Lettre mensuelle – **Janvier 2021** – **Disponible sur omc.ceis.eu**

Table des matières

ANALYSES (1/2)

SOLARWINDS : UNE ATTAQUE SANS PRÉCÉDENT ?

Au-delà des campagnes de *phishing* et de *ransomware* qui ont accompagné la crise sanitaire de 2020, l'année a également été marquée par l'attaque qui a visé l'entreprise SolarWinds au mois de décembre, touchant un nombre encore difficile à évaluer de près de 350 000 clients dont plusieurs agences gouvernementales des États-Unis. D'après la société FireEye, première victime confirmée, une multitude d'organisations, tant des administrations publiques que des multinationales, se sont déclarées touchées : départements stratégiques du gouvernement américain (défense, contrôle aérien, Trésor), Microsoft, FireEye, Cisco, Nvidia, Intel… et il est à craindre que cette liste ne cesse de s'allonger tant les utilisateurs des solutions de SolarWinds sont nombreux. En France, de nombreuses sociétés, dont plusieurs du CAC40, y ont par exemple recours.

Au-delà de son ampleur, si cette attaque a tant fait parler d'elle, c'est aussi parce qu'elle a rappelé, un peu brutalement, les enjeux des attaques sur la chaîne d'approvisionnement ou « *supply chain* », qui permettent aux assaillants d'atteindre des organisations réputées bien protégées en passant par leurs fournisseurs ou par leurs sous-traitants, qui en général le sont moins.

Une attaque d'une ampleur encore difficile à évaluer

En décembre 2020, la société FireEye a dévoilé une campagne d'espionnage numérique dont elle a été victime et qui aurait débuté en mars 2020, rendue possible par la compromission d'une mise à jour de la plateforme de gestion et de supervision Orion développée par SolarWinds. Ces mises à jour, contenant une *backdoor*, ont été téléchargées et installées par plus de 18 000 clients dans le monde (sur un total de 33 000), et ont permis d'introduire dans le code source du logiciel un code malveillant – une porte dérobée appelée « Sunburst », qui a permis aux attaquants d'accéder aux systèmes informatiques des victimes et à leurs données. Les dates de signatures des clients ayant téléchargé et installé les mises à jour s'étalent de mars à juin 2020. Elles indiquent donc que les attaquants ont eu six mois pour extraire des données ou poser des charges malveillantes en vue de compromissions ultérieures.

C'est bien le propre d'une attaque sur la *supply chain*, qui consiste à attaquer un acteur potentiellement moins bien défendu pour ensuite atteindre une cible réputée bien protégée qui lui est liée. Dans ce cas, c'est en s'attaquant à SolarWinds, fournisseur de géants industriels et d'agences gouvernementales, que les auteurs de cette campagne ont pu toucher des sociétés comme FireEye, Microsoft ou encore le département du Trésor américain. Cette technique permet aux attaquants de « rentabiliser » leurs efforts en ne compromettant qu'une seule cible facile d'accès, pour pénétrer ensuite le réseau de tous les clients de cette première cible.

La liste des victimes dressée à ce jour comprend des agences gouvernementales (18%), des entreprises de sécurité et autres entreprises technologiques (44%) et des organisations non gouvernementales[1].

Parmi les victimes institutionnelles figurent plusieurs agences gouvernementales telles que le département du Trésor américain, l'Administration nationale des télécommunications et de l'information (NTIA), des instituts nationaux de la santé (NIH), l'Agence de la cybersécurité et des infrastructures (CISA) ou le département de la Sécurité intérieure (DHS). Le Cyber Command et la NSA avaient installé des systèmes d'alerte précoce dans les réseaux étrangers pour détecter ce type d'attaque. Des systèmes qui, vraisemblablement, n'auraient

[1] « Piratage SolarWinds : la liste des victimes et des failles de sécurité s'allonge », *Channel News* [En ligne], 21 décembre 2021.

pas fonctionné[2]. L'agence du département de l'Énergie chargée de gérer le stock d'armes nucléaires a aussi été visée par les attaques. Dans une directive d'urgence publiée le 14 décembre, la CISA a ordonné la déconnexion ou la mise hors tension rapide des produits SolarWinds concernés des réseaux fédéraux[3]. Une unité de coordination mise au point par la CISA, le FBI et le directeur du renseignement est aujourd'hui chargée d'élaborer une réponse coordonnée du gouvernement en matière d'investigation et de remédiation [4].

Du côté des entreprises touchées, Microsoft a annoncé avoir repéré sur ses systèmes des versions d'Orion compromises, et a par la suite indiqué avoir détecté des traces confirmant que leur code source avait été en partie accédé – mais pas modifié[5]. L'entreprise a tout de même souhaité rassurer ses utilisateurs en expliquant que « *la sécurité de* [leurs] *produits ne dépend pas du secret du code source* ». Microsoft a en revanche révélé avoir trouvé des versions malveillantes de Orion dans ses systèmes affectant une quarantaine de ses clients. Si environ 80% de ces derniers se trouvent aux États-Unis, Microsoft a également identifié des victimes au Canada, au Mexique, en Belgique, en Espagne, au Royaume-Uni et en Israël.

Au-delà des victimes déjà identifiées, SolarWinds compte parmi ses clients plus de 425 des plus grandes entreprises américaines, d'importantes entreprises françaises du CAC 40 et plusieurs multinationales. Le nombre de victimes ne cesse de s'accroître, même un mois après l'identification de l'attaque, et il y a fort à parier que la liste de sociétés touchées continuera d'augmenter dans les mois à venir. En outre, il est probable que plusieurs mois de recul ne soient nécessaires pour comprendre l'étendue de la compromission, et encore plus longtemps pour chasser les pirates des réseaux des organisations victimes.

