COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

COMITÉ NATIONAL PILOTE D'ÉTHIQUE DU NUMÉRIQUE

Réflexions et points d'alerte sur les enjeux d'éthique du numérique en situation de crise sanitaire aiguë

Bulletin de veille n°2

Enjeux d'éthique dans la lutte contre la désinformation et la mésinformation

Mardi 21 juillet 2020

Les phénomènes de désinformation et de mésinformation ont été exacerbés à l'occasion de la crise engendrée par l'épidémie de SARS-CoV-2. Cela a conduit les plateformes numériques telles que les réseaux sociaux, moteurs de recherche, ou systèmes de partage de vidéos à développer des pratiques et des outils numériques pour contribuer à lutter contre leurs effets délétères tant sur le plan individuel que collectif.

Ce bulletin vise à identifier les tensions et enjeux éthiques résultant de ces diverses actions, ce qui nécessite de prendre en compte toute la complexité d'un tel phénomène aux implications transversales. Différentes questions peuvent alors émerger, par exemple : 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 mésinformation ? 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 interrogent la notion même d'éthique ? Ainsi, la distinction entre désinformation et mésinformation engendre une tension dans la nature des prises des positions éthiques qu'on est amenés à défendre. En effet, lorsqu'il s'agit d'acteurs qui agissent en toute conscience pour tromper leur cible, la réflexion éthique tend à interroger la responsabilité de chacun ; lorsqu'elle s'adresse plutôt à ceux qui, pris dans les flux d'informations, participent à la viralité de ces informations sans en être nécessairement conscients, elle appelle surtout à une meilleure maîtrise de leur rôle dans les mécanismes de viralité de l'information numérique. 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 mésinformation.

66, rue de Bellechasse 75007 Paris – 01 42 75 66 42
www.ccne.fr – contact@comite-ethique.fr

Ce bulletin s'inscrit dans la lignée d'un travail de veille engagé par le CNPEN depuis le début de la crise sanitaire[1]. Il entend contribuer à cette réflexion d'ensemble sous l'angle éthique en dressant un constat des actions et inactions mises en œuvre par les plateformes à l'occasion de la crise COVID-19. À l'aune de ce contexte spécifique, il formule des recommandations et identifie plusieurs points d'attention pour nourrir une réflexion qu'il conviendra de poursuivre sur ces phénomènes de désinformation et de mésinformation à l'ère numérique.

Emmanuel Didier, Serena Villata, Célia Zolynski
Rapporteurs du groupe de travail
Claude Kirchner
Directeur du comité national pilote d'éthique du numérique

1 https://www.ccne-ethique.fr/fr/actualites/comite-pilote-dethique-du-numerique-bulletin-de-veille-ndeg1

TABLE DES MATIÈRES

INTRODUCTION

S'il a toujours existé, le phénomène de la rumeur - c'est-à-dire la diffusion au sein du public d'informations à l'origine incertaine et à la véracité douteuse - se traduit dans le monde numérique par la propagation potentiellement massive, souvent délibérée ou automatisée, d'informations de tous types. Les intentions de leurs auteurs ou de leurs propagateurs peuvent être diverses. Certaines informations sont délibérément créées pour tromper subtilement, jeter le trouble, induire en erreur des personnes, des organisations ou l'opinion publique ou encore pour favoriser certains intérêts ; le fait de diffuser ces informations avec l'intention délibérée d'induire en erreur, de causer un préjudice public, ou encore de réaliser un gain économique peut être qualifié 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, constitués de rumeurs, d'informations mal comprises ou mal reformulées, d'inquiétudes non fondées ou insuffisamment fondées scientifiquement, massivement diffusées par les plateformes.[7] Leurs propagateurs humains n'ont généralement pas conscience des effets de cette transmission d'information, ni du fait qu'ils contribuent ainsi au modèle économique des plateformes. Cela relève alors de la « mésinformation »[2].

Produit et véhiculé sur Internet *via* les réseaux sociaux, les sites web, les forums ou les messageries instantanées, ce phénomène a pris une ampleur inédite depuis 2016 notamment avec la campagne de l'élection présidentielle américaine, la campagne du Brexit, et en 2017 avec la campagne présidentielle française. La crise sanitaire liée à la COVID-19 a exacerbé ce phénomène au point que les Nations Unies et plusieurs de ses agences (OMS, Unicef) évoquent désormais une véritable « infodémie »[3]. Confinement, isolement, anxiété, gravité de la situation ou encore multiplicité des facteurs d'incertitude constituent le terreau fertile de l'amplification de la désinformation et de la mésinformation, qui se joue tant à l'échelle individuelle que planétaire. La désinformation et la mésinformation ont ainsi pu concerner notamment l'origine et la prévention du virus SARS-CoV-2, 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 qui désorganisent sans fondement le fonctionnement de la société ou encore les publicités mensongères ou malveillantes.

2 Sur cette distinction, v. 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. CELEX 52020JC0008 FR TXT-1.pdf

3 « Les infodémies constituent une surabondance d'informations sur un problème donné, qui rend la définition d'une solution difficile. Lors d'une crise sanitaire, elles peuvent être sources de mésinformation, de désinformation et de rumeurs. Les infodémies peuvent faire obstacle à une réaction efficiente en termes de santé publique et susciter confusion et méfiance au sein de la population. » https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4

Cette nouvelle échelle de la désinformation est intimement associée à l'apparition des réseaux sociaux[4], moteurs de recherche[5], et systèmes de partage de vidéos[6], que nous désignons ici par le terme plateformes[7] numériques. Ces dernières augmentent significativement la capacité de leurs utilisateurs à jouir de la liberté d'expression propre à chacun, en contribuant à la diffusion, à la circulation et à l'échange d'informations. Chacun peut y exprimer son opinion librement, principe qui est revendiqué par les démocraties occidentales, ou telle est au moins la perception qu'en ont les utilisateurs. La défense de ce principe de liberté individuelle a permis aux plateformes de se présenter comme de simples diffuseurs d'informations sans responsabilité éditoriale. Or il est apparu, au moins depuis le milieu des années 2010, que cette liberté devait nécessairement être encadrée. Les scènes de violence - en particulier lorsqu'il s'agit d'actes terroristes - ou de pornographie diffusées sur les plateformes ont démontré que certains types de contenus pouvaient avoir des conséquences dangereuses pour certains groupes d'utilisateurs ou des populations entières. Des particularités culturelles font, par exemple, que la nudité fait partie de cette liste des contenus prohibés aux États-Unis d'Amérique. Dans le contexte actuel, certaines informations concernant l'épidémie, comme les publicités pour de faux remèdes, peuvent avoir des conséquences graves sur la santé. Elles peuvent aussi accentuer la défiance de la population à l'égard des autorités publiques et rendre plus difficile la gestion de la crise sanitaire. Ces effets réels et rapidement constatés ont poussé certaines plateformes à modérer davantage les contenus, voire à supprimer ou à promouvoir certaines informations.

Ce travail de modération est extrêmement complexe : toute information, quelle que soit son origine ou sa valeur de vérité, peut potentiellement devenir mésinformation ou désinformation selon le cadre dans lequel elle est présentée, la manière dont elle formulée ou le point de vue de son destinataire. En effet, l'information n'est pas seulement vraie ou fausse au sens d'une valeur de vérité ; elle est aussi inscrite dans un cours d'actions ou un contexte, dans une pragmatique, c'est-à-dire évaluée en fonction de ses sources et de ses effets avérés ou supposés. Cette évaluation comporte donc toujours une part d'incertitude et un aspect politique. Dans son actualité, l'information nécessite une mise en perspective : elle est nécessairement reçue et interprétée en fonction d'un ensemble de présupposés et d'effets sociaux et politiques propres à chaque destinataire. La valeur de vérité de l'information importe alors moins que la pragmatique de sa propagation, c'est-à-dire l'ensemble de ses conséquences et effets empiriques (à qui elle sert, comment elle permet de faire des alliances, quels types d'actions elle suscite, etc.)[8].

