



HAL
open science

Éthique de la lutte contre la désinformation

Serena Villata, Célia Zolynski, Emmanuel Didier

► **To cite this version:**

Serena Villata, Célia Zolynski, Emmanuel Didier. Éthique de la lutte contre la désinformation. Eric Germain; Claude Kirchner; Catherine Tessier. Pour une éthique du numérique, PUF, 2022, 978-2-13-083348-2. hal-03918915

HAL Id: hal-03918915

<https://cnrs.hal.science/hal-03918915>

Submitted on 2 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Éthique de la lutte contre la désinformation

Serena Villata, Université Côte d'Azur, CNRS, Inria, I3S

Célia Zolynski, Université Paris 1 Panthéon-Sorbonne, IRJS DReDIS

Emmanuel Didier, Centre Maurice Halbwachs, CNRS, ENS et EHESS

Produit et véhiculé sur Internet *via* les réseaux sociaux, les sites web, les forums ou les messageries instantanées, le phénomène des « fausses nouvelles » a pris une ampleur inédite depuis la crise sanitaire liée à la COVID-19. Les Nations-Unies et plusieurs de ses agences (OMS, Unicef) ont évoqué une véritable « infodémie », entretenue par les effets des confinements, de l'isolement, de la gravité de la situation ou encore la multiplicité des facteurs d'incertitude tenant aux situations individuelles et collectives. Il a en effet été constaté la diffusion d'une masse inégalée d'informations relatives notamment à l'origine et la prévention du virus de la COVID-19, la recherche de traitements, les conséquences de l'épidémie, les politiques de confinement et de déconfinement, l'éventuel traçage des chaînes de contamination, la discrimination de certaines populations, l'annonce de pénuries, les stratégies vaccinales, ou encore de publicités mensongères ou malveillantes.

Différencier la désinformation et mésinformation ou « Complexité du phénomène et enjeux éthiques »

S'il ressemble à celui bien connu de la rumeur – ou diffusion au sein du public d'informations à l'origine incertaine et à la véracité douteuse¹ –, ce phénomène a acquis des caractéristiques spécifiques dans le monde numérique. Il désigne alors la propagation potentiellement massive, souvent délibérée ou automatisée, d'informations de tous types qui peuvent avoir des conséquences graves. Les intentions de leurs auteurs ou de leurs propagateurs peuvent être diverses. Certaines informations sont délibérément créées et relayées pour tromper subtilement, jeter le trouble, induire en erreur des personnes, des organisations ou l'opinion publique ou encore pour favoriser certains intérêts ; on parlera alors de « désinformation ». D'autres informations peuvent s'avérer incertaines, incomplètes ou erronées alors qu'elles sont présentées comme sûres et diffusées de bonne foi par des propagateurs humains ; cela inclut tout un pan de contenus scientifiques, d'informations mal comprises ou mal reformulées, d'inquiétudes non fondées ou insuffisamment fondées scientifiquement, massivement diffusées par les plateformes numériques², qui relèvent alors de la « mésinformation »³.

¹ Voir notamment E. Morin, *La rumeur d'Orléans*. Nouv. [3.] éd. complétée avec *La Rumeur d'Amiens. L'Histoire immédiate*. Paris : Seuil, 1977.

² Les plateformes numériques sont par exemple les réseaux sociaux, les moteurs de recherche, et les systèmes de partage de vidéos.

Dès le début de la crise sanitaire, les plateformes ont très vite accepté le principe qu'il leur fallait modérer un certain nombre de contenus, mais ont été confrontées à deux difficultés liées qui étaient déjà connues. Tout d'abord, ce travail de modération est complexe car toute information, selon le cadre dans lequel elle est présentée, la manière dont elle est formulée ou le point de vue de son destinataire, est susceptible de relever finalement de la désinformation ou de la désinformation. Ensuite, le fait de sélectionner, de promouvoir ou de réduire la visibilité de certaines informations échangées sur les plateformes numériques entre en tension avec le respect des libertés d'information et d'expression qu'elles promeuvent par ailleurs. Différentes questions peuvent dès lors émerger, en particulier : que traduisent ces actions ou inactions dans le contexte de la COVID-19 ? Constate-t-on simplement un changement de volume ou, plus profondément, un changement de nature des solutions numériques conçues pour lutter contre la désinformation et de la désinformation ? Plus généralement, comment appréhender la complexité d'un tel phénomène dès lors que celui-ci appelle des analyses qui paraissent dépasser l'éthique, voire qui interroge la notion même d'éthique ? Ainsi, la distinction entre désinformation et désinformation engendre une tension dans la nature des positions éthiques que l'on est amenées à défendre. Les questions éthiques ne se posent en effet pas de la même manière selon que l'on traite d'acteurs qui agissent en toute conscience pour tromper leur cible ou que l'on s'adresse à ceux qui, pris dans les flux d'informations en format numérique, participent à la viralité de ces informations sans en être nécessairement conscients. Dans le premier cas, la réflexion éthique interroge la responsabilité ; dans le second, elle invite surtout à une prise de conscience. En toutes hypothèses, elle nécessite d'identifier, tout particulièrement dans le cadre numérique, les dimensions économiques, juridiques, sociales, politiques ou philosophiques des mécanismes de désinformation ou de désinformation.