Une attribution délicate

Le niveau de complexité de l'attaque évoque clairement un groupe soutenu par un gouvernement plutôt qu'un réseau de cybercriminels. Microsoft a par ailleurs confirmé que les méthodes utilisées évoquaient un acteur étatique, sans pour autant désigner de pays. Selon Kaspersky, la porte dérobée utilisée pour compromettre les clients de SolarWinds ressemble au logiciel malveillant Kazuar, utilisé par le groupe de pirates Turla, qui, selon les autorités estoniennes, opère au nom du FSB russe[6]. Si certaines hypothèses pointent vers des pirates d'origine russe, tels que le groupe APT29 (ou Cozy Bear), d'autres mentionnent des attaquants chinois ou nord-coréens. Certains estiment que l'attaque a pu être organisée sur le sol américain afin d'éviter les soupçons des autorités, ou depuis les bureaux de SolarWinds dans des pays d'Europe de l'Est tels que la Biélorussie, la République tchèque et la Pologne[7].

Le gouvernement américain a de son côté directement accusé la Russie et a rappelé qu'il se réservait le droit de répondre à l'attaque. Si l'ambassade de Russie aux États-Unis a rapidement réfuté ces accusations, le National Coordination Center for Computer Incidents russe (NKTsKI) a néanmoins émis une alerte sur de potentielles représailles américaines sur des organisations russes. On peut toutefois douter de la volonté des États-Unis de dénoncer ce qu'ils présentent être du cyberespionnage, une pratique largement utilisée de leur côté. Et ce d'autant plus qu'attribuer avec certitude une cyberattaque reste difficile. Lorsque des attaquants russes ont perturbé la cérémonie d'ouverture des Jeux Olympiques d'hiver en 2018 par exemple, ils ont

[2] « As Understanding of Russian Hacking Grows, So Does Alarm », *The New York Times* [En ligne], 5 janvier 2021.
[3] « Joint Statement by the FBI, the CISA, the ODNI, and the NSA », CISA, 5 janvier 2021.
[4] Ibid.
[5] « Microsoft Internal Solorigate Investigation Update », *Microsoft Security Response Center* [En ligne], 3 janvier 2021.
[6] « SolarWinds hackers linked to known Russian spying tools, investigators say », *Reuters* [En ligne], 11 janvier 2021.
[7] « Le piratage de SolarWinds pourrait être bien plus grave que ce que le FBI imaginait », *Siècle Digital* [En ligne], 3 janvier 2021.

délibérément imité un groupe nord-coréen. Dans le cas de SolarWinds, il est possible que les deux logiciels malveillants aient été déployés par le même groupe ou que Kazuar ait inspiré les pirates de SolarWinds.

Les motivations des attaquants soulèvent cependant encore de nombreuses questions. Ils semblent en effet procéder à la revente du fruit de leur attaque, de sorte à laisser penser que leurs motivations seraient d'abord pécuniaires. Le site SolarLeaks.net hébergerait ainsi des données volées aux victimes de l'attaque. Les propriétaires du site prétendent détenir des données provenant de Microsoft, FireEye, Cisco et SolarWinds. Ils proposent par exemple les données de Microsoft sous la forme d'un fichier de 2,6 Go pour 600 000 dollars, celles de Cisco pour 500 000 dollars (1,7 Go), de SolarWinds pour 250 000 dollars (612 Mo) et de FireEye pour 50 000 dollars (39 Mo)[8]. Les supposés attaquants proposent également un « pack » contenant l'intégralité des données dérobées pour la somme d'un million de dollars. Ils précisent par ailleurs que de nouvelles données suivront prochainement. Cependant selon le fondateur de Rendition Infosec, Jake Williams, leur prix est « *fantaisiste*[9] », ce qui tendrait à confirmer que le gain financier n'était pas le motif principal de cette attaque et qu'il s'agissait plutôt d'une campagne d'espionnage à des fins politiques ou géopolitiques.

La *supply chain* est-elle le maillon faible ?

L'affaire Solarwinds rappelle que les attaques sur la *supply chain* constituent aujourd'hui l'une des principales menaces auxquelles toute organisation doit faire face. Son « succès » pourrait d'ailleurs accélérer la tendance des groupes malveillants les plus avancés à privilégier ce type d'attaque indirecte. Elle rappelle aussi que nulle organisation n'est à l'abri, même la mieux protégée, si ses fournisseurs et ses sous-traitants ne le sont pas. Le serveur de mises à jour d'Orion Platform était par exemple protégé par un mot de passe on ne peut plus faible : « solarwinds123[10] ». Cette attaque souligne ainsi que toute organisation se doit d'être aussi exigeante avec ses fournisseurs, et à plus forte raison lorsqu'il s'agit de fournisseurs de logiciels, de sécurité ou de *cloud*, qu'avec elle-même. Il peut s'agir par exemple d'exiger de ses sous-traitants des garanties d'applications des bonnes pratiques, de rendre obligatoires des audits externes, voire des audits personnalisés, permettant de combler les éventuelles lacunes des processus de certification, d'imposer une mise en conformité complémentaire à celle déjà requise par son secteur d'activité...

Autre enseignement de cette attaque : l'utilisation des outils et des systèmes d'administration s'impose comme un moyen extrêmement efficace et discret de compromettre une victime. La protection de ces produits et de ces systèmes doit donc faire l'objet d'une attention particulière de la part de toute organisation.

Enfin, le piratage de SolarWinds rappelle surtout à quel point les organisations sont aujourd'hui interconnectées. Une attaque sur une seule d'entre elles peut avoir des effets dévastateurs au niveau mondial. L'accélération de la transformation numérique des organisations et leur dépendance grandissante aux fournisseurs de produits ou de services numériques ne fait qu'accentuer cette tendance. Pour les organisations directement ciblées par l'attaque Solarwinds, comme pour celles qui en dépendent, ainsi qu'en règle générale pour toutes les organisations dont les activités reposent en partie sur des fournisseurs et des sous-traitants, l'année 2021 posera de nouveau défis. Au regard du succès de cette attaque et des dégâts qu'elle a déjà pu causer, au regard de l'appétence de certains acteurs étatiques pour les opérations de déstabilisation, il est à craindre que le scénario SolarWinds ne se répète.

