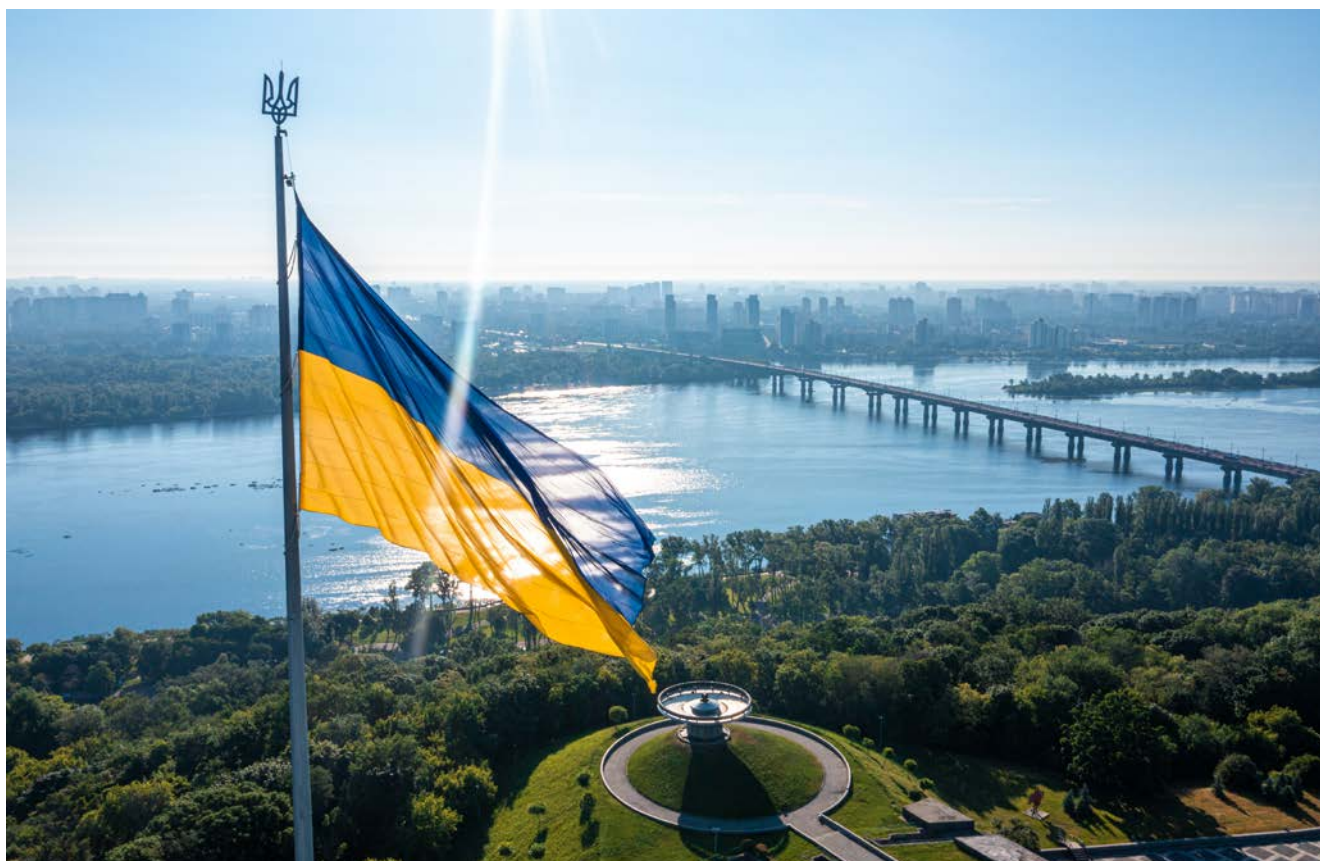
médias RTR Planeta, Russia 24 et TV Centre International. Ces mesures adoptées dans le cadre de la politique étrangère et de sécurité commune constituent à ce jour les premières mesures restrictives adoptées par l'Union visant spécifiquement à interdire la diffusion de certains médias établis sur le territoire de l'Union.

L'approche globale de l'Union contre les campagnes de manipulation de l'information

L'adoption de mesures restrictives dans le cadre d'une politique par nature externe de l'Union, mais visant spécifiquement à interdire la diffusion de certains médias établis sur le territoire de l'Union, démontre la particularité de la menace que représentent les campagnes de désinformation et de leur impact à la fois sur la sécurité intérieure et extérieure de l'Union et de ses Etats membres. Bien que l'adoption de ces mesures se justifie au regard de la nécessité de la mise en place d'une réaction rapide et efficace face au contexte exceptionnel lié à l'agression armée de l'Ukraine par la Fédération de Russie, elles mettent aussi en lumière la nécessité d'une approche globale en matière de lutte contre les campagnes de désinformation par la mobilisation d'instruments internes et externes de l'Union. En effet, par l'intermédiaire de retweet, de copies textuelles ou par d'autres canaux

de réseaux sociaux, les médias visés par ces mesures restrictives parviennent à contourner les interdictions de diffusion de leur contenu imposées par ces mesures. Ces moyens de contournement n'étant pas couverts par les mesures, l'Union s'en remet aux plateformes pour démanteler ces tentatives de contournement des campagnes de désinformation menées par la Russie. En cela, et dans l'attente de la mise en application pratique du Digital service act, le code de bonne pratique de l'Union en matière de désinformation de 2018, renforcé en juin 2022, un instrument non contraignant par lequel les signataires s'engagent à lutter contre la désinformation en ligne, constitue un complément certain aux tentatives de contournement des mesures restrictives par les médias ciblés.

Pour aller plus loin, accédez aux sources et contenus complémentaires disponibles en fin d'ouvrage ⁽¹⁾



Les enjeux pour les droits fondamentaux de la lutte contre la manipulation de l'information



Dr. Sandrine TURGIS

Maître de conférences en droit public
Université de Rennes
IODE/CREC DE SAINT-CYR



Si la lutte contre la manipulation de l'information a pour objet de protéger les droits fondamentaux, elle doit être strictement calibrée afin de ne pas entraîner leur violation.

Communiquer, informer et s'informer sont des activités garanties par les instruments de protection des droits fondamentaux, tant au niveau national, qu'europpéen et international. Ainsi de l'article 11 de la Déclaration des droits de l'homme et du citoyen (DDHC) selon lequel « la libre communication des pensées et des opinions est un des droits les plus précieux de l'homme », à l'affirmation de l'article 19 de la Déclaration universelle des droits de l'homme (DUDH) selon laquelle « tout individu a droit à la liberté d'opinion et d'expression », en passant par la formule « toute personne a droit à la liberté d'expression » de l'article 10 de la Convention européenne des droits de l'homme (CEDH), les formulations sont différentes mais confirment l'importance du droit à la liberté d'expression et de ses composantes. Ainsi la lutte contre la manipulation de l'information s'inscrit dans une démarche de protection des droits fondamentaux.

Cependant, comment éviter que le remède ne soit pire que le mal ?

Comment éviter que les mesures adoptées pour lutter contre la manipulation de l'information ne mettent à mal les droits fondamentaux qu'elles cherchent justement à défendre ? L'ambivalence de la lutte contre la manipulation d'information pour les droits fondamentaux est donc au cœur de la problématique. En effet, si la lutte contre la manipulation de l'information a pour objet de protéger les droits fondamentaux, elle doit être strictement calibrée afin de ne pas entraîner leur violation.

Lutter contre la manipulation de l'information pour protéger les droits fondamentaux

Le droit à la liberté d'expression « comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontière » (article 10 de la CEDH). La CEDH, adoptée en 1950, avait donc déjà pris en compte la dimension transfrontière de la liberté d'expression, qui a trouvé une nouvelle forme de concrétisation grâce au numérique. Cependant, les évolutions des moyens de communication et des canaux d'information, qui ont touché les médias traditionnels mais aussi favorisé l'émergence de nouveaux médias, ont fondamentalement changé le contexte. Or, la diffusion d'une information de qualité est particulièrement importante pour la concrétisation de l'ensemble des droits fondamentaux, à tel point que la presse a été qualifiée par la Cour de Strasbourg, qui veille au respect de la CEDH, de « chien de garde de la démocratie ».