4 Facebook, Tiktok, LinkedIn, Twitter, WhatsApp, Mastodon, etc.

5 Google Chrome, Qwant, DuckDuckGo, Ecosia, etc.

6 YouTube, DailyMotion, Snapchat, etc.

7 Sur cette terminologie, v. la Communication de la Commission européenne *Lutter contre la désinformation concernant la COVID-19 – Démêler le vrai du faux,* préc. ou encore le Rapport de la mission « Régulation des réseaux sociaux – Expérimentation Facebook », https://www.vie-publique.fr/sites/default/files/rapport/pdf/194000427.pdf

8 V. également l'avis n° 2018-37 du COMETS (le Comité d'éthique du CNRS) - *Quelles nouvelles responsabilités pour les chercheurs à l'heure des débats sur la post-vérité ?* 12/04/2018 : https://comite-ethique.cnrs.fr/wp-content/uploads/2019/10/AVIS-2018-37.pdf

Le problème repose alors sur l'intrication entre, d'une part, l'évaluation de la valeur de vérité des informations diffusées et, d'autre part, le degré de liberté d'expression laissé aux propagateurs relativement aux conséquences de leurs actes de parole.[9] Cette complexité de l'évaluation éthique est en outre aggravée par le fait que, sur les réseaux sociaux, tout individu ou tout groupe constitué rapidement ou spontanément (souvent uniquement en ligne) peut diffuser ses opinions à une échelle globale. Cette absence de sélection et la remise à plat des hiérarchies sociales est un facteur de premier plan dans l'analyse éthique. La situation créée par la crise Covid-19 a davantage participé à cette relativisation informationnelle : tout d'abord, le confinement a produit l'isolement des individus et les a rendus plus dépendants des réseaux numériques ; ensuite, la science, dont la temporalité est toujours plus lente que celle de l'actualité, a dû intégrer des facteurs d'incertitude radicale dans une communication sur des enjeux vitaux pour l'ensemble de la population.

Sans critère universel permettant de qualifier une information comme *fausse* puisque sa légitimité dépend de la perspective d'où elle est saisie et de la société dans laquelle elle est émise, les plateformes ont cherché à établir des cadres permettant de déterminer ce qu'il est possible ou non de diffuser. Pour cela, elles s'appuient parfois sur des associations de « *fact-checkers* », souvent organisées par les grands organes de presse, ou encore sur des autorités sanitaires ou gouvernementales nationales et internationales. Elles établissent aussi leurs propres critères permettant de discriminer des contenus jugés illicites ou dangereux des autres informations - cette détection s'appuyant souvent sur un recours massif à des outils numériques automatisés. La crise sanitaire de la Covid-19 a montré que ces ressources ne suffisaient pas : le caractère médico-scientifique des informations diffusées et des controverses dont elles pouvaient faire l'objet ont accentué la difficulté qu'il peut y avoir à identifier les autorités légitimes en matière d'information. Elle a aussi mis en évidence le fait que ces plateformes ne pouvaient définir seules de telles procédures de tri et de sélection des informations.

L'ensemble de ces phénomènes et actions, parfois amplifiés par la crise sanitaire, suscite différents questionnements de nature éthique. Il est tout d'abord nécessaire de s'interroger sur les risques potentiels d'atteintes disproportionnées aux libertés d'expression et d'information qui peuvent en résulter. Il est en effet essentiel, même en période de crise, de garantir les principes fondamentaux de nos démocraties[10] tels l'accès à l'information, la liberté d'expression, l'indépendance des médias et la délibération ouverte. Il est en outre possible d'interroger la légitimité et les effets du pouvoir, numérique et politique, qui semble ainsi acquis par les plateformes sous couvert de l'objectif de lutte contre la désinformation ou la mésinformation. L'articulation de ce pouvoir avec celui des autorités préexistantes (État, juridictions…), devrait être pensée. Plus généralement, il paraît essentiel de s'interroger sur la responsabilité éthique qui devrait incomber aux

9 J.L. Austin, *Quand dire, c'est faire*, Paris, le Seuil, 1991.

10 Sur ce point, v. la déclaration sur la liberté d'expression et d'information en temps de crise par le Comité d'experts du Conseil de l'Europe sur l'environnement des médias et la réforme (MSI-REF), 21 mars 2020 – également *Respecter la démocratie, l'état de droit et les droits de l'homme dans le cadre de la crise sanitaire du COVID-19. Une boîte à outils pour les États membres*, 7 avril 2020, SG/Inf(2020)11.

différents acteurs contribuant à la diffusion de ces contenus au moyen d'outils numériques.

De telles interventions peuvent en effet donner lieu à des positions divergentes. On pourrait considérer que toute suppression est une atteinte à la liberté, ou bien que le dommage engendré par la suppression de certaines formes d'art ou d'humour propres aux réseaux sociaux est moindre que le dommage suscité par la diffusion de fausses informations. On pourrait également craindre que de telles pratiques induisent, par effet mimétique, le même type de censure ou d'autocensure hors des réseaux sociaux, conduisant à terme à l'appauvrissement de la vie sociale en général. Ou encore considérer que ce changement adviendra vraisemblablement, mais qu'il ne nous appartient pas d'en juger car ce sera aux générations futures de dire s'il est bon ou mauvais de leur point de vue. Il est possible de résumer la situation comme résultant des tensions entre trois éléments : premièrement, le respect de la liberté d'expression ; deuxièmement, l'identification d'autorités, nouvelles ou anciennes, ayant la légitimité de déterminer les contours de cette liberté, ainsi que les limites nécessaires au pouvoir de celles-ci ; troisièmement, les procédures concrètes de modération des échanges entre utilisateurs et plateformes traduisant en acte les décisions de ces autorités. De ce triangle de tensions émergent de nombreuses questions éthiques.

L'objectif de ce bulletin n'est pas d'évaluer la valeur de vérité de certaines informations ni les conséquences immédiates de leur diffusion ; cette tâche est celle que les acteurs du web, plateformes et autorités compétentes. Il vise à expliciter et analyser les enjeux éthiques soulevés par les choix institutionnels que les plateformes ont mis en œuvre pour lutter contre les phénomènes de désinformation et de mésinformation. Autrement dit, il ne s'agira pas d'analyser les modes de production de désinformation mais d'envisager les choix faits ou délibérément non réalisés par ces différents acteurs pour réagir au phénomène nouveau d'« infodémie » afin d'identifier les tensions éthiques que leurs actions ou inactions peuvent soulever.

Ces tensions concernent en premier lieu la mise œuvre des outils de modération et les mécanismes de lutte contre la viralité auxquels les plateformes ont recours (I) ; elles interrogent par ailleurs les rapports que ces opérateurs entretiennent avec différentes autorités étatiques, judiciaires, scientifiques ainsi qu'avec la presse (II).

I. OUTILS DE MODÉRATION ET MÉCANISMES DE VIRALITÉ

Pour répondre au risque de désinformation, particulièrement accru dans le contexte de la crise COVID-19, les différentes plateformes proposent une variété de moyens pour agir sur les contenus diffusés, que ce soit pour les supprimer, réduire leur visibilité ou les promouvoir.