[8] Cf. SolarLeaks.net

[9] Cf. tweet de Jake Williams, *Twitter* [En ligne], 12 janvier 2021.

[10] « SolarWinds : l'étau se resserre sur l'éditeur piraté », *Silicon* [En ligne], 6 janvier 2021.

ANALYSES (2/2)

FAKE NEWS ET MANIPULATIONS DE L'INFORMATION : LA TECHNOLOGIE, ALLIÉE OU ENNEMIE ?

Le présent article est une synthèse du deuxième module de l'événement ***Cyberdéfense et Stratégie****, organisé entre les 15 et 17 décembre 2020 par CEIS au profit du Commandement de la cyberdéfense du ministère des Armées. Dédié au sujet* ***« Fake news et manipulations de l'information : la technologie, alliée ou ennemie ? »****, cet atelier eu lieu dans un format webinar dont le replay est disponible ici.*

Les échanges ont réuni les intervenants suivants (par ordre alphabétique) : ***Lukas Andriukaitis*** *(Directeur associé au Digital Forensic Research Lab de l'Atlantic Council),* ***Vincent Claveau*** *(Informaticien à l'Institut de recherche en informatique et systèmes aléatoires/IRISA),* ***Paula Gori*** *(Secrétaire générale de l'European Digital Media Observatory),* ***Chine Labbé*** *(Rédactrice en chef Europe à NewsGuard), le* ***colonel Philippe de Montenon*** *(adjoint au Commandant de la cyberdéfense) et* ***Benoît Raphaël*** *(Chief Robot Officer à Flint).*

Les manipulations de l'information permettent d'obtenir des effets auxquels aucun conflit armé ou outil militaire ne peut prétendre, tels que renverser une opinion publique ou délégitimer un responsable politique. Bien orchestrée et « crédibilisée », une fausse information peut même semer la panique au sein d'une population sélectionnée, ce qui pourrait alors l'apparenter à une forme d'attaque armée, en raison des potentielles violences, morts ou destructions d'outils de production associées.

La France ne peut ainsi se désintéresser des manipulations de l'information. Et ce, d'autant plus que le droit international n'offre aucune garantie contre leur utilisation à des fins de politique étrangère. Elle doit ainsi se tenir prête, dans le respect du droit international, à pouvoir les détecter, les caractériser et y répondre selon des cas à préciser. Le champ informationnel étant désormais investi par les moyens numériques[11], la France doit demeurer à l'état de l'art des technologies associées à la création et la détection des *fake news* afin de mieux les combattre. En effet, l'adaptation aux évolutions technologiques est cruciale pour appréhender les attaques informationnelles, dont celles situées au-dessus du seuil estimé de riposte ou d'attaque armée.

[11] Ministère des Armées, *Actualisation stratégique*, Janvier 2021, p. 19.

Accédez à l'allocution d'ouverture du colonel Philippe de Montenon ici :

Qu'il s'agisse de propagande ou de désinformation, les manipulations de l'information exploitent un large éventail de méthodes et de techniques de conception, de falsification et d'amplification de contenus. Parmi les technologies qui en sont à l'origine, celles de l'intelligence artificielle (IA) sont sans doute aujourd'hui les plus efficaces.

Par leur nombre élevé d'utilisateurs, les réseaux sociaux servent à l'ère numérique de caisse de résonance aux fausses informations. En 2018, le ministère de la Culture a indiqué dans une étude qu'ils s'imposaient désormais comme le premier moyen pour suivre l'actualité au quotidien de 71% des jeunes (15-34 ans)[12], au détriment des médias classiques. La crise Covid-19 et les confinements successifs ont renforcé cette tendance. Les réseaux sociaux constituent ainsi un terrain privilégié des manipulations de l'information, pratiques facilitées par la multiplication sur les plateformes des contenus en formats courts (texte, clip vidéo, image), de plus en plus populaires auprès des utilisateurs. Les possibilités de falsification de ce type de contenus, comme de façon générale des contenus vidéo diffusés sur les réseaux sociaux, ont été démultipliées par les progrès de l'IA.

Mais si la technologie nourrit la fabrique de la désinformation, elle est en parallèle de plus en plus efficace dans la lutte contre les manipulations de l'information. Par exemple, si l'IA permet de créer et de falsifier des contenus à des fins de désinformation (photomontage, *deepfake*, génération automatique de texte, etc.), elle offre également, avec d'autres technologies, des capacités utiles à leur détection (analyse des mouvements, reconnaissance faciale, détection d'altérations d'une image, etc.). Avec des méthodes de *machine learning* et de *deep learning*, un nombre croissant de solutions d'IA émerge et permet d'identifier des contenus manipulés, à partir entre autres d'algorithmes et d'important volumes de données.

La technologie permet donc à la fois de créer et de se prémunir des fausses informations. Dans quelle mesure et de quelle(s) façon(s) contribue-t-elle à la production et à la diffusion, mais aussi à la détection de contenus dans le cadre de manipulations de l'information ?

1. Le modeste rôle de la technologie dans les manipulations de l'information

1.1. Une fabrique de fausses informations essentiellement artisanale

À l'ère numérique, les manipulations de l'information sont essentiellement réalisées à partir de technologies considérées comme « artisanales » – c'est-à-dire peu sophistiquées. La réalisation d'un « faux » capable de tromper une audience ne nécessite pas forcément d'engager des moyens conséquents. Bien que l'émergence de technologies de pointe doive être surveillée dans les prochaines années, celles-ci ne constituent aujourd'hui pas le principal moteur de création et de falsification de contenus inauthentiques. La « désinformation

[12] Ministère de la Culture, *Les jeunes et l'information (synthèse),* Juillet 2018, p. 14.

artisanale » reste le risque le plus prégnant et ne disparaîtra pas de sitôt. Les exemples ci-dessous montrent que les principaux vecteurs des manipulations de l'information découlent de technologies simples mais efficaces.