Le Comité national pilote d'éthique du numérique (CNPEN) a publié, en juillet 2020, un bulletin de veille afin de dresser un état des lieux des tensions et enjeux éthiques résultant de la lutte contre la désinformation en ligne⁴. Il nous est apparu nécessaire de prendre en compte toute la complexité d'un tel phénomène qui a des implications transversales dans différents domaines tels que l'informatique, le droit, la sociologie.

Les difficultés posées par la modération des fausses nouvelles peuvent être regroupées en trois ensembles : celles qui sont associées aux algorithmes ; celles qui relèvent du phénomène de la viralité ; enfin, celles qui sont posées par l'identification et les relations avec des autorités légitimes.

Comprendre les limites des solutions algorithmiques

Les mécanismes de lutte contre la désinformation et la désinformation développés par les plateformes reposent en partie sur des outils automatisés, compte tenu du volume considérable d'informations à analyser. Ils sont néanmoins supervisés par des modérateurs humains. Chaque crise interroge le degré de cette supervision. Durant le confinement, celle-ci a été largement réduite car les conditions de télétravail, souvent non anticipées, pouvaient amener à utiliser des

³ Voir sur cette distinction la communication de la Commission européenne *Lutter contre la désinformation concernant la COVID-19 – Démêler le vrai du faux*, 10 juin 2020, JOIN(2020) 8 final, p. 4&s.

⁴ CNPEN, « Enjeux d'éthique dans la lutte contre la désinformation et la désinformation », *Bulletin de veille* n°2, 21 juil. 2020.

réseaux non sécurisés pour transférer de tels contenus, potentiellement délictueux, ou à devoir les modérer dans un contexte privé difficilement maîtrisable. Or les risques d'atteintes disproportionnées à la liberté d'expression se sont avérés plus importants en l'absence de médiation et de validation humaines, seules à même d'identifier voire de corriger les erreurs de classification ou les biais algorithmiques. L'emploi massif d'outils automatiques indépendamment de tout contrôle humain exercé *a posteriori* a en outre interrogé la possibilité de recours offerte à l'auteur d'un contenu ayant été retiré par la plateforme. Ces difficultés montrent clairement l'importance pour la société civile que les plateformes soient plus transparentes sur les critères algorithmiques de classification de la désinformation ainsi que sur les critères qu'elles retiennent pour définir leur politique de modération, qu'ils soient d'ordre économique ou relèvent d'obligations légales. Ces politiques de modération doivent être mieux explicitées et factuellement renseignées dans les rapports d'activités périodiques qu'elles sont tenues de publier depuis la loi contre la manipulation de l'information, dite « loi Infox » promulguée le 22 décembre 2018⁵. Plus généralement, il apparaît qu'une réflexion sur la constitution de bases de données communes pour améliorer les outils numériques de lutte contre la désinformation et la mésinformation devrait être menée pour aboutir à un partage des métadonnées associées aux données qu'elles collectent à cette.