Suppression de contenus : La plupart des plateformes suppriment les contenus pouvant causer un danger imminent ou être préjudiciables à la santé publique (contestant, par exemple, une décision ou une recommandation d'un organisme de santé publique ou un fait scientifique) ou, plus généralement, étant susceptibles de porter atteinte à l'intégrité d'autrui ou à l'ordre public. Elles peuvent aussi refuser de diffuser les publicités identifiées comme trompeuses, mensongères ou jouant sur la panique. En temps de crise COVID-19, certaines suspendent par exemple les faux comptes d'utilisateurs se faisant passer pour des organismes de santé, ou ceux identifiés comme diffusant des informations erronées et potentiellement dangereuses pour la santé.

Réduction de la visibilité de contenus : Plusieurs plateformes réduisent la diffusion de certains contenus en les rétrogradant dans leur ordre d'apparition. Elles peuvent aussi signaler à l'utilisateur les informations douteuses et le rediriger vers des articles ou des pages de vérification des faits sur le sujet. D'autres limitent le nombre possible de transferts de contenus ou bloquent les comptes à partir desquels sont effectués des transferts en masse.

Promotion de contenus : Différentes plateformes promeuvent des informations sous la forme de bandeaux, de contenus éditorialisés ou de fils d'actualité provenant de sources qu'elles estiment de confiance tels que des organismes de santé publique, des ministères, ou des sites de vérification des faits. Elles peuvent également promouvoir ces informations en les mettant en avant dans leur référencement ou en amplifiant leur visibilité par des annonces publicitaires gratuites.

Si ces réponses, qui s'appuient sur des moyens techniques et humains, peuvent paraître *a priori* adaptées pour lutter contre la désinformation, elles posent dans le même temps différents problèmes éthiques relatifs, d'une part, à l'emploi d'outils automatiques pour détecter ce type d'informations (1.1) et, d'autre part, aux mécanismes de viralité, c'est-à-dire la diffusion rapide et imprévisible de ces contenus, qui alimentent leur propagation (1.2).

A. Les outils automatiques

Le recours à différents outils automatiques s'explique par la nécessité d'un passage à l'échelle dans la détection de contenus relevant de la désinformation ou de la mésinformation. En effet, compte tenu du nombre considérable d'informations qui circulent *via* Internet, et notamment à travers les plateformes, le recours aux outils automatiques paraît seul permettre de rendre la détection de la désinformation ou de la mésinformation plus efficace par rapport à ce que peuvent réaliser les vérificateurs

humains (*fact-checkers*) en qualifiant la plupart de ces informations en très peu de temps. L'automatisation de ces procédures pose toutefois de nombreuses questions éthiques qui tiennent tant à la fiabilité de ces outils automatiques qu'à leurs effets s'agissant du respect de la liberté d'expression.

Tout d'abord, lorsqu'ils sont totalement automatisés, ces outils peuvent comporter des biais algorithmiques ou des erreurs de classification. Quant à leur fonctionnement, ces outils utilisent divers algorithmes de détection automatique de la désinformation ou de la mésinformation adaptés à différents supports (texte, vidéos, images). Ces algorithmes se fondent notamment sur la reconnaissance de mots clés dans les textes, la détection de la réutilisation d'images anciennes à travers l'analyse de la chronologie ou encore la détection de l'angle du visage, de son teint et de son expression, l'éclairage et d'autres informations importantes pour vérifier l'authenticité de vidéos de personnes.

Pour aller plus loin

Évaluer la véracité d'une information est une tâche complexe et lourde, même pour des experts qualifiés tels que les fact-checkers *humains[1]. Par exemple, une première étape pour identifier les contenus relevant de la désinformation ou de la mésinformation consiste à analyser ce que les autres sources d'information disent sur le sujet. Cette tâche automatique est appelée "*stance detection*" et consiste à estimer la position relative de deux morceaux de texte par rapport à un sujet. Cette approche permet d'établir la cohérence de ces contenus (*consistency*).*

Il existe différentes stratégies d'étiquetage ou de classement pour la détection des contenus relevant de la désinformation ou de la mésinformation. Dans la plupart des études, la détection de ces informations est formulée comme un problème de classification ou de régression. L'approche la plus courante consiste à formuler la tâche de détection de ces informations comme un problème de classification binaire (désinformation ou non). Cependant, classer tous ces contenus en deux classes est difficile parce qu'il y a des cas où les contenus ne relèvent que partiellement de la désinformation (seulement une portion du contenu relève de la désinformation ou de la mésinformation). La détection de ces informations peut également être formulée comme une tâche de régression, où le résultat est un score numérique de véracité.

Ces approches algorithmiques de détection automatique de la désinformation ou de la mésinformation sont confrontés à plusieurs défis. Un premier défi important est lié à la disponibilité et la qualité des données : pour que les classificateurs atteignent de bonnes performances, ils doivent disposer de suffisamment de données étiquetées. Or l'étiquetage fiable d'un large volume de données implique un travail long et complexe de la part d'experts qualifiés. La détection du contexte représente un autre défi important. Elle suppose de mettre au point des algorithmes en mesure d'analyser efficacement des informations à long terme et de transition de contenu en utilisant des connaissances de base. Enfin, un troisième défi consiste dans le croisement des données multimodales. En effet, pour être détectés efficacement, certains contenus supposent de croiser différents types d'informations tels que du texte, des images et les métadonnées associées à ces contenus

.

Compte tenu de ces difficultés, l'automatisation de la détection de contenus relevant de la désinformation ou de la mésinformation peut conduire à un certain nombre d'erreurs de classification, appelées faux positifs, c'est-à-dire d'informations classifiées à tort comme relevant de la désinformation ou de la mésinformation. Cela concerne notamment les contenus humoristiques, ironiques ou caricaturaux, ou encore les informations nécessitant la prise en compte de diverses connaissances antérieures pour permettre une classification appropriée. Cette détection automatique peut aussi être source de faux négatifs : des informations relevant bien de la désinformation ou de la mésinformation peuvent ne pas être pas détectées par l'algorithme et continuer à être diffusées sur les plateformes. De plus, d'éventuels biais algorithmiques de fonctionnement peuvent influencer la détection de la désinformation ou de la mésinformation. Ces biais sont dus à des choix conscients ou inconscients des développeurs ou à des biais provenant des données et peuvent entraîner une discrimination indirecte d'une partie des utilisateurs.

Bien qu'elle soit rendue nécessaire par le volume considérable d'informations à analyser, cette vérification algorithmique peut donc induire des risques de censure et d'atteintes disproportionnées à la liberté d'expression. Ces risques de biais algorithmiques et d'erreur de classification sont d'autant plus importants en l'absence de mécanismes de médiation et de validation finale par un être humain. Or, à l'occasion de la crise sanitaire, il est apparu que les plateformes n'ont pas été en mesure de laisser leurs équipes de modérateurs, en situation de télétravail, accéder à toutes les informations nécessaires du fait de leur contenu potentiellement intrusif ou dérangeant (contenus violents, propos haineux, …). En effet, les conditions de télétravail, souvent non anticipées, pouvaient amener à utiliser des réseaux non sécurisés pour transférer de tels contenus (potentiellement délictueux) ou à devoir les modérer dans un contexte privé difficilement maîtrisable. Cela a donc engendré un moindre contrôle humain de ces processus de suppression, réduction ou promotion de contenus, alors que la désinformation et la mésinformation progressaient fortement. Par ailleurs, l'emploi massif d'outils automatiques indépendamment de tout contrôle humain exercé *a posteriori*, qui peut induire un risque de censure automatique, interroge la possibilité de recours offert à l'auteur d'un contenu ayant été retiré par la plateforme.