- **Le photomontage : des logiciels de traitement d'image accessibles à tous**

En novembre 2020, un internaute a produit à l'aide d'un logiciel de montage probablement grand public (voire de traitement de texte) une fausse note d'information de la Direction générale de la sécurité civile du ministère de l'Intérieur. Dans le contexte de crise Covid-19, ce document visait à faire croire à la programmation quatre mois à l'avance d'un « *troisième confinement* » en mars 2021, suivi d'une « *campagne de vaccination massive* ». Son auteur l'a assortie d'un logo de la République française afin de lui donner l'authenticité d'une note officielle.

Le ministère de l'Intérieur a rapidement démenti l'information en indiquant qu'il s'agissait d'un « faux ». Il demeure que le montage a été publié puis partagé des millions de fois sur le média social Facebook. Son objectif visait à discréditer le gouvernement français et à alimenter les théories du complot. Très exposé, il a fait l'objet, de la part de plusieurs médias traditionnels, de travaux de *fact-checking* (vérification des faits) afin de réfuter la véracité de l'information véhiculée par le montage.

- **Les *deepfakes* : la démocratisation d'une arme de désinformation**

Un *deepfake* est une synthèse d'images reposant sur l'IA. Déjà utilisés à des fins artistiques ou pour la production audiovisuelle[13] (jeux vidéo, télévision, etc.), les *deepfakes* deviennent de plus en plus accessibles et sophistiqués à mesure que la recherche en IA progresse et que les technologies associées se démocratisent. Cette technique est toutefois régulièrement dévoyée à des fins malveillantes de désinformation. Sa principale utilisation consiste à cet égard à attribuer de faux propos ou comportements à des personnalités publiques.

S'ils suscitent des craintes depuis quelques années, les *deepfakes* – et plus largement les vidéos truquées – ont eu un impact limité dans le débat politique en 2020, puisque peu de campagnes de manipulations de l'information à des fins militantes y ont eu recours. Chine Labbé rappelle que l'une des vidéos les plus populaires sur les réseaux sociaux a été cette année-là un montage sur Nancy Pelosi, qui a simplement été ralentie et tronquée de sorte à faire apparaître la Présidente de la Chambre des représentants des États-Unis dans un état d'ébriété. Ce trucage montre qu'un modeste niveau technologique permet d'obtenir des effets importants.

- **Le traitement automatique des langues : l'invention et le recyclage de textes**

Domaine de l'IA, le traitement automatique des langues (NLP) se décline à des fins de manipulations de l'information dans, entre autres, la traduction automatique et la génération automatique de textes.

Selon Lukas Andriukaitis, la traduction automatique d'un contenu erroné, puis sa diffusion en ligne, demeure l'un des vecteurs les plus récurrents des manipulations de l'information. Cette méthode de « recyclage » de narratifs vers plusieurs audiences étrangères est privilégiée par les acteurs aux moyens techniques limités. Chine Labbé prend en exemple le courriel d'un prétendu lanceur d'alerte canadien en octobre 2020. Le mail, écrit dans un anglais approximatif, révélé sous la forme d'un article sur un blog, faisait état d'une feuille de route alarmiste du gouvernement canadien quant à l'arrivée d'un nouveau virus appelé « Covid-21 ». Cette information a circulé sur les réseaux sociaux dans plusieurs langues dont le français, l'espagnol, l'italien, le russe et le polonais. Sur Facebook, la version tagalog a suscité plus de trois millions de vues (Philippines).

Quant à la génération automatique de textes, l'exemple ci-dessous est le résultat d'un logiciel utilisé par l'Institut de recherche en informatique et systèmes aléatoires (IRISA). Celui-ci, à partir d'une « accroche » factuellement véridique (ici, l'événement annuel Cyberdéfense & Stratégie organisé par le COMCYBER),

[13] « Plus belle la vie : un épisode utilise l'effet spécial du "deepfake" », *franceinfo* [en ligne], 17 novembre 2020.

constitue, à partir d'informations authentiques recueillies sur Internet et portant sur des sujets similaires, un texte dont le contenu est totalement factice. Factice car ne correspond pas à la réalité mais n'en demeure pas moins cohérent, détaillé et bien construit. Un tel contenu peut facilement tromper le lecteur non ou peu averti.

Exemple de génération automatique de texte – Source : IRISA

1.2. Le rôle central des utilisateurs des réseaux sociaux dans la propagation des contenus

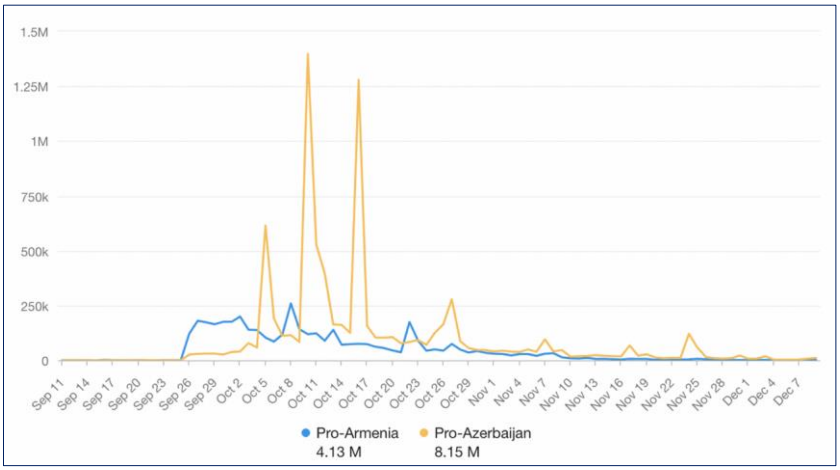
Plus que par leur sophistication technologique, les *fake news* gagnent en popularité grâce aux réseaux sociaux, qui font office de caisse de résonance. Automatisés voire semi-automatisés par des technologies de l'IA, des robots peuvent néanmoins être programmés pour promouvoir activement certaines informations, en partageant systématiquement les publications de comptes sélectionnés (*social bots*). Pour autant, la propagation des *fake news* s'effectue essentiellement de manière « naturelle » en passant de compte en compte, de page en page, de site de mésinformation en site de mésinformation. Les internautes conservent ainsi un rôle central dans la promotion de contenus sur les sites d'information et sur les réseaux sociaux.