Lutter contre les menaces hybrides

Afin de défendre à la fois la liberté d'expression et la démocratie, les Etats ne doivent pas seulement s'abstenir de manipuler eux-mêmes l'information mais doivent aussi agir pour éviter que de telles manipulations se produisent. Ainsi la mise en place d'un cadre de lutte contre la manipulation de l'information s'inscrit dans cette démarche – éventuellement sous le vocable de lutte contre les « menaces hybrides » – et permet aux Etats de respecter leurs obligations positives en la

matière. Les concrétisations normatives sont diverses, qu'il s'agisse de l'adoption de la loi de 2018 sur la manipulation de l'information – dont l'ARCOM vient à nouveau de souligner les limites dans son rapport de 2022 – ou de l'adoption par l'Union européenne du règlement européen sur les services numériques ou Digital Services Act (DSA) de 2022 visant à une responsabilisation de tous les intermédiaires en ligne (plateformes, moteurs de recherche ...). Cependant, cette lutte doit être calibrée afin de respecter les droits fondamentaux.

Les risques pour les droits fondamentaux de la lutte contre la manipulation de l'information

Les moyens de la lutte contre la manipulation de l'information pourraient cependant porter atteinte à cette liberté d'expression qu'ils brandissent comme justification. D'ailleurs, il a pu être avancé que ceci pouvait éventuellement être « le véritable effet final recherché par les puissances étrangères à l'origine des manipulations de l'information : non pas tant de convaincre la population de tel ou tel récit que d'inciter les gouvernements à prendre des mesures contraires à leurs valeurs démocratiques et libérales, ce qui suscitera des réactions (d'une autre partie de la classe politique et de la société civile) et in fine contribuera à approfondir les divisions de la société » (Jean-Baptiste Jeangène Vilmer, Alexandre Escorcía, Marine Guillaume et Janaina Herrera, *Les manipulations de l'information : un défi pour nos démocraties*, CAPS et IRSEM, 2018, p. 25.).

Les risques pour les droits fondamentaux, à commencer par la liberté d'expression, des réponses de la lutte contre la manipulation de l'information sont régulièrement signalés par les organes de contrôle des droits de l'homme qui rappellent l'importance pour les Etats de respecter les principes de légalité, de nécessité et de proportionnalité de toute restriction (notamment Rapporteur spécial des Nations Unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression, Représentant de l'Organisation pour la sécurité et la coopération en Europe pour la liberté des médias, Rapporteur spécial de l'Organisation des Etats Américains pour la liberté d'expression et Rapporteur spécial sur la liberté d'expression et l'accès à l'information de la Commission africaine des droits de l'homme et des peuples, Déclaration conjointe du vingtième anniversaire : les défis clés pour la liberté d'expression au cours de la prochaine décennie, 2020).

Un équilibre à trouver

Ainsi, même si l'exercice semble parfois ressembler à un numéro d'équilibriste, les réponses et les mesures

adoptées pour lutter contre la manipulation de l'information doivent s'inscrire dans le cadre du respect de la liberté d'expression et des autres droits fondamentaux, comme le rappelle le Comité des ministres du Conseil de l'Europe dans sa recommandation sur les effets des technologies numériques sur la liberté d'expression de 2022. C'est uniquement à ces conditions que la lutte contre la manipulation de l'information interviendra effectivement au profit de la protection des droits et des valeurs qu'elle doit défendre.

Dispositifs étatiques pour lutter contre la manipulation d'information



Dr. Emmanuel BRESSON

Directeur d'engagement
Capgemini



Prenant acte des risques inhérents aux campagnes de manipulation d'information, l'État a mis en place des réponses législatives, organisationnelles mais aussi militaires, pour les opérations de lutte d'influence menées sur les théâtres extérieurs.

Une loi « anti fake news »

La loi française n°2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information⁽¹⁾ vise à mieux protéger la démocratie contre les diverses formes de manipulation ; elle vient, comme souvent, compléter des lois existantes (la loi de 1881 sur la liberté de la presse comportait déjà un volet pénal punissant la diffusion de fausses nouvelles). Son champ d'application concerne les plateformes numériques de diffusion massive d'information. Elle contraint notamment ces plateformes à fournir aux internautes des moyens de signaler les contenus erronés ou manipulés et à lutter contre les comptes utilisateurs envoyant des fausses informations de façon massive. Elle les oblige aussi à publier avec transparence les rémunérations reçues pour la promotion de contenus liés à un débat d'intérêt général. La loi du 22 décembre 2018 comporte également un volet destiné à protéger les processus électoraux, cible de choix des adversaires de nos démocraties. Ainsi, dans les trois mois qui précèdent un scrutin national, lorsque des allégations trompeuses de nature à altérer la sincérité dudit scrutin sont diffusées de manière délibérée, artificielle ou automatisée et massive, le juge des référés peut être saisi et prescrire toutes mesures proportionnées et nécessaires pour faire cesser cette diffusion. En parallèle, la loi autorise le Conseil supérieur de l'audiovisuel (CSA, aujourd'hui ARCOM) à ordonner la suspension de la diffusion de

ces informations par tout procédé de communication électronique jusqu'à la fin des opérations de vote. Il n'est évidemment pas question de contrôler l'expression d'idées contradictoires et d'empêcher les débats d'opinions ; les informations exagérées, satiriques ou humoristiques ne sauraient être considérées, en tant que telles, comme des manipulations. La cible de la loi du 22 décembre 2018 concerne les informations qui, d'une part, sont présentées comme vraies mais dont on peut démontrer la fausseté de manière objective, et qui, d'autre part, véhiculent un contenu de nature à troubler manifestement l'ordre public. La loi ne fait pas de distinction quant à l'origine, prétendue ou supposée, de ces informations.

Un service à compétence nationale

Une autre réponse étatique aux manipulations d'information s'est concrétisée dans la création du service VIGINUM (service de vigilance et de protection contre les ingérences numériques étrangères). L'histoire et la réflexion qui ont mené à la naissance de VIGINUM sont intéressantes à plus d'un titre. Le constat de la réalité et du sérieux des menaces contre le fonctionnement des démocraties s'est accéléré dans la dernière décennie, de plus en plus de pays ayant subi des campagnes d'influence caractérisées par la diffusion d'information sciemment manipulées (élections américaines en

2016, referendum sur le Brexit, élection présidentielle française de 2017...). Le but évident de ces campagnes est d'influencer la décision de vote de certains groupes de citoyens et leur vecteur d'attaque dépasse désormais celui des processus électoraux pour couvrir l'ensemble des débats d'opinion de la vie sociale des pays visés. En France, en parallèle de la réponse législative, le besoin d'une coordination interministérielle destinée à caractériser la menace (comprendre les mécanismes de construction et de diffusion des fausses informations, anticiper et mettre en place des mesures de protection...) a été rapidement identifié. De fait, après les attaques terroristes islamistes ayant eu lieu en octobre 2020 (à Conflans-Sainte-Honorine puis à Nice), la France a été la cible d'une campagne particulièrement virulente de désinformation et de discrédit, principalement sur les réseaux sociaux. Cette attaque, structurée et coordonnée depuis la Turquie, a parfaitement illustré la réalité de la menace. Elle a mis en lumière la nécessité de caractériser rapidement ce type d'opérations hostiles, afin de réagir. Le président de la République a instauré VIGINUM par décret⁽²⁾ le 13 juillet 2021. Service à compétence nationale, il est rattaché au SGDSN, organisme interministériel coordinateur des questions de défense et de sécurité nationale. Le service se concentre sur les opérations orchestrées par des acteurs étrangers et portant atteinte aux intérêts fondamentaux de la Nation. Le périmètre des missions qui incombent au service –détecter et caractériser les opérations d'ingérence numérique étrangères– se définit à travers 4 critères : l'objectif de porter atteinte aux intérêts de la Nation ; le contenu imprécis, trompeur ou inexact de l'information ; l'attitude consistant à diffuser de façon massive, automatisée et non-authentifiée ; et enfin l'origine étrangère (étatique ou non) de l'orchestration de l'opération. VIGINUM n'a pas vocation à se prononcer sur la véracité des informations, mais plutôt à signaler et mettre en lumière leur origine et mode de propagation suspects. L'équipe est encore en phase de croissance, avec un objectif de plusieurs dizaines d'experts (data analystes, spécialistes des média numériques...). Dans le cadre de ses missions, VIGINUM est autorisé à réaliser un traitement automatisé de données personnelles⁽³⁾. Étant donné la sensibilité du sujet, l'activité de VIGINUM est encadrée de façon très stricte. Un comité éthique dédié, composé de plusieurs personnalités qualifiées (scientifiques, journalistes, diplomates...) et présidé par un membre du Conseil d'État, qui a pour mission de veiller à la conformité des activités du service avec le décret⁽³⁾.