Ensuite, le recours massif aux outils automatiques interroge la transparence et l'explicabilité des algorithmes mis en place pour détecter la désinformation ou la mésinformation. Cette problématique recouvre deux aspects distincts. D'une part, elle est relative à l'explication du résultat produit par l'algorithme et les principaux éléments pris en compte pour y parvenir (ex. les caractéristiques utilisées par le système de classification supervisé pour discriminer les contenus relevant de la désinformation, le degré de fiabilité estimé par l'algorithme du résultat obtenu, les données d'apprentissage les plus influentes pour la tâche de classification, les caractéristiques d'entrée les plus marquantes ou les caractéristiques significatives à l'intérieur des couches d'un réseau de neurones). D'autre part, le manque de transparence concerne aussi les critères retenus par les plateformes pour définir leur politique de modération (qu'ils soient d'ordre économique ou qu'ils relèvent d'obligations légales, etc.). Les solutions algorithmiques mises en place dans ces

outils peuvent ainsi induire des biais “de décision” (ou volontaires) qui influencent la modération de contenus. La question se pose alors de savoir si la plateforme doit être transparente sur ces différents critères à l’égard de ses utilisateurs et des régulateurs, dans le prolongement des obligations qui lui sont imposées par la loi du 22 décembre 2018 s’agissant de la lutte contre la mésinformation en période électorale[11].

Recommandations :

1.1 Garantir, y compris après la crise, l’existence d’une modération par l’humain permettant de vérifier les résultats produits automatiquement par les algorithmes d’analyse de contenus.

1.2 Maintenir, y compris au cours de la crise, les instruments de recours ouverts à l’auteur d’un contenu s’agissant des décisions de suppression ou promotion de contenus prises par la plateforme sur la base d’un traitement algorithmique.

1.3 Promouvoir la transparence et l’explicabilité, et donc l’« auditabilité » des algorithmes de détection de désinformation ou mésinformation et de recommandation de contenus utilisés par les plateformes, plus encore en période de crise. Exposer aux utilisateurs les critères qui fondent la décision algorithmique afin de protéger la liberté d’expression face aux trois réponses proposées par les plateformes : suppression de contenus, réduction de la visibilité de contenus, promotion de contenus.

Un autre problème tient encore à l’inégalité de performance dans la tâche de détection automatique des contenus relevant de la désinformation ou de la mésinformation entre la recherche publique et les plateformes en raison d’un accès très limité aux données de ces dernières. Ces données sont d’une grande utilité, tant pour les vérificateurs humains que pour l’amélioration des algorithmes automatiques. Les annotations visant à identifier le type de désinformation ou de la mésinformation (par exemple une citation tronquée), les méta-informations telles que la source, la date et l’heure de publication ou encore les partages de ce contenu sur Internet peuvent en effet contribuer à améliorer ces algorithmes de détection. Cela interroge alors la gouvernance de l’ensemble des données liées à la désinformation ou à la mésinformation identifiées et collectées par les plateformes et sur l’utilité qu’il y aurait à favoriser le partage de ces données entre

11 Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, art. 11. V. également la recommandation du CSA n° 2019-03 du 15 mai 2019 : “le Conseil encourage les plateformes à assurer à chaque utilisateur :
- la traçabilité de ses données exploitées à des fins de recommandation et de hiérarchisation des contenus, qu’elles soient fournies sciemment ou collectées par l’opérateur de la plateforme en ligne ;
- une information claire, suffisamment précise et facilement accessible sur les critères ayant conduit à l’ordonnancement du contenu qui lui est proposé et le classement de ces critères selon leur poids dans l’algorithme ;
- une information claire et précise sur sa faculté, si elle existe, de procéder à des réglages lui permettant de personnaliser le référencement et la recommandation des contenus ;
- une information claire et suffisamment précise sur les principaux changements opérés dans les algorithmes de référencement et de recommandation, ainsi que sur leurs effets ;
- un outil de communication accessible permettant l’interaction en temps réel entre lui et l’opérateur, et offrant à l’utilisateur la possibilité d’obtenir des informations personnalisées et précises sur le fonctionnement des algorithmes”. V. également, s’agissant des relations entre plateformes et utilisateurs consommateurs, les obligations imposées aux opérateurs au titre des articles L. 11-7 &s. du Code de la consommation.

différents acteurs[12]. La question se pose tout particulièrement lorsque l'État, comme en période de crise COVID-19, entend associer les plateformes à une politique publique de lutte contre la désinformation. Il conviendrait par exemple de mettre au point des mécanismes de partage mieux adaptés et synchronisés à l'échelle continentale ou internationale, tout en offrant aux scientifiques, aux citoyens, et à la société civile la possibilité de contribuer à leur développement et mise à jour.

<u>Point d'attention :</u>

1.a Comme le préconise le rapport de la mission « Régulation des réseaux sociaux – Expérimentation Facebook »[13] et la Commission Européenne[14], il convient d'encourager à ce que soit menée une réflexion d'ampleur sur la constitution de bases de données communes pour améliorer les outils numériques de lutte contre la désinformation et la mésinformation et d'inciter les plateformes à partager les métadonnées associées aux données qu'elles collectent à cette fin (ex. source, sujet, citations, partages sur les plateformes, contre-arguments publiés en tant que commentaire de ces contenus). De telles bases de données permettraient en outre de faciliter la recherche scientifique dans ce domaine.

Au-delà, se pose également la question de la mise en œuvre effective de ces moyens techniques et humains par les plateformes pour lutter contre la désinformation et la mésinformation. Il a pu être notamment constaté une ambiguïté dans l'attitude de certaines plateformes qui, en dépit des annonces faites sur les actions menées pour lutter contre la désinformation sur les enjeux de la crise COVID-19, laissaient leurs outils publicitaires à la disposition de certains sites qui en étaient à la source[15]. S'agissant de la lutte contre la haine en ligne, la loi visant à lutter contre les contenus haineux sur internet entendait y remédier en imposant aux plateformes de rendre compte au CSA des moyens matériels et humains qu'elles mettent en œuvre pour lutter contre la diffusion de contenus visés par la liste des infractions définies par le texte[16]. La loi du 22 décembre 2018, qui impose aux plateformes de prendre des mesures en vue de lutter contre la diffusion de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un

12 V. l'article 7-III de la loi visant à lutter contre les contenus haineux sur internet (version votée par Parlement mais jugée contraire à la Constitution par la décision du Conseil constitutionnel n°2020-801 DC du 18 juin 2020) encourageant les plateformes, sous l'égide du CSA, à mettre en œuvre des outils de coopération et de partage d'informations, dans un format ouvert entre ces opérateurs, dans la lutte contre les infractions visées par le texte (contenus haineux).

13 Rapport de la mission « Régulation des réseaux sociaux – Expérimentation Facebook » – https://www.vie-publique.fr/sites/default/files/rapport/pdf/194000427.pdf, p. 18.

14 A ce titre, v. la Communication "Lutter contre la désinformation concernant la COVID-19 – Démêler le vrai du faux" préc., section 5.2.