Maîtriser la viralité

L'ampleur prise à ce jour par les phénomènes de désinformation et de mésinformation tient à l'accroissement de mécanismes de viralité qui se déploient à partir des outils offerts par les plateformes. La viralité est d'abord l'effet du modèle économique de nombre de ces opérateurs, qui sont rémunérés par les publicitaires en fonction des interactions avec les utilisateurs qu'ils obtiennent et ont donc intérêt à générer des clics ou toute autre réaction aux contenus. Elle relève ensuite du rôle joué par leurs utilisateurs eux-mêmes dans la propagation virale de la désinformation et de la mésinformation, que ces derniers y contribuent délibérément ou par simple négligence ou ignorance. La lutte contre la désinformation doit donc nécessairement être l'affaire de tous les utilisateurs, à qui il appartient de se montrer plus scrupuleux avant de décider de partager des informations et ainsi de contribuer à leur propagation virale. Cette remarque va d'ailleurs dans le même sens que le programme *#MarquonsUnePause*⁶ désormais promu par l'ONU. Mais ceci n'est possible que si les plateformes mettent à disposition de leurs utilisateurs un certain nombre d'informations et d'outils afin de les mettre en mesure de prendre conscience, de réfléchir et de maîtriser le rôle qu'ils jouent dans la chaîne de viralité de l'information⁷. Sous la pression des gouvernements, et particulièrement depuis le début de la crise sanitaire, les

⁵ Voir le bilan d'activité pour 2019 publié par le CSA le 30 juillet 2020 et sa recommandation du 15 mai 2019, ou encore les propositions formulées dans le cadre du Digital Services Act publié par la Commission européenne le 15 décembre 2020 (Proposition de Règlement relatif à un marché intérieur des services numériques et modifiant la directive 2000/31/CE, COM(2020) 825 final)

⁶ Le programme *#MarquonsUnePause* est une campagne lancée par l'ONU pour encourager les citoyens à réfléchir avant de partager une information sur les réseaux sociaux, un appel à changer de comportement, mais aussi à diffuser des informations fiables. Elle s'inscrit dans le cadre de l'initiative *Vérifié* qui a pour objectif de lutter contre la désinformation liée à l'épidémie de COVID-19, afin que chacun ait accès à des informations précises et fiables.
<https://unric.org/fr/covid-19-marquonsunepause-avant-de-partager-des-informations/>

⁷ *Ibid.* voir sur ce point les recommandations du CSA.

plateformes ont mis en œuvre diverses pratiques visant à indiquer explicitement qu'une information reçue a été massivement partagée et invitent leurs utilisateurs à être vigilants avant de repartager des contenus ayant fait l'objet de signalements. Il serait toutefois possible d'aller plus loin mais cela semble difficile sans un changement drastique du modèle économique fondé sur la viralité. Plus fondamentalement, il est important de renforcer l'esprit critique des utilisateurs, ce qui suppose tout particulièrement que ceux-ci puissent être sensibilisés aux sciences et technologies du numérique afin de mieux maîtriser le fonctionnement de ces plateformes et les effets induits par ces mécanismes de viralité.

Promouvoir une gouvernance plurielle de la modération

Il faut relever que, si la modération des contenus et le contrôle de la viralité jouent un rôle majeur dans le contrôle pragmatique de la désinformation et de la mésinformation, ces opérations ne peuvent *in fine* être accomplies sans référence à des autorités indépendantes établissant, ne serait-ce que temporairement, la validité des arguments échangés dans l'espace public. Les plateformes n'ont en effet aucune qualité ni compétence pour déterminer, par exemple, l'efficacité d'un vaccin ou le bienfondé d'une mesure de santé publique. Afin de contrôler et vérifier les informations circulant sur Internet, en particulier en période de crise sanitaire, ces opérateurs doivent pouvoir les comparer à des informations émanant de sources considérées comme étant sûres ou légitimes à l'image des services de « vérification des faits » privés ou liés à des organes gouvernementaux, ou encore du Service statistique public qui fournit l'immense majorité des chiffres concernant la pandémie. En France, depuis le début de la crise sanitaire, de nombreuses plateformes se sont ainsi rapprochées de services gouvernementaux, en particulier du secrétariat d'État chargé du numérique ou du Service d'information du Gouvernement. Dans le même temps, ces mêmes plateformes se sont éloignées d'autres gouvernements, comme en atteste leur modération des contenus publiés par Jamir Bolsonaro ou Donald Trump. En l'occurrence, on peut comprendre ces choix. Ils n'en restent pas moins arbitraires dans la mesure où ils ne reposent pas sur des principes explicites régulant les relations entre les plateformes et les gouvernements. À cet égard, une réflexion d'ensemble sur la responsabilité des plateformes ainsi que sur le contrôle à exercer s'agissant de leur politique de modération de contenus paraît devoir être menée. À notre sens, ce contrôle ne peut être dévolu à l'État seul et devrait relever d'une autorité indépendante, incluant les représentants de diverses associations, scientifiques et acteurs de la société civile dans l'établissement des procédures de sélection d'informations à promouvoir, tout particulièrement en période de crise sanitaire.