Des efforts analogue en Europe et aux États-Unis

La France n'est pas le seul pays à avoir mis en place des réponses de niveau étatique ou institutionnel

en réponse aux attaques basées sur la manipulation d'information : L'Union européenne a ainsi mis en place en 2015 une East StratCom Task Force (ESCTF) consacrée à la lutte contre campagnes de manipulation venant de Russie⁽⁴⁾. Aux États-Unis, le Global Engagement Center au sein du Département d'État a pour mission de « [...] coordonner les efforts du gouvernement fédéral pour détecter, comprendre, exposer et contrer la propagande étatique et non-étatique [...] visant à saper ou influencer sur la politique, la sécurité ou la stabilité des États-Unis, de leurs alliés et de leurs partenaires »⁽⁵⁾.

Une doctrine militaire adaptée

Enfin, pour compléter le paysage de la lutte contre la manipulation d'information, il faut mentionner l'effort fait par les forces armées pour intégrer cette notion dans les doctrines de combat numérique⁽⁶⁾. En effet, les armées françaises, dans leurs opérations, sont la cible d'attaques informationnelles menées par des acteurs ou des États hostiles : on se souvient de la campagne, sur Facebook notamment, laissant entendre que les militaires français pilleraient les ressources minières au Mali. L'élaboration d'une doctrine de lutte informatique d'influence (L2I) offre un cadre et des outils pour l'action des armées : détection des attaques, réaction et dénonciation, actions de déception, recueil de renseignements. Précisons que les actions de L2I des forces armées concernent exclusivement les opérations à l'extérieur du territoire national, et qu'elles sont planifiées et conduites, comme l'ensemble des opérations militaires françaises, dans le strict respect du droit international et humanitaire⁽⁷⁾.

Pour aller plus loin, accédez aux contenus complémentaires en fin d'ouvrage ⁽⁸⁾

Pôle d'excellence cyber

Lexique





Notre lexique de la lutte contre la manipulation de l'information a été construit de la manière suivante :

1. Élaboration collective d'une liste de termes à définir ;
2. Consolidation des définitions avec par ordre de priorité les sources suivantes : Académie française, Larousse ou Robert, lexiques d'organismes professionnels, wikipédia, et définitions créées ou utilisées par nos rédacteurs
3. Relecture collaborative afin de consolider l'ensemble des avis et contributions.

Avatar

1. Chacune des incarnations successives du dieu hindou Vishnou. Fig. Chacune des formes diverses que prend une personne ou une chose.

2. (Informatique) Apparence que prend une personne dans un environnement graphique informatique (image dans un forum internet ou un clavardoir, objet 3D dans un univers virtuel). Cette image peut illustrer autre chose que l'utilisateur lui-même.

3. (Renseignement) entité manipulée clandestinement pour prendre des contacts, diffuser des informations ou en recueillir.

Sources : académie et wikipedia

B2B – B2B2C – B2C

Ces concepts sont souvent utilisés pour décrire les cibles ou le marché auquel s'adresse une entreprise.

B2B : B2B est un raccourci très couramment utilisé pour le terme anglais "business to business". Le B2B, désigne donc les activités commerciales et marketing réalisées entre entreprises, ou organisations. On parle également de commerce inter-entreprises.

Le B2B2C ou BtoBtoC est un terme utilisé pour désigner la double problématique marketing qui consiste à devoir s'adresser aussi bien à une cible de professionnels (distributeurs ou prescripteurs) qu'au client final consommateur.

Le terme B to C pour "business to consumer" désigne les activités commerciales ayant pour cible un client consommateur qui n'est pas une entreprise.

Source definitions-marketing.com

Bulle de filtres et chambres d'écho

Principes propres aux réseaux sociaux et à la manière dont les audiences peuvent être confinées dans des sphères d'opinions identiques.

L'émergence de ces mécanismes est venue de l'usage de technologies et de méthodologies de personnalisation de contenus publicitaires s'appuyant sur des cookies, et appliqués à des contenus de réseaux sociaux.

Bulle de filtre : désigne le mécanisme de filtrage de

l'information parvenant à un usager de l'Internet. Selon le spécialiste de l'Internet Eli Pariser, les bulles de filtre résultent des dispositifs de personnalisation des contenus en ligne et auraient pour conséquence d'isoler intellectuellement les internautes et de réduire la diversité des informations auxquelles ils sont exposés.

Chambre d'écho : Pour de nombreux analystes, la radicalisation des opinions sur l'Internet serait largement le fait de ce que l'on appelle des chambres d'écho. Sur l'Internet, les individus auraient tendance à échanger préférentiellement avec les personnes qui s'intéressent aux mêmes sujets qu'eux et qui partagent des opinions proches des leurs. C'est ainsi que se formeraient des communautés virtuelles au sein desquelles les internautes partageraient et recevraient des informations focalisées sur leurs centres d'intérêt et conformes à leurs croyances. Si ces communautés sont qualifiées de chambres d'écho, c'est que la voix de chacun de leurs membres y ferait essentiellement écho à celle des autres. Fonctionnant comme des caisses de résonance de la vision du monde de ces individus, les chambres d'écho seraient un lieu de radicalisation des esprits.

Source : fondationdescartes.org

Communication

Action de faire savoir, faire connaître.

(Militaire) Communication stratégique ou Stratcom : mise en cohérence et coordination des messages avec les actions, au service d'une stratégie

(Sciences sociales) La communication est l'action de transformer l'autre par la simple transmission d'une information et non par l'action d'une force mettant en œuvre une énergie. Il ne faut pas comprendre le verbe transformer comme une volonté de manipulation, mais comme l'expression de l'intention de l'émetteur qui cherche à ce que son message soit intégré par le récepteur.

Sources : académie et rédacteurs de l'ouvrage.

Cybernétique

κυβερνητικός, du grec : gouverner, piloter.

Entre 1935 et 1939, Oscar Schmidt, ingénieur allemand, exprime le premier les principes généraux des mécanismes auto-régulés. Par l'utilisation du calcul binaire pur pour l'informatique, Louis Couffignal a la prescience du phénomène cybernétique. Parmi les premiers cybernéticiens figurent également Grey Walter, W.S. McCulloch et surtout William Ross Ashby (1903-1972), psychiatre et ingénieur qui, entre 1945 et 1950 conçoit l'appareil cybernétique le plus remarquable : l'homéostat, son « cerveau artificiel ».

Platon puis André-Marie Ampère (1775-1836) ont employé le terme de cybernétique, mais il s'agit ici de simples convergences sémantiques ; dans « Essai sur la philosophie des sciences », 1834, Ampère le définit comme l'« étude des moyens de gouvernement ».

Cybernétique : Mesure de la diversité des choix dans un répertoire de messages possibles. Voir notamment l'ouvrage « la théorie de l'information » de Shannon et Weaver qui théorisent notamment la manière de séparer le code porteur de sens du bruit, autrement dit le rapport signal/bruit. A relier à cette citation : L'information c'est la différence qui fait la différence. (Gregory Bateson)

Cyber : Relating to or characteristic of the culture of computers, information technology, and virtual reality : the cyber age ; origin : 1980s ; abbreviation of cybernetics. Source : Académie, Wikipedia, Oxford dictionary et rédacteurs.

Déception, opérations de déception

(Militaire) Ensemble des mesures actives ou passives visant à tromper l'adversaire et à le faire réagir de façon préjudiciable à ses propres intérêts.