Pour illustrer cette circulation naturelle des *fake news*, Chine Labbé prend l'exemple d'un article du site peu fréquenté GreatGameIndia, indiquant que le SARS-CoV-2 (Covid-19) a été créé dans un laboratoire canadien. La publication précise que, dans la perspective de développer une arme biologique, des espions chinois auraient amené ce virus à Wuhan (Chine) où il aurait fuité. Passant inaperçu, l'article a pourtant été dupliqué *in extenso* par le blog américain de mésinformation ZeroHedge, doté d'une plus grande popularité, qui l'a ensuite partagé sur ses réseaux sociaux. L'engagement suscité (nombre de partages, *likes* et commentaires) aurait alors été multiplié par quinze. La viralité de l'article a contribué à sa reprise par RedStateWatcher, dans le top 150 des sites les plus consultés aux États-Unis, qui lui a offert une exposition considérable.

Lukas Andriukaitis observe que des comptes authentiques peuvent activement contribuer à l'amplification de narratifs sur les réseaux sociaux. Pendant la seconde guerre du Haut Karabagh de 2020 par exemple, les deux camps belligérants – arméniens et azerbaïdjanais – se sont livrés à une bataille sur Twitter pour la conquête « des cœurs et des esprits » dans les sociétés occidentales[14]. Leur méthode consistait alors à mettre en avant des *tweets* favorables à son propre camp et hostiles à celui de l'adversaire. Des comptes venant de Turquie, alliée de l'Azerbaïdjan, ont participé à un « *hashtag activism*[15] » visant à nuire à l'Arménie, par des

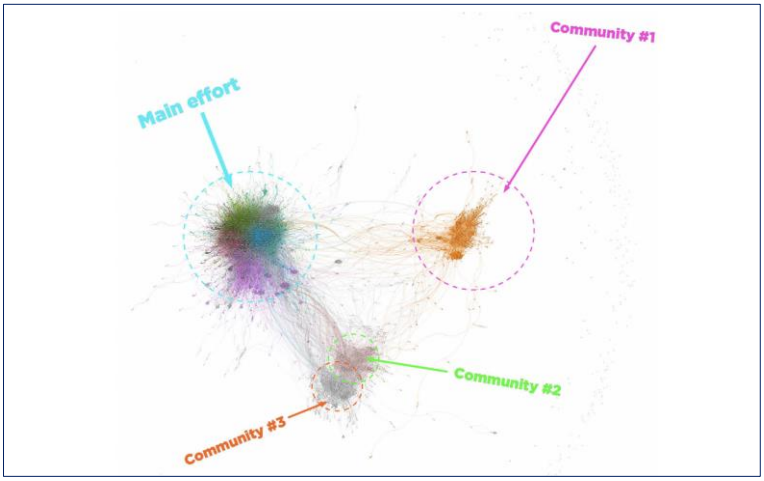
[14] DFRLab, « Turkish pop culture Twitter accounts mobilize to support Azerbaijan », *Medium* [en ligne], 15 décembre 2020.
[15] Voir « Fake news et manipulations de l'information, la prochaine épidémie ? », *OMC* [en ligne], Janvier 2021.

tweets utilisant une douzaine de *hashtags* hostiles dont #ArmenianTerrorism et #ArmenianAgression. Le graphique ci-dessous montre l'utilisation des *hashtags* « pro-Arménie » et « pro-Azerbaïdjan » lors du conflit :



Hashtags dans la seconde guerre du Haut Karabagh – Source : Lukas Andriukatis (DFRLab)

Plusieurs pics d'utilisation des *hashtags* « pro-Azerbaïdjan » peuvent être observés. L'un d'eux correspond au cessez-le-feu du 10 octobre et un autre à l'annonce d'une deuxième tentative de cessez-le-feu le 16 octobre. Le fait que ces pics s'étalent sur de courtes périodes similaires de cinq jours laisse supposer que les *hashtags* ont été promus de manière coordonnée. La modélisation des comptes ayant diffusé les #ArmenianTerrorism et #ArmenianAgression permet de dégager quatre réseaux distincts d'utilisateurs « pro-Azerbaïdjan » :



Visualisation des hashtags #ArmenianAggression et #ArmenianTerrorism – Source : Lukas Andriukatis (DFRLab)

Les comptes les plus actifs se regroupent au sein d'une communauté principale (*Main effort*). Leur date récente d'inscription sur Twitter suggère qu'ils ont spécifiquement été créés pour promouvoir une vision hostile à l'Arménie en utilisant les *hashtags* #ArmenianTerrorism et #ArmenianAgression. Pour autant, rien n'indique qu'il s'agit de faux-nez : leurs publications ne relèvent pas de *bots* car ils publient des *tweets* personnalisés et leurs activités ne se limitent pas au partage d'autre *tweets*. La recherche d'image inversée de leurs photos de profil montre que celles-ci ne sont ni fausses, ni volées. Les trois autres communautés dites de « soutien » (#1, #2 et #3) rassemblent des utilisateurs, majoritairement turcs, ayant contribué à la popularité des deux

hashtags en partageant de manière spontanée certains *tweets* de la communauté principale. Sur la base des mêmes critères que pour l'analyse de la communauté principale, il semblerait que ces comptes soient authentiques et qu'ils n'ont pas été créés dans le seul but d'amplifier des narratifs « pro-Azerbaïdjan ».

Accédez aux interventions de Lukas Andriukaitis en cliquant ici et ici :

Selon Chine Labbé, les utilisateurs des réseaux sociaux s'adaptent pour contourner le contrôle des plateformes visant à limiter la viralité d'un *hashtag*. Le *typosquatting* d'un hashtag, voire son remplacement par un autre, fait partie des moyens les plus employés. Ces techniques de contournement montrent les limites des plateformes : leur technologie ne peut ni détecter, ni contrer seule toutes les tentatives de manipulations de l'information.