15 Sur ce point, v. notamment le rapport de l'ONG Tech Transparency Project : https://www.techtransparencyproject.org/articles/google-profiting-coronavirus-conspiracy-sites

16 Loi visant à lutter contre les contenus haineux sur internet, article 5 jugé contraire à la Constitution par la décision du Conseil constitutionnel préc. S'agissant des critiques adressées face à l'inaction de certaines plateformes, v. par ex. l'action intentée par quatre associations (Union des étudiants juifs de France (UEJF), J'accuse, SOS-Racisme et SOS-Homophobie) contre Twitter : https://www.lemonde.fr/pixels/article/2020/05/12/twitter-assigne-en-justice-pour-son-inaction-massive-face-aux-messages-haineux_6039412_4408996.html

des scrutins, précise que ces mesures et leurs modalités de mise en œuvre doivent être rendues publiques[17]. Par ailleurs, la Commission européenne a annoncé en juin 2020 qu'il serait demandé aux plateformes de remettre un rapport mensuel sur leurs politiques et actions visant à lutter contre la désinformation liée à la COVID-19 précisant notamment les données sur les flux de publicité liés à la désinformation. Les autorités de contrôle et leurs utilisateurs pourront ainsi en évaluer l'efficacité ainsi que la réalité des promesses faites par les opérateurs[18].

Recommandation :

1.4 Mettre en place les mécanismes permettant de s'assurer que les plateformes diffusent un rapport d'activité périodique, accessible aux autorités de contrôle ainsi qu'à leurs utilisateurs, exposant de façon claire, loyale, précise et transparente leur politique de gestion de lutte contre la désinformation et la mésinformation, les moyens matériels et humains mis en œuvre à cette fin et les données relatives aux flux de publicité liés à la désinformation.

Point d'attention :

1.b Une réflexion d'ampleur devrait être menée à l'avenir sur les ressources à allouer à la modération humaine et la répartition des coûts en résultant, la qualification des modérateurs humains, le traitement des biais culturels éventuels voire les critères sur lesquels ils fondent leur décision et leur éventuel contrôle par une autorité indépendante, notamment le juge comme garant des libertés fondamentales.

B. Les mécanismes de viralité

L'ampleur prise à ce jour par les phénomènes de désinformation et de mésinformation tient à l'accroissement de la diffusion de ces contenus par les mécanismes de viralité qui se déploient à partir des outils offerts par les plateformes et les moteurs de recherche. Leur modèle économique peut amplifier un tel phénomène dès lors qu'il repose sur l'économie de l'attention qui encourage la viralité, source de bénéfices (1.2.1). Les utilisateurs jouent également un rôle dans ce phénomène puisque les micro-actions (réexpédier, partager, etc.) en sont le déclencheur initial (1.2.2).

a) Modèle économique des plateformes favorisant la viralité

Le modèle économique de certaines plateformes est fondé sur la rémunération au nombre de clics - donc sur la promotion des *Clickbaits* (« pièges à clics ») - et repose sur la captation et la valorisation de l'attention de leurs utilisateurs auprès des annonceurs. Ce modèle s'appuie sur un système algorithmique de priorisation des contenus qui met en avant les

17 Loi n°2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, article 11.

18 En ce sens, v. les propositions faites par la Commission européenne dans sa Communication du 10 juin 2020 préc., p. 10-11.

contenus suscitant le plus de réactions et de conversations. Ainsi, il participe grandement aux mécanismes de viralité. Un tel fonctionnement peut produire des effets délétères : il conduit à faire primer des contenus concentrant davantage l'attention par des effets de coordination virale (souvent des contenus choquants, haineux ou des fausses nouvelles)[19], par rapport aux contenus publiés par la presse. Ces informations sont alors souvent placées au second rang[20], afin d'offrir plus de visibilité aux contenus viraux qui génèrent davantage de revenus publicitaires. Cet « engagement métrique », qui a pour objectif principal la captation, la rétention et la monétisation de l'attention des utilisateurs des plateformes, est formalisé dans le code sous une forme algorithmique, ce qui accentue le risque de promotion de fausses nouvelles si elles sont attrayantes. Ce modèle paraît en outre conduire à une forme d'hyper personnalisation des contenus qui enferme les internautes dans des bulles filtrantes et amplifie leurs biais sociologiques et cognitifs, en particulier l'homophilie - tendance à former des liens avec des individus qui nous ressemblent - et le biais de confirmation - tendance à privilégier les informations confirmant nos hypothèses[21]. Les actions mises en œuvre par les plateformes pour accompagner la lutte contre la crise Covid-19 ne doivent pas occulter les limites de ces modèles économiques.

Point d'attention :

1.c Il sera souhaitable d'approfondir l'analyse des mécanismes sur lesquels reposent les marchés de publicité en ligne dont les ressorts et les critères de fixation des prix sont pour l'heure opaques et soulèvent un certain nombre d'enjeux éthiques.

b) Viralité et rôle des utilisateurs

Si les plateformes jouent bien un rôle d'amplificateur dans la diffusion des contenus, il convient d'analyser les différentes causes des mécanismes de viralité et, en particulier, le rôle d'utilisateurs individuels ou de groupes d'utilisateurs dans ce mécanisme[22].

Le rôle des utilisateurs se joue en effet à deux niveaux : d'une part, en leur qualité de destinataires de l'information, d'autre part, en leur qualité d'agents de la viralité dès lors qu'ils participent à la propagation des informations en publiant, republiant et commentant différents contenus textuels ou visuels (mème, GIF, etc.). A ce titre, il faut distinguer deux types d'utilisateurs – sachant qu'un individu ou un groupe peut passer d'un type à l'autre en fonction des circonstances. D'une part, on trouve ceux qui participent intentionnellement à cette propagation, la plupart du temps pour des motifs idéologiques ou cupides. Ce type de comportements pose la question de la responsabilité et des

19 D. Cardon, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data,* Seuil, 2015, p. 91.

20 B. Patino, *La civilisation du poisson rouge. Petit traité sur le marché de l'attention*, Grasset, 2019, p. 141.

21 Ce dernier phénomène est toutefois discuté : v. not. F. Tarrisan, *Au coeur des réseaux*, Le Pommier, 2019, p. 115 &s.

22 En ce sens, v. le rapport visant à renforcer la lutte contre le racisme et l'antisémitisme sur Internet, remis au Premier ministre en septembre 2018 par L. Avia, K. Amellal et G. Taieb, https://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2018/09/rapport_visant_a_renforcer_la_lutte_contre_le_racisme_et_lantisemitisme_sur_internet_-_20.09.18.pdf, pt. 6.1.

sanctions applicables à de tels comportements sur le plan juridique. D'autre part, on trouve les acteurs qui peuvent participer à la viralité par simple négligence ou ignorance des effets néfastes que celle-ci peut engendrer.

Il paraît nécessaire d'inciter ces derniers à être scrupuleux avant de décider de partager des informations et ainsi de contribuer à leur propagation virale. Si ce partage relève, sur le plan éthique, de leur responsabilité en tant qu'acteurs de la diffusion d'une information douteuse ou erronée, la promotion d'une conduite plus responsable suppose que les plateformes mettent ces utilisateurs en mesure de prendre conscience, voire de maîtriser, le rôle qu'ils jouent dans la chaîne de viralité de l'information. A ce titre, la loi du 22 décembre 2018 sur la lutte contre la mésinformation impose aux plateformes de mettre en place des dispositifs appropriés permettant aux utilisateurs d'être informés sur la nature, l'origine et les modalités de diffusion des contenus. Le CSA[23] recommande ainsi aux opérateurs de plateforme en ligne de veiller à :

- distinguer clairement les contenus sponsorisés des autres contenus et encourager le développement d'outils permettant à l'utilisateur d'identifier les critères qui ont conduit la plateforme à lui proposer de tels contenus ;
- appeler les utilisateurs à faire preuve de vigilance concernant les contenus qui ont fait l'objet de signalements[24] ;
- identifier de façon claire l'origine des contenus diffusés et l'afficher de manière visible ;
- préciser les modalités de diffusion des contenus en indiquant dans la mesure du possible les conditions de leur publication telles que l'existence de contreparties financières, l'ampleur de la diffusion (nombre de vues, type de population ciblée, etc.), et s'ils ont été générés de manière automatisée ou non.