La déception englobe la dissimulation et la simulation.

Source : Wiktionary.

Deep learning

L'apprentissage profond est un procédé d'apprentissage automatique utilisant des réseaux de neurones possédants plusieurs couches de neurones cachées. Ces algorithmes possédant de très nombreux paramètres, ils demandent un nombre très important de données afin d'être entraînés. Source : CNIL

Deepfake, Hypertrucage

Le deepfake, ou hypertrucage est une technique de synthèse multimédia reposant sur l'intelligence artificielle. Elle peut servir à superposer des fichiers vidéo ou audio existants sur d'autres fichiers vidéo (par exemple changer le visage d'une personne sur une vidéo) ou audio (par

exemple reproduire la voix d'une personne pour lui faire dire des choses inventées). Les hypertrucages vidéos sont aussi appelés infox vidéo ou vidéotox.

Source : Wikipedia.

Diffusion massive

Ensemble des formes, des conditions et des effets d'une émission régulière, continue, fréquente (et même permanente), vers un public ad hoc, d'éléments d'information, de formation, d'expression intellectuelle, de pression mentale et de thérapie sociale.

Source : Thésaurus de l'activité gouvernementale du Québec.

Fait

Acte, phénomène, action ; chose ou événement qui se produit, dans le temps et dans l'espace ; ce qui est reconnu comme certain, incontestable ; tout événement susceptible de produire des effets de droit, d'avoir des conséquences juridiques.

Source : Larousse.

Fake news – Infox – Hoax - Canular

Les infox, fausses nouvelles, fausses informations, informations fallacieuses, canards, en anglais fake news, sont des nouvelles mensongères diffusées dans le but de manipuler ou de tromper le public en s'appuyant en particulier sur des procédés de tromperie (falsification de son, d'image ou de personne).

Dans un contexte militaire, ceci est particulièrement illustré par les fausses nouvelles antifrancaises au Mali en 2022 ou par une partie de la communication autour de l'invasion de l'Ukraine par la Russie.

Source : wikipedia.

Influence

Action, généralement continue, qu'exerce quelqu'un ou quelque chose sur quelque chose ou sur quelqu'un ; ascendant de quelqu'un sur quelqu'un d'autre ; pouvoir social et politique de quelqu'un, d'un groupe, qui leur permet d'agir sur le cours des événements, sur des décisions prises.

Militaire : Les opérations militaires d'influence regroupent l'ensemble des activités dont l'objet est d'obtenir un effet sur les comportements d'individus, de groupes ou d'organisations (appelés infocibles) afin de contribuer à l'atteinte des objectifs politiques et militaires.

Source : Larousse et rédacteurs.

Information

Action de porter des nouvelles à la connaissance du public, de faire part des événements, des faits marquants de l'actualité.

Élément de connaissance traduit par un ensemble de signaux selon un code déterminé, en vue d'être conservé, traité ou communiqué.

Guerre de l'information : Dans la guerre de l'information, l'information constitue un enjeu à acquérir et à protéger. Elle représente aussi une arme pour conquérir, dissuader, déstabiliser, convaincre les autres acteurs de prendre des décisions qui soient favorables à celui qui prend l'initiative. « L'information c'est la différence qui fait la différence. » (Gregory Bateson)

Source : Académie, Wikipedia, Oxford dictionary et rédacteurs.

Ingérence & Contre-ingérence

Action de s'ingérer dans les affaires d'autrui. Ingérer : S'immiscer dans une affaire indûment ou sans titre, sans en être requis.

L'ingérence numérique étrangère est un phénomène inauthentique affectant le débat public numérique qui combine :

1. Une atteinte potentielle aux intérêts fondamentaux de la Nation ;
2. Un contenu manifestement inexact ou trompeur ;
3. Une diffusion artificielle ou automatisée, massive et délibérée ;
4. L'implication, directe ou indirecte, d'un acteur étranger (étatique, paraétatique ou non étatique).

Ingérences étrangères dans l'espace de l'information : efforts coercitifs et trompeurs déployés par un acteur d'un État étranger ou des agents de celui-ci dans le but d'entraver la formation et l'expression libres de la volonté politique des individus.

Contre-ingérence : s'opposer à des actions représentant une menace. Les agissements en cause peuvent être délibérément ou non dirigés contre les intérêts d'un État et émaner de toutes sortes d'acteurs : un autre État, une entreprise, une organisation non gouvernementale, un groupe criminel ou terroriste, des individus isolés.

Source : Académie, rédacteurs, Viginum et Commission européenne (Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions relative au plan d'action pour la démocratie européenne, 3 décembre 2020, COM(2020) 790 final, pp. 21-22.).

Intégrité

État d'une chose qui est dans son entier, qui n'est pas entamée ou altérée.

(Informatique) Propriété d'exactitude et de complétude des biens et informations (i.e. une modification illégitime d'un bien doit pouvoir être détectée et corrigée)

Intégrité des données : fait référence à la fiabilité et à la crédibilité des données durant tout leur cycle de vie. Elle peut être représentative de l'état des données (valides ou non) ou du processus visant à garantir et préserver la validité et l'exactitude des données.

Dans le cadre d'un processus, le contrôle et la validation des erreurs sont, par exemple, des méthodes courantes de protection de l'intégrité des données.

Source : Académie, ANSSI, rédacteursd.

Lii – L2I - Lutte informatique d'influence

La lutte informatique d'influence (L2I) désigne les opérations militaires conduites dans la couche informationnelle du cyberspace pour détecter, caractériser et contrer les attaques, appuyer la StratCom, renseigner ou faire de la déception, de façon autonome ou en combinaison avec d'autres opérations.

Source : ministère de la défense, doctrine L2I.

Man-in-the-middle (attaque)

Une attaque de type « Man in The Middle » (MiTM) est une attaque dans laquelle l'agresseur intercepte et relaie secrètement des messages entre deux parties qui croient communiquer directement l'une avec l'autre. L'attaque est une forme d'écoute clandestine dans laquelle l'attaquant contrôle l'ensemble de la conversation et peut interagir avec un des participants afin de perturber les échanges à son profit.

Source : actualiteinformatique.fr

Manipulation (de l'information) – Désinformation – Méinformation (angle juridique)

Manipulations de l'information : Actions informationnelles délibérées (qui supposent l'intention de nuire) et clandestines (les victimes en sont inconscientes) dans l'objectif de créer des effets politiques et/ou militaires.

Manipulation de l'information : diffusion intentionnelle et massive de nouvelles fausses ou biaisées à des fins politiques hostiles. Les auteurs évitent volontairement de recourir à l'expression galvaudée de « fake news » jugée trop vague et polémique.

Malinformation : La malinformation est une information qui se fonde sur la réalité, mais qui est utilisée pour porter préjudice à une personne, une organisation ou un pays. Désinformation : des contenus faux ou trompeurs diffusés avec l'intention de tromper ou dans un but lucratif ou

politique et susceptibles de causer un préjudice public
 Méinformation : des contenus faux ou trompeurs transmis sans intention de nuire, même si leurs effets peuvent néanmoins être préjudiciables ; c'est notamment le cas lorsque des personnes partagent de bonne foi de fausses informations avec des amis ou des membres de leur famille.

Source : un.org et Commission européenne, mediadefence.org

Média – médium – mass-media

- Médium : Support ; intermédiaire.
- Le terme média désigne tout moyen de distribution, de diffusion ou de communication interpersonnelle, de masse ou de groupe, d'œuvres, de documents, ou de messages écrits, visuels, sonores ou audiovisuels (comme la radio, la télévision, le cinéma, Internet, réseaux sociaux, la presse, les télécommunications, etc.). Ce terme est souvent utilisé comme l'abréviation du terme anglais mass-media ou médias de masse en français.

Source : wikipedia.

Mème

Concept (texte, image, vidéo) massivement repris, décliné et détourné sur l'Internet de manière souvent parodique, qui se répand très vite, créant ainsi le buzz.

Source : Larousse.

Nudge

La théorie du nudge (ou théorie du paternalisme libéral) est un concept des sciences du comportement, de la théorie politique et d'économie issu des pratiques de design industriel qui fait valoir que des suggestions indirectes peuvent, sans forcer, influencer les motivations et inciter à la prise de décision des groupes et des individus, de manière au moins aussi efficace que l'instruction directe, la législation ou l'exécution.