2. La détection des *fake news* : la recherche d'un équilibre entre l'humain et la technologie

2.1. L'humain au cœur de la lutte contre les fausses informations

L'enjeu est de trouver le bon équilibre entre l'humain et la technologie face aux manipulations de l'information. Le colonel Philippe de Montenon rappelle que la lutte contre ces dernières s'appuie sur une diversité de métiers, à l'image du caractère protéiforme de ce phénomène. Une équipe équilibrée doit réunir des :

- techniciens pour développer, maîtriser et adapter des outils de détection de *fake news* à la situation ;
- linguistes pouvant comprendre et analyser les contenus en langue étrangère diffusés par les adversaires ;
- experts capables d'appréhender la culture et l'environnement des auteurs des manipulations de l'information, notamment en géopolitique si des forces armées en opération extérieure sont impliquées ou ciblées ;
- psychologues, en mesure d'analyser les enjeux des manipulations de l'information pour l'humain.

Ces compétences étant difficilement automatisables, l'humain doit conserver un rôle prépondérant. D'autant plus que l'IA identifie difficilement les contenus de propagande, de désinformation et de mésinformation, contrairement à ceux à caractère pornographique ou violent. La détection des manipulations de l'information requiert en effet, outre un esprit critique et de la culture générale, d'apprécier le contexte d'un contenu diffusé.

Aux États-Unis, lors de la campagne électorale de 2020, des comités d'action politique ont financé des sites de propagande qui, profitant du vide laissé par la fermeture de médias, se sont fait passer pour des médias locaux et ont diffusé des contenus orientés visant à renforcer les intérêts de leurs bailleurs. Le lecteur non averti peut être manipulé à son insu et la technologie est démunie face à ce type de situation. Seul un œil humain peut décrypter les effets recherchés par les auteurs de ces manipulations et ainsi mieux les appréhender.

NewsGuard : labéliser les sites d'information pour lutter contre les *fake news*

NewsGuard s'efforce de placer l'humain au cœur de ses activités. Elle évalue la fiabilité des sites d'information et alerte les internautes sur ceux qui ne le sont pas. Concrètement, elle met à disposition des internautes une extension de navigateur qui fait apparaître à côté de la source d'information un écusson de couleur rouge (source peu fiable) ou verte (source fiable).

Cette extension propose aussi une « étiquette nutritionnelle » du site d'information visité, qui présente une analyse détaillée des neuf critères journalistiques sur lesquels les équipes de NewsGuard fondent et mettent régulièrement à jour leurs évaluations de sites.

Les sites devant être notés sont identifiés par un algorithme qui analyse le nombre de contenus qu'ils publient et partagent en ligne. Il peut donc s'agir de médias reconnus mais aussi de blogs.

Retrouvez la présentation de NewsGuard et le Focus Entreprise dédié.

Accédez aux interventions de Chine Labbé en cliquant ici, ici et ici :

Face à la surabondance d'informations disponibles, Benoît Raphaël estime que les experts humains doivent s'attacher à développer une IA capable d'identifier et de collecter les « bonnes » informations, plutôt qu'une IA cherchant à contrer les (bien trop nombreuses) « fausses » informations. La technologie doit être considérée comme un outil qui doit permettre aux individus de prendre de la distance avec les *fake news* en sélectionnant pour eux des informations de qualité. C'est tout l'objectif du média Flint qu'il a créé.

Flint : l'IA contre les bulles d'information

Pour Flint, l'objectif n'est pas de lutter contre les *fake news* mais d'aider l'humain à mieux s'informer en le faisant sortir de la « bulle filtrante » dans laquelle les algorithmes des plateformes d'intermédiation l'enferment. La machine qui permettra de faire face aux réseaux sociaux doit mêler l'intelligence artificielle (IA) avec l'intelligence individuelle et collective des internautes.

Flint combine ainsi trois intelligences : artificielle, collective et individuelle. Pour fournir à chaque lecteur un contenu personnalisé et de qualité, son algorithme d'apprentissage automatique est entraîné par ses lecteurs et par ses quatre cents experts thématiques, qui sont aussi ses clients.

Cet apprentissage comporte deux volets : un « naturel », activé par la lecture des articles proposés par Flint, et un « actif » déclenché par les réactions « j'aime/j'aime pas » à ces mêmes articles.

À partir de ces deux volets, l'IA de Flint détermine les articles qui conviennent à chaque lecteur, tant en termes de thématiques que de critères qualitatifs, grâce à un algorithme combinant :

- Un modèle social, qui repose sur l'intelligence collective et le travail des milliers d'experts qui valident la pertinence des contenus ;
- Un modèle sémantique, qui s'attache à comprendre le sens et la signification des mots et des articles, et créer entre eux des connexions dans le but de développer une véritable culture générale au fil des apprentissages.

Pour plus d'informations sur Flint, accédez au Focus Entreprise.

Accédez aux interventions de Benoît Raphaël en cliquant ici et ici :

Plusieurs initiatives mobilisant des experts aux profils multiples dans la lutte contre la désinformation ont émergé ces dernières années, tant au niveau national qu'international. Parmi eux, l'European Digital Media Observatory (EDMO) vise à rassembler des spécialistes du *fact-checking*, de l'éducation aux médias, ainsi que des chercheurs universitaires, pour comprendre et analyser les fausses informations. L'EDMO a pour projet de se doter d'une plateforme collaborative et sécurisée de fact-checking d'envergure européenne.

Accédez à la présentation de l'EDMO en cliquant ici :

2.2. Les outils technologiques au service de l'expertise humain

Si l'humain doit rester au centre de la lutte contre les manipulations de l'information, il a à sa disposition un certain nombre d'outils et de technologies permettant la détection des manipulations de l'information :

- **Les outils forensiques : la détection des manipulations graphique d'une image**

La forensique recouvre l'ensemble des sciences informatiques relatives à l'analyse des traces numériques. Elle propose des outils capables d'analyser efficacement les manipulations locales d'une image, en repérant par exemple la présence de traces non cohérentes avec l'ensemble du contenu ou des traces laissées par

une compression. Néanmoins, la forensique perd en efficacité face aux images dégradées ou retouchées, ce qui est souvent le cas de celles en circulation sur les réseaux sociaux (petite dimension, compressées, etc.).