Afin de permettre à l'utilisateur de mesurer et maîtriser son rôle dans la chaîne de viralité, il s'agirait par ailleurs de demander aux plateformes d'offrir à leurs utilisateurs la possibilité de :

- mesurer, lorsqu'ils utilisent un réseau social, la tribune qu'ils offrent aux contenus qu'ils diffusent directement (par publication sur son propre profil ou par "*retweet*" ou "*share*") ou indirectement (par "*like*") ;
- réaliser que les informations qu'ils diffusent ou relaient sont autant d'indications permettant notamment aux réseaux sociaux de les profiler plus précisément.

23 Recommandation du CSA n° 2019-03 du 15 mai 2019 aux opérateurs de plateforme en ligne dans le cadre du devoir de coopération en matière de lutte contre la diffusion de fausses informations, n°5.

24 Par exemple, le « flaggage » de fausses nouvelles par les plateformes.

Recommandations aux plateformes :

1.5 Mettre en œuvre les recommandations formulées par le CSA concernant la mise en place de dispositifs appropriés permettant aux utilisateurs d'être informés sur la nature, l'origine et les modalités de diffusion des contenus, et tout particulièrement demander aux plateformes :

- d'indiquer explicitement à leurs utilisateurs qu'une information reçue a été massivement partagée ;
- d'être vigilant avant de repartager des contenus ayant fait l'objet de signalement

1.6 Développer et mettre à disposition les outils permettant à leurs utilisateurs de prendre conscience de la tribune qu'ils offrent aux contenus qu'ils diffusent, directement ou indirectement, par l'intermédiaire de la plateforme.

Points d'attention :

1.d S'il est important, comme le souligne le CSA, de promouvoir les outils permettant à l'utilisateur d'identifier les critères qui ont conduit la plateforme à lui proposer de tels contenus (comprendre ce qu'il voit), un autre enjeu éthique conduira à développer les outils permettant d'assurer aux utilisateurs la capacité de préciser les critères de mise en visibilité de certains contenus à leur égard, pour tenir compte de leurs propres intérêts.

1.e Par ailleurs, et compte tenu du volume et de la vitesse de propagation de l'information circulant sur les plateformes, en particulier sur les réseaux sociaux, une réflexion pourrait être menée sur l'utilité de penser le ralentissement d'une telle circulation par le recours à des moyens numériques. Il s'agirait ainsi de rendre l'utilisateur plus scrupuleux et de l'inciter à analyser d'avantage son contenu lorsqu'il entend partager une information de façon spontanée.

Plus généralement, il conviendrait de favoriser le développement de l'esprit critique des utilisateurs afin qu'ils soient en mesure de partager un contenu en connaissance de cause. Les plateformes pourraient ainsi rappeler aux utilisateurs l'intérêt d'évaluer la pertinence et la fiabilité de leurs sources, par exemple en comparant une information aux autres informations dont ils disposent, et en menant une recherche, même brève, sur les sources et les analyses relatives à cette information.

Il s'agirait de l'inviter notamment à :

1.f chercher à identifier la source de l'information, s'interroger sur la confiance qu'on peut lui accorder, vérifier son référencement et croiser différentes sources sur le même sujet.

1.g tenir compte, avant de la partager, du caractère potentiellement incertain d'une information en particulier dans un contexte de crise sanitaire où beaucoup d'inconnues subsistent[25].

Renforcer l'esprit critique de l'utilisateur supposerait surtout de développer sa culture numérique tant sur le plan scientifique[26] que s'agissant du fonctionnement des outils numériques. A cette fin, de nouvelles ressources innovantes ont été développées par divers acteurs en lien avec la crise sanitaire, qui devraient être complétées par des actions à plus long terme[27].

Par ailleurs, une formation aux outils numériques devrait être apportée aux publics les plus vulnérables, en particulier les plus jeunes et les plus âgés, mais également être assurée tout au long de la vie[28]. A ce titre, une attention toute particulière devrait être portée à la compréhension des interfaces utilisées par les plateformes pour identifier une information erronée et permettre aux utilisateurs d'associer la bonne sémantique aux symboles utilisés. L'État pourrait ainsi lancer des campagnes d'éducation à large échelle. Il conviendrait également d'impliquer les plateformes dans ce processus de formation de l'ensemble de leurs utilisateurs.

Recommandation à l'État et aux plateformes :

1.7 Diffuser une infographie claire et accessible à tous détaillant les étapes de la réflexion à mener concernant la qualité de l'information avant de la repartager.

Points d'attention :

1.h La crise COVID-19 a confirmé l'importance de sensibiliser et d'éduquer l'ensemble des citoyens sur les enjeux de la désinformation et de la mésinformation, particulièrement amplifiée par l'usage d'outils numériques. Il convient par conséquent de mener une réflexion d'ampleur sur le sujet.

1.i Il semble particulièrement important de concevoir des campagnes de formation relatives aux outils numériques, dans le cadre de la formation initiale et tout au long de la vie, y compris pour les plus âgés. Ces formations doivent permettre aux utilisateurs de mieux évaluer la qualité des sources d'information et de maîtriser leur rôle viral.

25 À ce titre, v. par exemple l'infographie diffusée par *Infographie International Federation of Library Association and Institutions (IFLA)* : french_-_how_to_spot_fake_news_0.pdf

26 Dans sa résolution du 21 février 2017, sur les sciences et le progrès dans la République, l'Assemblée nationale plaide pour que « la culture scientifique soit le ferment indispensable pour des citoyens éclairés et responsables ».

27 Par exemple, l'AMCSTI (réseau professionnel des cultures scientifiques, techniques et industrielles – www.amcsti.fr).

28 En ce sens, v. CNCDH, Avis relatif à la proposition de loi visant à lutter contre les contenus haineux sur internet, 9 juil. 2019, p. 8 : final_avis_relatif_a_la_ppl_lutte_contre_la_haine_en_ligne.pdf. V. également la Recommandation du CSA n° 2019-03 du 15 mai 2019 préc., n°6.

II. LE RÔLE DES AUTORITÉS

Si la modération des contenus et le contrôle de la viralité jouent un rôle prépondérant dans le contrôle pragmatique de la désinformation et de la mésinformation, ces opérations soulèvent d'autres questionnements éthiques relatifs au rôle joué par différentes autorités dans ce processus. Cela interroge tout d'abord l'autorité ainsi acquise par les plateformes et le contrôle qui devrait en résulter (2.1). Ensuite, il apparaît que ces opérations ne peuvent se passer d'instances qui identifient les informations recevables et celles qui ne le sont pas. Différentes questions émergent alors s'agissant de la légitimité dont jouissent ces instances dès lors qu'elles sont considérées par les plateformes comme contribuant à établir, *hic et nunc*, la vérité (2.2).

A. L'autorité acquise par les plateformes

Premières responsables de l'identification de désinformation ou de mésinformation et des réponses à y apporter, les plateformes acquièrent une très grande autorité sur le partage d'information. Or leurs pratiques habituelles de modération ont été en partie transformées par la crise COVID-19 (voir supra). Différents niveaux de questionnement peuvent en résulter.