Source : wikipedia.

Opinion

Jugement, avis, sentiment qu'un individu ou un groupe émet sur un sujet, des faits, ce qu'il en pense ; Ensemble des idées d'un groupe social sur les problèmes politiques, économiques, moraux, etc.

Source : Larousse.

Perception

Action de percevoir par les organes des sens ; idée, compréhension plus ou moins nette de quelque chose ; événement cognitif dans lequel un stimulus ou un objet, présent dans l'environnement immédiat d'un individu, lui

est représenté dans son activité psychologique interne, en principe de façon consciente ; fonction psychologique qui assure ces perceptions.

La manœuvre des perceptions, grâce à des actions d'influence efficaces, vise à modifier le sentiment général d'une cible par rapport à une situation ou un événement donnés.

Source : Larousse.

Plate-forme

Une plate-forme est, en informatique, une base de travail à partir de laquelle on peut écrire, lire, utiliser, développer un ensemble de logiciels et sites.

Une « plateforme de formation et d'entraînement à la sécurité numérique » est un ensemble de « ressources » matérielles ou immatérielles (moyens techniques, services, contenus, ressources humaines) permettant d'appréhender de manière générale tous les besoins liés à la formation et à l'entraînement en cybersécurité. Elle permet de répondre à des besoins de montée en compétences et de validation de savoir et savoir-faire (sensibilisation, formation, entraînement, exercice).

Source : wikipedia et ANSSI.

Populisme

Attitude, comportement d'un homme ou d'un parti politique qui, contre les élites dirigeantes, se pose en défenseur du peuple et en porte-parole de ses aspirations, avançant des idées le plus souvent simplistes et démagogiques.

Idéologie (ou mouvement politique) qui fait la promotion du «peuple» -imaginaire ou réel, majoritaire ou identitaire- en développant un discours fondé sur une triple méfiance :

1. à l'endroit de certaines élites (partis, députés, fonctionnaires) ;
2. à l'endroit d'un prétendu système caché (complot) qui trahirait les intérêts fondamentaux du peuple ;
3. à l'endroit d'entités ou de mouvances internationales -entreprises, organisations, migrations, etc.

Associé parfois à la droite et parfois à la gauche, le populisme est aussi un style politique : doctrine simple, chef fort et charismatique, organisations de masse structurées.

Source : Académie, université de Sherbrooke.

Renseignement – OSINT

Un renseignement est une information estimée pour sa valeur et sa pertinence. Le renseignement se définit ainsi par opposition à la donnée brute et au fait. Le renseignement est fondé sur la recherche, l'analyse et l'exploitation de données d'origine très variée, recoupées, idéalement d'origines techniques et humaines. Il est fourni à des commanditaires (gouvernements, institutions) pour guider des prises de décision et des actions.

Le renseignement de sources ouvertes ou renseignement d'origine sources ouvertes (acronyme ROSO, en anglais open source intelligence, OSINT) est un renseignement obtenu par une source d'information publique. D'une manière générale, l'OSINT désigne « un ensemble hétéroclite de pratiques d'investigation et d'analyse visant à dévoiler une information préalablement dissimulée en récoltant, croisant ou analysant des données numériques disponibles en source ouverte ».

Source : wikipedia.

Réseaux sociaux

Site Web ou application mobile qui permet aux utilisateurs de se constituer un réseau d'amis, de relations ou d'abonnés, et qui favorise les interactions sociales entre les individus, groupes d'individus ou organisations. Les réseaux sociaux appartiennent à la famille des médias sociaux.

Source : Mercator.

Troll, ferme ou usine à trolls

En argot Internet, un troll caractérise un individu recherchant l'attention d'un public par la création de ressentis négatifs, ou par un comportement qui vise à générer des polémiques. Il peut s'agir d'un message, par exemple sur un forum, d'un débat conflictuel dans son ensemble, surtout politique, ou plus couramment de la personne qui en est à l'origine.

Une usine à trolls (en anglais « troll factory »), ou une ferme à trolls (de l'anglais « troll farm »), est une organisation qui regroupe et coordonne des trolls sur l'Internet, voire des hackers payés, ou des intelligences artificielles (trolls virtuels) programmées pour diffuser de manière massive des informations partielles, partiales ou totalement mensongères sur les réseaux sociaux. Leur but est la déstabilisation géopolitique ou politique, le lobbying ou la propagande politique. Ces usines à trolls peuvent soit être issues d'initiatives privées, soit être constituées et coordonnées par un gouvernement, un parti politique ou tout autre groupe de pression. Les méthodes utilisées incluent la rémunération de commentateurs sur les réseaux sociaux, le trolling agressif de journalistes, ou la création de bots informatiques et de médias diffusant massivement des informations mensongères.

Source : wikipedia.

Lexique rédigé en collaboration avec :

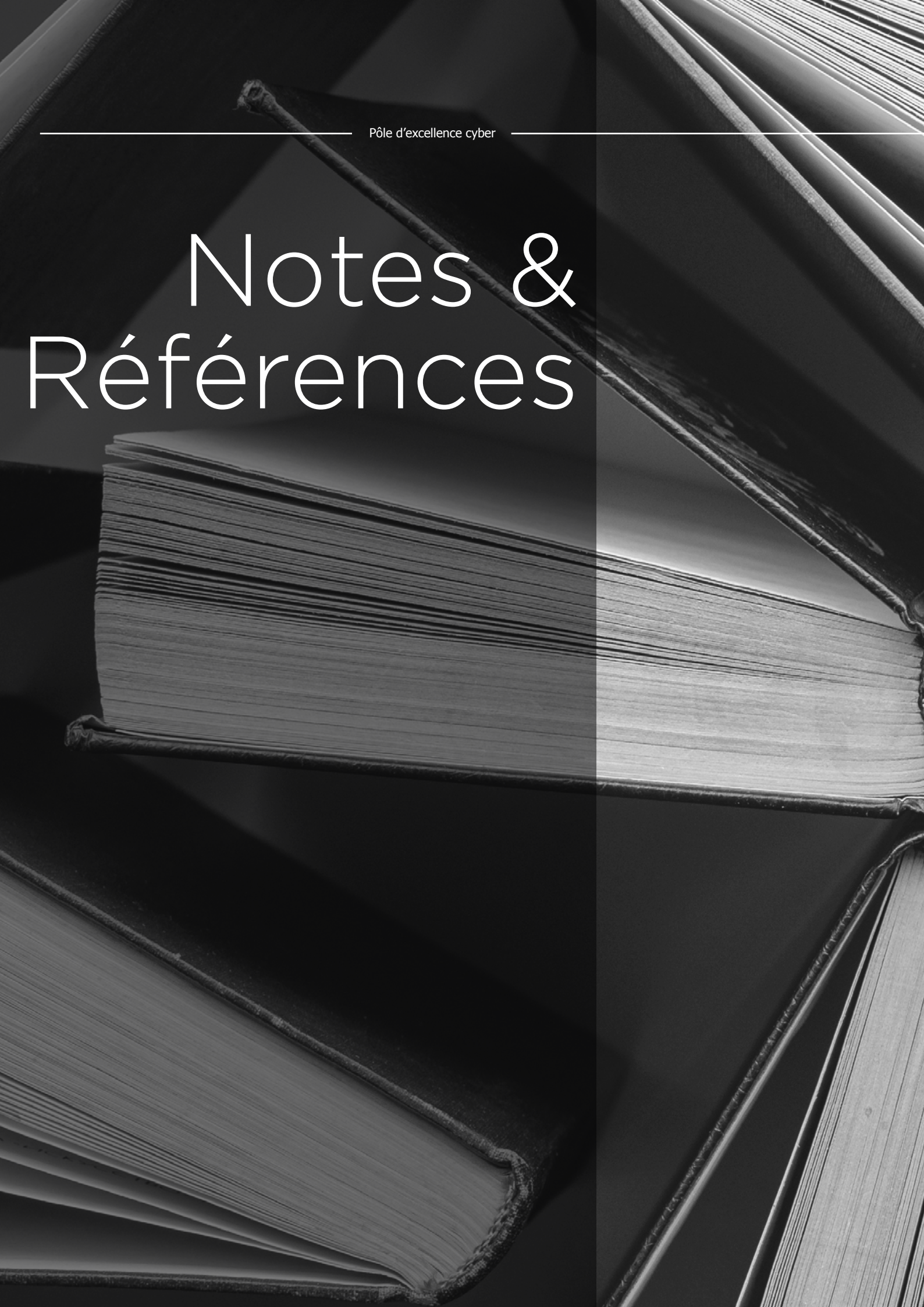
Dr. Jean-Philippe Riant
Directeur conseil en intelligence
artificielle et communication digitale