Le cas échéant, l'une des méthodes pour démontrer qu'une image a été manipulée consiste à retrouver l'image authentique – celle qui a servi de support au montage – puis de les comparer. Outre la recherche inversée, des technologies d'IA permettent de décrire des images sous forme de texte afin de faciliter les recherches.

- **Le *machine learning* contre les articles de désinformation**

Le *machine learning* propose des techniques connues et performantes pour identifier des articles véhiculant de fausses informations. Vincent Claveau indique que les « faux » articles se caractérisent notamment par leur stylistique, la ponctuation, l'usage récurrent des pronoms à la première personne, la longueur des textes et le vocabulaire. Un algorithme bien entraîné, alimenté par un grand volume d'articles issus à la fois de médias considérés comme fiables et de sites connus de désinformation, peut être interrogé sur ces attributs pour distinguer les articles susceptibles de véhiculer de « faux » contenus. En revanche, la génération automatique de texte ne cesse de se perfectionner et risque d'être de moins en moins discernable à l'avenir.

- **Le défi technologique de la décontextualisation**

La décontextualisation consiste à renforcer l'impact d'un texte en l'associant à une image qui n'illustre pas le même fait – ou inversement. Il est en l'état difficile de détecter le « décalage » entre les deux supports utilisés car ils font rarement l'objet de manipulations de part et d'autre. L'un des principaux enjeux aujourd'hui est de pouvoir décrire le texte et l'image dans un même espace de représentation.

Accédez à l'intervention de Vincent Claveau en cliquant ici :

Conclusion

Le rôle de la technologie ne doit pas être surestimé au détriment de celui de l'humain dans les manipulations de l'information. Les contenus utilisés sont souvent produits à partir de moyens peu sophistiqués mais qui demeurent efficaces pour tromper autrui (photomontage, *deepfakes*, etc.). De plus, la propagation de ces contenus relève moins de *bots* que de l'action d'internautes, qui contribuent activement à leur circulation sur les réseaux sociaux et sur les sites d'information. La viralité d'un contenu tient ainsi davantage à son auteur et aux individus qui la relaient qu'à la technologie employée pour sa création et sa diffusion.

Face à la complexité des manipulations de l'information, l'humain ne peut se remettre entièrement à une machine pour décider si un contenu est faux ou non. La technologie – dont celles de l'IA – doit se cantonner à lui faciliter la tâche. Cela par le biais de solutions capables, outre de repérer les traces de falsification d'un contenu, de « scanner » la toile et son immense volume de contenus afin d'identifier ceux potentiellement faux.

Le développement d'outils et de solutions permettant de lutter contre la menace grandissante des manipulations de l'information relève d'une responsabilité collective. L'État ne peut en effet détenir et contrôler seul leur utilisation. La société civile, la recherche et l'industrie doivent contribuer à l'émergence non pas d'une mais d'une myriade de solutions à la mesure des différents enjeux. Outre le soutien à cette démarche, l'État doit de son côté exiger davantage de transparence de la part des plateformes d'intermédiation privées, notamment pour le référencement et le contrôle de leurs contenus.

FOCUS INNOVATION

Sahar : rendre la *data* pleinement accessible et actionnable

Entretien avec Antoine Franz et Gauthier Schweitzer (co-fondateurs).

Présentation

Antoine Franz et Gauthier Schweitzer ont fondé Sahar en 2019. *Data scientists*, les deux associés s'efforcent de mettre leur appétence technique pour la donnée au service de l'intérêt général. Leur objectif est de rendre « actionnables » les données de l'Internet ouvert, en plus de sensibiliser les citoyens, les métiers et les décideurs publics sur le potentiel de celles-ci. Le Grand débat national[16] a constitué à cet égard un élément déclencheur, en donnant l'opportunité de mettre en cohérence leur expertise dans la recherche textuelle, ainsi que dans l'analyse et la visualisation de données massives, à travers le projet democratie.app (cf. *infra*). Depuis, la startup développe ses propres produits et accompagne ses clients dans leur stratégie *data*.

Solution

Sahar est spécialisée dans l'analyse des données textuelles en sources ouvertes. Composé essentiellement de *data scientists* et de développeurs web, son personnel traduit deux ambitions claires quant à la donnée : demeurer à l'état de l'art des algorithmes de traitement et être en mesure de la restituer en temps réel.

Au service de l'utilisateur, la technologie de la startup permet l'acquisition, le traitement et la visualisation des données relatives aux publications écrites sur les sites d'information et les réseaux sociaux. Elle s'appuie sur des techniques d'intelligence artificielle dont le traitement automatique des langues, le *machine learning* et surtout le *deep learning*. Inspirée de l'Elastic Stack, elle s'appuie sur des algorithmes spécifiquement dédiés aux cas d'usage, qui sont développés en surcouches (analyse de sentiments, viralité, etc.).

[16] Débat public lancé par le Président de la République française pour recueillir les attentes des Français quant aux politiques publiques dans le contexte du mouvement des Gilets jaunes.

Applications

Sahar est à l'origine de nombreux projets plaçant tous l'intérêt collectif au cœur de leur développement dont :

- democratie.app, un moteur de recherche citoyen et apolitique dans le cadre de l'élaboration des politiques publiques liées au Grand débat national de 2019. La plateforme fait apparaître la tendance et les avis des Français quant à un sujet relatif à quatre thèmes (transition écologique, fiscalité, démocratie et citoyenneté, organisation de l'État et des services publics) et par zone géographique. Elle a notamment permis aux administrations locales de recueillir les perceptions relatives à leur action ou à leur communication ;
- anteviral (site désormais inactif), qui accompagnait ses utilisateurs dans la détection et l'analyse des *fake news* relatives à la Covid-19. Le projet cherchait à limiter la propagation des fausses informations avant qu'elles ne sapent le débat public et la confiance envers les institutions. La plateforme suivait, mesurait et classifiait les informations textuelles en ligne afin de faire remonter celles virales et jugées suspectes. Elle transmettait ensuite ces dernières à des journalistes partenaires dans le cadre d'initiatives de *fact-checking* ;
- des exercices de crise de « communication » à l'École nationale d'administration (ENA). Par la génération automatique de texte en langue française sur les réseaux sociaux, impliquant une forte composante technique, ces séminaires simulent une situation de crise politique à laquelle les élèves doivent réagir.