Les multiples mesures mises en œuvre par les plateformes dans le cadre de la lutte contre la crise sanitaire (suppression, réduction de visibilité et promotion de contenus) interrogent ainsi l'évolution possible de leur fonction d'intermédiaire technique. En effet, ces mesures remettent largement en cause la position longtemps défendue par ces plateformes consistant à dire qu'en tant que simples diffuseurs de contenus sans rôle éditorial, elles n'auraient pas à intervenir sur ce qui est publié par leurs abonnés, permettant ainsi à tout un chacun d'exprimer ses idées sans sélection. Mais cet argument a commencé à être battu en brèche à la suite de la vague d'actes terroristes des dernières années. Par l'appel de Christchurch, en mai 2019, ces entreprises se sont en effet engagées à ne pas diffuser sur leurs plateformes des contenus à caractère terroriste. Cette évolution est confirmée par la crise sanitaire de la COVID-19. La mise en visibilité et l'affichage de certains contenus ont en effet atteint des proportions jamais atteintes auparavant. L'accentuation de ces pratiques éditoriales peut avoir plusieurs conséquences. Par exemple, si des choix éditoriaux, comme la promotion d'informations officielles ne sont pas rendus visibles pour l'utilisateur, la neutralité des moteurs de recherche pourrait être remise en question.

Recommandation aux plateformes :

2.1 Indiquer clairement à leurs utilisateurs que certaines propositions d'information résultent de choix éditoriaux, qui peuvent changer en temps de crise.

Plus généralement, l'autorité que peuvent exercer les plateformes lorsqu'elles définissent leur politique de modération de contenus peut faire l'objet de diverses tensions éthiques.

On peut considérer que chaque plateforme doit être en mesure d'agir comme elle l'entend, sous réserve de se conformer à ses obligations légales. Cela peut alors les conduire à prendre des positionnements différents, conformément à leurs propres intérêts économiques et politiques, tout en arguant du rôle qu'elles pourraient jouer en tant que gardienne de la démocratie ou de la liberté d'expression de leurs utilisateurs (voir les différences de traitement des messages du président Trump par Facebook et Twitter pendant l'actuelle campagne présidentielle). En ce sens, on rappellera que ces opérateurs économiques ne sont pas qualifiés de média et ne sont donc pas soumis à une obligation de pluralisme ; ils se distinguent également des opérateurs de communication électroniques qui se voient imposer une obligation de neutralité dans l'acheminement des contenus[29].

On peut encore considérer que certaines plateformes – comme les grands réseaux sociaux – constituent de nouvelles agoras numériques, des lieux de l'expression publique. De plus, ces agoras voient l'interaction d'un nombre croissant de *bots* (utilisateurs artificiels) ayant un grand pouvoir de persuasion à travers les contenus qu'ils génèrent et font, pour certains d'entre eux, courir le risque de diffuser massivement des contenus pouvant avoir des finalités de déstabilisation économique et politique. Ceci pourrait commander une révision de leur statut, d'autant plus lorsque ces plateformes deviennent l'un des instruments d'une politique de santé publique comme en période de crise COVID-19.

On peut également interroger la légitimité des plateformes à évaluer la licéité d'un contenu, et à décider seules de son éventuel retrait, dès lors que cela revient à consacrer une forme de justice privée et à accroître des phénomènes de censure constituant autant d'atteintes à la liberté d'expression. Cela conduit alors à penser le rôle du juge qui, en sa qualité de garant des libertés fondamentales, ne saurait être relégué au second plan. Mais cela impose alors d'évaluer l'effectivité de son contrôle compte tenu de la masse des contenus visés et de leur vitesse de propagation, ou encore eu égard à sa portée territoriale potentiellement limitée alors que la plupart des plateformes opèrent à échelle mondiale et ne se limitent pas à un territoire national.

Par ailleurs la responsabilité de ces opérateurs peut être interrogée, tant lorsqu'ils procèdent à ce type de choix éditoriaux que lorsqu'ils contribuent, par défaut, à la propagation de désinformation et de mésinformation[30]. Cette responsabilité est actuellement limitée dès lors qu'ils bénéficient de la qualification d'hébergeur au sens de

29 Sur ce point, v. l'étude annuelle du Conseil d'État, Numérique et droits fondamentaux, 2014, p. 217&s.

30 En ce sens, Créer un cadre français de responsabilisation des réseaux sociaux, rapport préc.

la loi n° 2004-575 du 21 juin 2004 pour la confiance de l'économie numérique transposant la directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 sur le commerce électronique. Différentes discussions sur une redéfinition possible de la responsabilité de ces plateformes sont d'ailleurs engagées, en Europe[31] et aux États-Unis[32]. Quoiqu'il en soit, leur responsabilité en cas de retrait ou de non retrait de contenus devrait alors être pensée dans le respect de liberté d'expression, ce que rappelle la récente censure par le Conseil constitutionnel de la loi visant à lutter contre les contenus haineux sur internet[33].

Point d'attention :

2.a Promouvoir la réflexion sur la redéfinition de la responsabilité des plateformes aux niveaux national et européen dans le respect de la protection de la liberté d'expression.

Une autre difficulté porte sur le contrôle de ces nouvelles autorités. La question n'est certes pas nouvelle mais pourrait se reposer à l'aune de l'accroissement du rôle joué par les plateformes dans le traitement de l'information à l'occasion de la crise Covid-19. L'Union européenne se montrait jusqu'alors plutôt favorable à une autorégulation également promue par les plateformes (v. la promotion des guides de bonne conduite[34]) alors que certains États, à l'image de la France, ont préféré instituer un contrôle par l'autorité publique (v. s'agissant de la désinformation, la loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information). D'autres encore proposent que ce contrôle soit exercé par des autorités indépendantes à la fois des plateformes et de l'autorité publique[35]. L'existence d'un tel contrôle soulève de nombreux questionnements éthiques. Il conviendrait alors de prendre en compte les bénéfices et les risques soulevés par des actions de contrôle, notamment en matière de liberté d'expression ainsi que les limites à ne pas dépasser. Il faudrait en outre évaluer les effets et les limites de l'autorégulation des plateformes s'agissant de la lutte contre la désinformation et la mésinformation et se demander si une éthique inspirée de la déontologie journalistique pourrait être appliquée aux réseaux sociaux. Il serait nécessaire de penser de nouvelles formes de régulation, notamment par une autorité indépendante, bien que son rôle puisse être délicat.

31 A cet égard, v. la consultation en cours sur la future législation sur les services numériques : https://ec.europa.eu/commission/presscorner/detail/fr/ip_20_962

32 https://www.lemonde.fr/pixels/article/2020/05/28/dans-sa-charge-contre-twitter-donald-trump-veut-changer-le-regime-de-responsabilite-des-reseaux-sociaux_6041052_4408996.html

33 Conseil constitutionnel, Décision 2020-801 DC du 18 juin 2020, préc.

34 Lutter contre la désinformation en ligne : une approche européenne, Communication de la Commission européenne COM(2018) 236 final, 26 avril 2018

35 En ce sens, v. notamment le rapport Créer un cadre français de responsabilisation des réseaux sociaux : agir en France avec une ambition européenne, préc., quatrième pilier.

Points d'attention :

2.b Mener une réflexion d'ensemble sur le contrôle des plateformes et en particulier sur la consécration d'une nouvelle autorité chargée de leur régulation, ou sur le renforcement du rôle d'une autorité indépendante existante en charge de leur régulation comme le CSA qui pourrait devenir un Conseil supérieur de l'audiovisuel et du numérique.

2.2 Permettre aux utilisateurs et à la société civile de s'organiser pour s'imposer comme un interlocuteur à part entière de ces plateformes numériques dans un souci d'autonomie et de responsabilisation de tous les acteurs, citoyens, associations, entreprises aux côtés des institutions démocratiques.