Patrick Erard
délégué général adjoint
Pôle d'excellence cyber

Général (2S) Bruno Courtois
Conseiller Défense et Cyber
Sopra Steria

Pôle d'excellence cyber

Notes & Références





La lutte contre la manipulation de l'information : définitions et concepts clefs (p.12)

- (1) - « Exemples de campagnes de manipulation de l'information ». Quai d'Orsay [consulté le 19/12/2022]. Disponible sur : <https://disinfo.quaidorsay.fr/encyclopedia/introduction/exemples/fr>
- (2) - MOREIRA Paul. 5 mars 2007. LAFFONT Robert, « Les nouvelles censures dans les coulisses de la manipulation de l'information », 288 pages
- (3) - DE LESPINOIS Jérôme. « La cyberstratégie en tant que stratégie du milieu », Mooc cours de stratégie de l'École de Guerre (Partie II), 28/11/2022.
- (4) - « Lutter contre les manipulations de l'information en temps de crise » [en ligne]. Veille, le 14/10/2022. Disponible sur : https://www.veillemag.com/Agenda-Lutter-contre-les-manipulations-de-l-information-en-temps-de-crise--le-14-octobre-2022-a-Paris_a4504.html
- (5) - « Viginum : des cas d'ingérences étrangères en ligne détectés lors des élections de 2022 » [en ligne]. Vie publique, le 28/10/2022. Disponible sur : <https://www.vie-publique.fr/en-bref/286939-viginum-cas-dingerences-etrangeres-en-ligne-detectes-elections-202>

Exploitation du désir de confort

La victoire de l'entre-soi et mise en danger de la démocratie (p.18)

Deux lectures pour aller plus loin :

- Propaganda « Comment manipuler l'opinion en démocratie » d'Edward Bernays, ouvrage qui expose les grands principes de la manipulation mentale de masse - Editions Zones Déc 2021
- Toxic Data de David Chavalarias « Comment les réseaux manipulent nos opinions » - Editions Flammarion mars 2022.

Enseignement supérieur et désinformation : quel bilan dans les écoles d'ingénieur ? (p.20)

- (1) - Bojjireddy, Sirisha, Soon Ae Chun, et James Geller. 2021. « Machine Learning Approach to Detect Fake News, Misinformation in COVID-19 Pandemic ». In DG.O2021: The 22nd Annual International Conference on Digital Government Research, DG.O'21, New York, NY, USA: Association for Computing Machinery, 575-78. <https://dl.acm.org/doi/10.1145/3463677.3463762>.

- (2) - Info, France. 2023. « [PODCAST] TikTok, Roblox, Fortnite : la jeunesse dans le viseur des complotistes ». Conspiracy Watch | L'Observatoire du conspirationnisme. <https://www.conspiracywatch.info/podcast-tiktok-roblox-fortnite-la-jeunesse-dans-le-viseur-des-complotistes.html>.
- (3) - Gupta, Ankur et al. 2022. « Combating Fake News: Stakeholder Interventions and Potential Solutions ». IEEE Access 10: 78268-89.
- (4) - de Oliveira, Nicollas R. et al. 2021. « Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges ». Information 12(1): 38.
- (5) - Directorate-General for Parliamentary Research Services (European Parliament), Trisha Meyer, et Chris Marsden. 2019. Regulating Disinformation with Artificial Intelligence: Effects of Disinformation Initiatives on Freedom of Expression and Media Pluralism. LU: Publications Office of the European Union. <https://data.europa.eu/doi/10.2861/003689>.
- (6) - Wardle, Claire, et Hossein Derakhshan. 2018. Les désordres de l'information : Vers un cadre interdisciplinaire pour la recherche et l'élaboration des politiques. Strasbourg: Editions du Conseil de l'Europe. <https://rm.coe.int/rapport-les-desordres-de-l-information-/1680935bd4>.
- (7) - « La mésinformation scientifique des jeunes à l'heure des réseaux sociaux ». Fondation Jean-Jaurès. <https://www.jean-jaures.org/publication/la-mesinformation-scientifique-des-jeunes-a-lheure-des-reseaux-sociaux/>.
- (8) - Goetgheluck, Nicole Devillard et al. 2019. « Atelier : Travailler des compétences transversales en information : une nécessité avec les étudiants d'aujourd'hui ». <https://hal.science/hal-03133967>.
- (9) - « Journalisme, fake news & désinformation : manuel pour l'enseignement et la formation en matière de journalisme - UNESCO Bibliothèque Numérique ». <https://unesdoc.unesco.org/ark:/48223/pf0000372695>.
- (10) - Latour, Bruno. 2007. « La cartographie des controverses ». Technology Review, N.O, pp. 82-83.
- (11) - Badouard, Romain. 2020. « Chapitre 1. Fake news, complotisme, désinformation : quels enjeux pour l'éducation aux médias ? » In Éducation critique aux médias et à l'information en contexte numérique, Papiers, éd. Sophie Jehel et Alexandra

Saemmer. Villeurbanne : Presses de l'enssib, 27-36.
<http://books.openedition.org/pressesenssib/11143>.

Moyens technologiques actuels pour lutter contre les manipulations d'informations (p.24)

- (1) - <https://factuel.afp.com/>
- (2) - <https://twitter.com/AirbusDefence/status/1536388025655209984>
- (3) - <https://www.claimreviewproject.com/>
- (4) - <https://tineye.com/>
- (5) - https://www.bellingcat.com/app/uploads/2015/05/Forensic_analysis_of_satellite_images_EN.pdf
- (6) - Projets Invid <https://www.invid-project.eu/> et We Verify <https://weverify.eu/about-us/overview/>
- (7) - <https://www.m82-project.com/post/disarm-une-matrice-pour-d%C3%A9criver-les-campagnes-d-influence>

Les bots sociaux et la manipulation de l'information (p.32)

- (1) - S. C. Woolley et P. N. Howard (dir.), Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media, New York, Oxford University Press, 2018.
- (2) - TwiBot-22 - <https://arxiv.org/pdf/2206.04564.pdf> - 2022
- (3) - SGBot - <https://arxiv.org/pdf/1911.09179.pdf> - 2019
- (4) - DeeProBot - <https://link.springer.com/article/10.1007/s13278-022-00869-w> - 2022
- (5) - RoBERTa - <https://arxiv.org/abs/1907.11692> - 2019
- (6) - T5 - <https://arxiv.org/abs/1910.10683> - 2019
- (7) - Kouvela et al. - <https://dl.acm.org/doi/abs/10.1145/3415958.3433075> - 2020
- (8) - BotRGCN - <https://arxiv.org/abs/2106.13092> - 2021

Usurpation d'identité par le biais des Deep Fakes dans le Metaverse (p.38)

- (1) - Nicolas M. Müller, Karla Pizzi, and Jennifer Williams. 2022. Human Perception of Audio Deepfakes. In Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia (DDAM '22). Association for Computing Machinery, New York, NY, USA, 85-91. <https://doi.org/10.1145/3552466.3556531>
- (2) - F. Kilger, A. Kabil, V. Tippmann, G. Klinker and M. -O. Pahl, "Detecting and Preventing Faked Mixed Reality," 2021 IEEE 4th International Conference on

Multimedia Information Processing and Retrieval (MIPR), Tokyo, Japan, 2021, pp. 399-405, doi: 10.1109/MIPR51284.2021.00074