Actualité

Sahar a été lauréate du hackathon organisé par l'Assemblée nationale en 2019 pour son projet democratie.app. La startup est passée depuis mars 2020 de deux à treize employés, confirmant ainsi sa montée en puissance. En octobre 2020, Sahar a rejoint la Cyberdéfense Factory du ministère des Armées où elle développe un outil d'analyse de l'influence.

CALENDRIER

10/02/2021 : PETIT-DÉJEUNER THÉMATIQUE « ATTAQUES PAR REBONDS : LA SUPPLY CHAIN EST-ELLE LE MAILLON FAIBLE ? »

Organisé par **CEIS (Avisa Partners)** au profit du **Commandement de la cyberdéfense**, un petit-déjeuner-débat en visioconférence sur le sujet « **Attaques par rebonds : la *supply chain* est-elle le maillon faible ?** » aura lieu le **mercredi 10 février de 8h30 à 10h00**.

Si l'année écoulée a été marquée par les campagnes de *ransomware* et de *phishing* qui ont accompagné la généralisation du télétravail et l'accélération de la transformation numérique, l'attaque "Solarwinds" de décembre 2020 a rappelé brutalement les risques et les conséquences des attaques indirectes.

Ces menaces sérieuses aux conséquences en série et potentiellement dévastatrices ne sont pas une nouveauté, mais cristallisent les pires situations. On se souviendra par exemple de NotPetya en 2017.

L'accroissement du nombre et de la gravité des cyberattaques dites "*supply chain*", consistant à atteindre une cible en infiltrant les systèmes d'information (SI) de ses partenaires, prestataires numériques ou encore éditeurs de solutions logicielles, rappelle ainsi que l'ensemble de l'écosystème d'une organisation constitue aussi une porte d'entrée majeure dans son SI.

Depuis la découverte de l'attaque "Solarwinds" en décembre 2020, chaque jour apporte son lot de révélations sur les modes d'action des attaquants, les outils utilisés, les victimes touchées et les dommages causés, qui dépeignent une attaque d'une ampleur sans précédent. L'impact et les ramifications de cette affaire soulèvent aussi des doutes quant à l'implication d'acteurs étatiques ou sponsorisés par des États.

Quels enseignements tirer de l'épisode "Solarwinds" ? Faut-il craindre les conséquences de cette attaque d'ampleur ? Comment se préparer pour faire face à ce type d'action sur la *supply chain* ? Faut-il ré-internaliser certains développements, est-ce d'ailleurs possible ? Comment se protéger sur l'ensemble de la chaîne, jusqu'aux fournisseurs, et notamment, pour les Armées, comment appréhender la protection de la BITD et des fournisseurs de services numériques ?

La liste des intervenants sera communiquée prochainement.

Le lien de la visioconférence vous sera transmis dans les jours précédant l'évènement. Il est également possible d'assister physiquement à son enregistrement (nombre de places limité). En cas d'intérêt, merci de contacter Julien Tran Van Nhieu (julien.tran@avisa-partners.com*).*

ACTUALITÉ

LE MINISTÈRE DES ARMÉES CRÉE L'AGENCE DU NUMÉRIQUE DE DÉFENSE (AND)

Le ministère des Armées va se doter en ce début d'année 2021 d'une Agence du numérique de défense (AND). Cette dernière, rattachée au Délégué général pour l'armement (DGA), sera chargée de conduire les projets numériques complexes ou à forts enjeux de l'ensemble du ministère. Sa création s'effectue dans le cadre des grands travaux de transformation numérique initiés, depuis quelques années, par la Direction générale du numérique et des systèmes d'information et de communication (DGNum).

L'AND a vocation à fédérer et à mutualiser les capacités existantes pour devenir la référence interne en matière de gestion de projets numériques. Outre la diffusion des meilleures pratiques dans le domaine, elle veillera à « *la cohérence technique et à l'optimisation des moyens des projets qui lui sont confiés* », en apportant notamment « *un cadre de méthode et de soutien* ». Elle assurera ainsi la maîtrise d'ouvrage de bout-en-bout des projets numériques sur l'ensemble du cycle de vie (conception, réalisation, déploiement et retrait).

Cette nouvelle organisation sera dirigée par un conseil d'orientation et de gestion associant l'ensemble des acteurs internes : les états-majors, les directions et les services du ministère des Armées.

Retrouvez plus d'informations sur le site du ministère des Armées en cliquant ici.

La **Direction générale des relations internationales et de la stratégie (DGRIS)** propose des analyses politiques et stratégiques contribuant à renforcer l'appréciation des situations et l'anticipation. Elle soutient la réflexion stratégique indépendante, en particulier celle menée par les instituts de recherche et les organismes académiques français et étrangers. Elle contribue au maintien d'une expertise extérieure de qualité sur les questions internationales et de défense.

À ce titre, la **DGRIS** a confié à **CEIS** la réalisation de cet **Observatoire du Monde Cybernétique (OMC)**, qui s'inscrit dans le contrat-cadre n°2018-02. Les opinions développées dans cette étude n'engagent que leur auteur et ne reflètent pas nécessairement la position du Ministère de la Défense.

……

Ministère des Armées
Direction générale des relations internationales et de la stratégie
60 boulevard du général Martial Valin | 75015 Paris

……

CEIS
Tour Montparnasse | 33 avenue du Maine | 75015 Paris
E-mail : omc@ceis.eu