- (3) - Pooyandeh M, Han K-J, Sohn I. Cybersecurity in the AI-Based Metaverse: A Survey. Applied Sciences. 2022; 12(24):12993.
<https://doi.org/10.3390/app122412993>

L'Union européenne et la lutte contre les campagnes de manipulation de l'information dans le contexte du conflit armé en Ukraine (p.44)

Actes du Conseil de l'Union européenne

- Décision (PESC) 2022/351 du Conseil du 1er mars 2022 modifiant la décision 2014/512/PESC concernant des mesures restrictives eu égard aux actions de la Russie déstabilisant la situation en Ukraine, JOUE n° L 65, 2 mars 2022, p. 5-7.
- Règlement (UE) 2022/350 du Conseil du 1er mars 2022 modifiant le règlement (UE) no 833/2014 concernant des mesures restrictives eu égard aux actions de la Russie déstabilisant la situation en Ukraine, JOUE n° L 65, 2 mars 2022, p. 1-4.
- Décision (PESC) 2022/884 du Conseil du 3 juin 2022 modifiant la décision 2014/512/PESC concernant des mesures restrictives eu égard aux actions de la Russie déstabilisant la situation en Ukraine, JOUE n° L 153, 3 juin 2022, p. 128-138.
- Règlement (UE) 2022/879 du Conseil du 3 juin 2022 modifiant le règlement (UE) no 833/2014 concernant des mesures restrictives eu égard aux actions de la Russie déstabilisant la situation en Ukraine, JOUE n° n° L 153 du 3 juin 2022, p. 53-74.

Arrêt de la Cour de justice de l'Union européenne

- Trib. UE (ord.), 30 mars 2022, RT France c. Conseil, aff. T-125/22, non encore publiée.
- Trib. UE, 27 juillet 2022, RT France c. Conseil, aff. T-125/22, ECLI:EU:T:2022:483.

Actions du Service européen pour l'action extérieure

https://www.eeas.europa.eu/countering-disinformation/tackling-disinformation-information-work-eeas-strategic-communication_en
<https://euvsdisinfo.eu/fr/>

Articles :

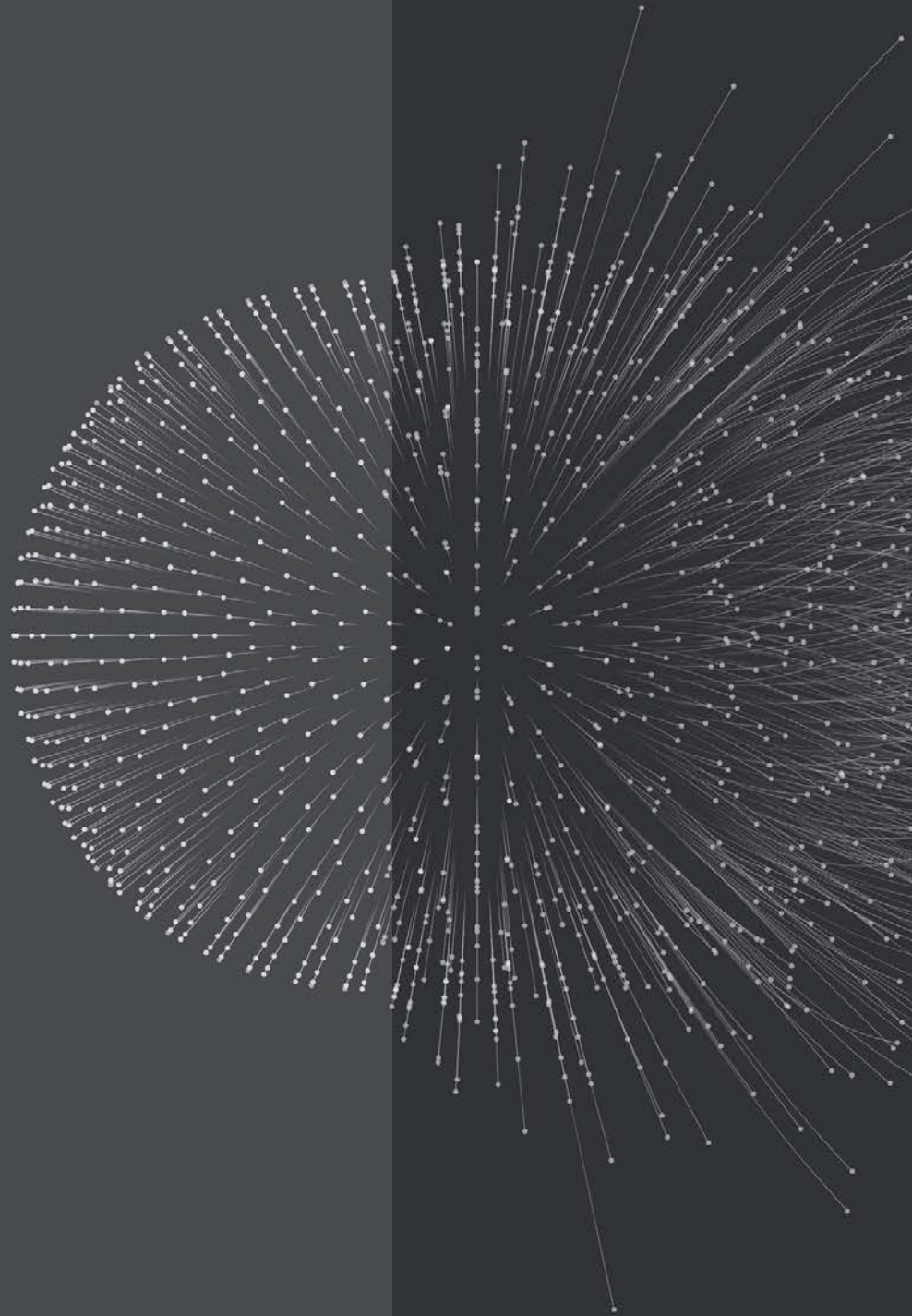
<https://www.euractiv.fr/section/economie/news/les-medias-lies-au-kremlin-contournent-les-sanctions-de-lue-selon-un->
<https://www.euractiv.fr/section/economie/news/meta-demantele-des-campagnes-de-desinformation-menees-par-la-russie-et-la-chine/>

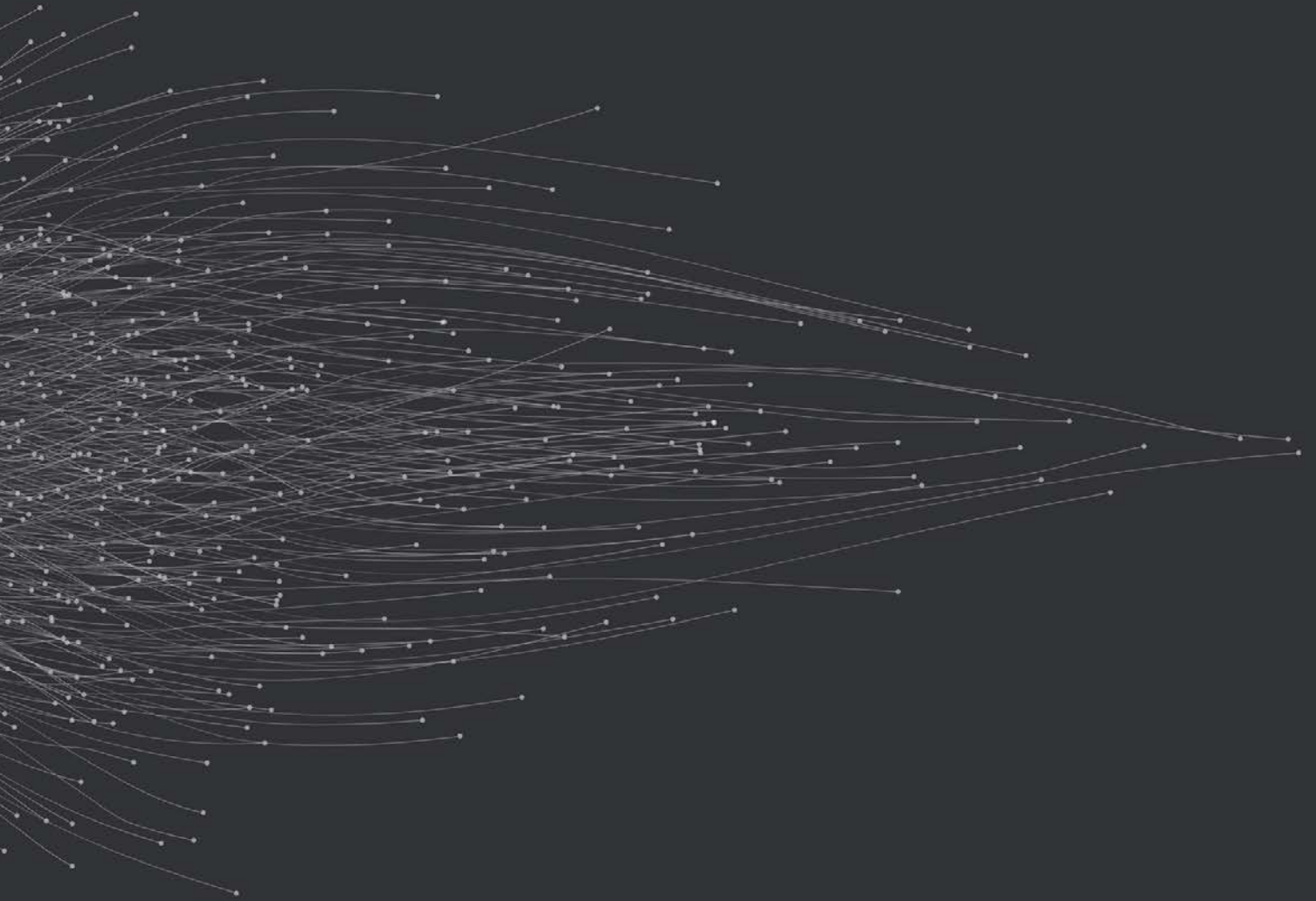
Dispositifs étatiques pour lutter contre la manipulation d'information (p.48)

Références :

- (1) - Loi n°2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information. Journal officiel n°0297 du 23 décembre 2018. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559>.
- (2) - Décret n° 2021-922 du 13 juillet 2021 portant création, auprès du secrétaire général de la défense et de la sécurité nationale, d'un service à compétence nationale dénommé « service de vigilance et de protection contre les ingérences numériques étrangères ». Journal officiel n°0162 du 14 juillet 2021. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043788361>.
- (3) - Décret n° 2021-1587 du 7 décembre 2021 portant autorisation d'un traitement automatisé de données à caractère personnel dans le but d'identifier les ingérences numériques étrangères. Journal officiel n°0286 du 9 décembre 2021. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044454057>.
- (4) - Conclusions du Conseil européen des 19 et 20 mars 2015, §13 <https://www.consilium.europa.eu/media/21888/european-council-conclusions-19-20-march-2015-en.pdf>.
- (5) - Global Engagement Center. Département d'État, Sous-secrétaire à la diplomatie publique et aux affaires publiques, 2016. <https://www.state.gov/bureaus-offices/under-secretary-for-public-diplomacy-and-public-affairs/global-engagement-center/>.
- (6) - Éléments publics de doctrine militaire de lutte informatique d'influence (L2I). Ministère des Armées, Commandt de la cyberdéfense. <https://www.defense.gouv.fr/ema/actualites/armees-se-dotent-dune-doctrine-militaire-lutte-informatique-dinfluence-l2i>.
- (7) - Déclarations 2022 des opérateurs de plateformes en ligne (dailymotion, google, linkedin, meta, microsoft, pinterest, snapchat, tiktok, twitter, webedia, wikimedia, yahoo !). Site de l'ARCOM. <https://www.arcom.fr/vos-services-par-media/internet-et-reseaux-sociaux/lutte-contre-la-manipulation-de-linformation-declarations-des-operateurs-de-plateformes-en-ligne-et-questionnaires-de-larcom>.
- (8) - Les manipulations d'information, un défi pour nos démocraties. IRSEM, ministère des Armées ; CAPS, ministère de l'Europe et des Affaires étrangères, 2018. https://www.diplomatie.gouv.fr/IMG/pdf/les_manipulations_de_l_information_2_cle04b2b6.pdf.
- (9) - Les opérations d'influence chinoises : un moment machiavélien. IRSEM, ministère des Armées, sept. 2021. <https://www.irsem.fr/rapport.html>.
- (10) -Rapport d'activité de l'année #1. Service VIGINUM, 2022. <http://www.sgdsn.gouv.fr/uploads/2022/10/20221025-viginum-annee1.pdf>.

Conclusion







Conclusion



Jean-Luc GIBERNON
Vice-président
Développement Industriel
du Pôle d'excellence cyber



Dr. Jean-Philippe Riant
Directeur conseil
en intelligence artificielle
et communication digitale



Il est assez inévitable que cet ouvrage possède un certain côté anxiogène. Nous avons voulu élaborer cet ouvrage pour d'une part sensibiliser et d'autre part bien saisir toutes les dimensions de la lutte qui s'opère. Mais au-delà de la prise de conscience et de la bonne compréhension des mécanismes de la désinformation, aussi importantes soient-elles, il s'agit désormais de lancer des actions concrètes.

A travers les réflexions échangées lors de la rédaction de cet ouvrage, notre groupe de travail a pu esquisser les premières pistes pour aller vers des actions concrètes de façon à construire une réponse aux enjeux de la lutte.

Tout d'abord, il nous semble indispensable de mettre en place un observatoire de la désinformation, chargé d'apporter un éclairage à la fois opérationnel et technologique. Puisque l'échelle de l'Europe est le bon niveau, ce dispositif devra se positionner d'emblée au niveau européen. Il pourrait tout à la fois éclairer sur les actualités du domaine, publier des retours d'expérience et des études, organiser des événements et conférences, mettre du matériel pédagogique à la disposition du grand public ou des journalistes.

Ensuite des sensibilisations destinées au grand public devraient être largement mises en place. La prise de conscience de nos concitoyens peut être accompagnée par des rencontres terrain pour que les mécanismes de la désinformation soient très largement partagés. Certains projets en ce sens existent déjà dans le domaine plus large de l'éducation au numérique, il s'agit de les faire accélérer.

Il nous est clairement apparu que les réseaux sociaux eux-mêmes doivent également être des promoteurs de la lutte contre la désinformation. Travailler avec les principaux réseaux sociaux sera donc indispensable, avec leur coopération. Pour cela, les outils juridiques devront jouer pleinement leur rôle.

D'un point de vue général, rien ne sera possible sans mener des actions fortes dans la dimension juridique, que ce soit au niveau français et au niveau européen. Le droit est aujourd'hui insuffisant ou trop imprécis, les instruments juridiques français et européens doivent être renforcés, à la fois pour donner des instruments de contrôle et de répression.

Par ailleurs, il est nécessaire de développer des outils et des plates-formes pour aider pour instrumenter la lutte contre la désinformation. Ces outils peuvent être par exemple des outils de détection de faux : détection de fausses vidéos, d'images retouchées, voire créées de toute pièce par un logiciel. D'autres outils seront nécessaires, il s'agit maintenant de pousser plus loin les réflexions et d'étudier les outils techniques à développer. La recherche technique et opérationnelle dans ce domaine mérite d'être développée et aidée par des subventions, elle permettra le développement de concepts, de plates-formes, d'outils logiciels, etc. L'intelligence artificielle doit en être un champ important : si les technologies de l'IA permettent de créer des « fake », elle permet aussi de les détecter.

Enfin, pour accompagner les actions de long terme, la création de chaires sur des thèmes techniques et sur des thèmes juridiques doit être envisagée.

Le Pôle d'excellence cyber estime qu'il est essentiel de lancer immédiatement ces actions, de façon à se donner les capacités de mener cette lutte contre les manipulations de l'information. Il est important de garder en objectif une perspective européenne, parce que c'est à cette échelle que les actions seront réellement efficaces.

La France peut devenir le moteur d'une dynamique européenne. La lutte sera de longue haleine, il est donc important de la démarrer rapidement.

PÔLE D'EXCELLENCE
CYBER

www.pole-excellence-cyber.org