B. Les autorités sur lesquelles s'appuient les plateformes

Afin de contrôler et vérifier les informations circulant sur Internet, en particulier en cas de crise sanitaire, les plateformes numériques doivent pouvoir les comparer à des informations émanant de sources considérées étant sûres ou légitimes : "vérificateurs de faits", les services gouvernementaux, et le Service statistique public qui fournit l'immense majorité des chiffres concernant la pandémie. Or les rapports avec ces instances peuvent, eux-aussi, être source d'un certain nombre de difficultés.

Pour mieux discriminer les informations faisant autorité des autres, les plateformes peuvent par exemple s'appuyer sur le travail des vérificateurs de faits (« *fact checkers* »). Différents acteurs s'organisent au niveau national à l'image, en France, des Décodeurs du Monde, *Checknews* de Libération, *Fake off* de 20 minutes ou Factuel de l'AFP. Des associations internationales voient également le jour comme le *European Digital Media Observatory* (EDMO), soutenu par l'Union Européenne[36]. Il s'agit souvent d'émanations d'organes de presse, d'universitaires, ou d'acteurs de la société civile. Ces acteurs enquêtent sur des contenus viraux et tentent d'évaluer leur degré de véracité. Pour cela, ils cherchent à établir les faits ou, au contraire, à montrer que les informations manquent de fondement et s'efforcent d'identifier les sources et les réseaux qui ont l'habitude de produire et ou de diffuser de fausses informations. Ils n'en sont pas moins pris dans des tensions complexes. D'abord, certains groupes sont financés par les plateformes elles-mêmes ce qui fragilise leur indépendance. La vérification des faits est en effet coûteuse dès lors qu'elle implique d'entretenir d'importantes bases de données et de rémunérer des équipes qualifiées pour les gérer. D'autre part, ces vérificateurs n'ont pas accès à la totalité des informations qui circulent sur les plateformes – ils ne peuvent voir, en particulier, les informations partagées au sein de groupe privés ou de messageries – ce qui constitue un grand nombre d'angles morts. L'Union Européenne promeut en ce sens des échanges plus fluides d'informations entre les plateformes et les vérificateurs. L'autorité sous l'égide de laquelle les vérificateurs agissent peut en outre être discutée et

36 Communication de la Commission européenne Lutter contre la désinformation concernant la COVID-19 préc., partie 5.2. Soutien aux vérificateurs de faits et aux chercheurs, p.11.

générer chez l'utilisateur le rejet de la qualification du contenu en « désinformation » ou « mésinformation ». Ainsi, l'utilisateur pourrait considérer que la volonté même de cacher l'information, émanant d'une autorité contestée, atteste de sa légitimité. Enfin, de nombreuses associations de vérificateurs sont constituées principalement de journalistes, qui peuvent être désarmés face à certaines informations ou controverses d'ordre scientifiques, en particulier pendant la crise Covid-19.

Recommandations :

2.3 Faciliter le transfert d'informations entre les plateformes et les vérificateurs.

2.4 Renforcer le pluralisme des équipes de vérificateurs afin que les chercheurs et la société civile puisse y être représentés.

S'agissant tout particulièrement de la promotion de certains contenus, il est également possible d'interroger la neutralité de l'État dès lors que les informations qu'il promeut sont aussi celles qui légitiment l'action du gouvernement. Laisser les entreprises en discussions exclusives avec lui fait courir d'importants risques de censure. Les arguments qui questionneraient certaines des décisions politiques du gouvernement pourraient être écartés ou supprimés injustement. L'exemple de la page « Désinfox information » en est une alerte. Celle-ci a été mise en place par le gouvernement, mais retirée quelques heures après le dépôt d'un référé liberté auprès du Conseil d'État par le syndicat national des journalistes qui y voyait une « atteinte grave au pluralisme"[37].

Par ailleurs, cette exclusivité risquerait d'engendrer un certain nombre d'ambiguïtés dans les relations entre ces plateformes et les autorités étatiques, en particulier celles avec le gouvernement. Les plateformes pourraient par exemple avoir recours au gouvernement pour valider ou modérer certaines informations, et ce dernier leur demander de promouvoir certains contenus visant à lutter contre la crise sanitaire. Il existerait alors un réel risque de connivence qui fournirait aux plateformes des appuis qu'elles pourraient mobiliser lorsqu'il serait question de contrôler leurs pratiques dans d'autres domaines (s'agissant par exemple de leurs pratiques concernant l'information des utilisateurs sur leurs modes de fonctionnement).

Recommandation :

2.5 Publier les mécanismes de modération de contenus mis en œuvre lors de la crise sanitaire par les plateformes, en particulier ceux qui garantissent la transparence des interactions entre ces opérateurs et les autorités publiques, et exercer un contrôle *ex post* de ces mécanismes par l'autorité compétente, dont le juge en tant que garant des libertés individuelles.

37 refere-liberte---04-05-2020.pdf, demande rejetée par l'ord. Conseil d'État, 8 juin 2020 relevant que "le Premier ministre a supprimé cette page Internet, à compter du 5 mai 2020, soit postérieurement à l'introduction de la requête » ce dont il résulte que « les conclusions de cette requête ont perdu leur objet et il n'y a plus lieu d'y statuer » tout en condamnant l'État aux frais de procédure – v. également l'intervention du Ministre de la Culture Franck Riester annonçant le retrait de cette page : Questions au Gouvernement, 5 mai 2020.

Par ailleurs, la crise sanitaire que nous traversons a la particularité d'être, à un niveau rarement atteint, perçue à travers quantité de chiffres et des statistiques pour une très grande part produits par le service statistique public (SSP), comme le nombre de morts, de personnes infectées, de personnes soignées, à l'hôpital, dans les EHPAD, en population générale. La plupart des mesures politiques, des discours et des réflexions individuelles sur cette épidémie est orientée et confortée par des outils quantitatifs qui reposent sur des définitions et des méthodes qui en spécifient la portée à l'image du nombre de morts communiqué, qui n'est, au mieux, qu'une approximation de la réalité. Les limites de la portée des outils statistiques ne sont que rarement mises en avant alors que ces chiffres sont relayés très largement. Le manque de mise en contexte de ces chiffres peut être source d'interprétations participant à une forme de désinformation.

Recommandations :

2.6 Accompagner la communication des statistiques relatives à l'épidémie d'un discours méthodologique rappelant notamment le contexte et les limites des résultats obtenus.

2.7 Publier des réflexions non seulement sur les méthodes de production des statistiques, mais aussi sur leurs usages et les transformations qu'elles subissent au fur et à mesure de ces réappropriations : comment elles sont utilisées, transmises et parfois déformées, comment elles influencent les comportements du gouvernement ou du public.

Enfin, le recours à des autorités établies pour promouvoir des contenus scientifiques présentés comme certains ne doit pas conduire à sous-évaluer le caractère controversé de ceux-ci[38]. L'OMS, par exemple, semble admise comme une source sûre de résultats scientifiques par les plateformes, alors que d'autres autorités scientifiques la contestent, parfois à bon droit[39]. D'autres acteurs, comme les utilisateurs, les scientifiques ou encore les associations pourraient dès lors être impliqués dans la sélection des informations mises en avant[40].

38 Sur les enjeux éthiques liés aux contre-vérités scientifiques, la post-vérité et la communication de la science dans la sphère publique, v. déjà l'avis n° 2018-37 du COMETS « Quelles nouvelles responsabilités pour les chercheurs à l'heure des débats sur la post-vérité ? » publié le 12 avril 2018.

39 A cet égard, v. la tribune publiée dans Le Monde et signée par de nombreuses autorités qui appelle à développer le « le Forum sur l'information et la démocratie, créé en novembre 2019 par onze organisations, think tanks et centres de recherche de neuf pays, pour mettre en œuvre le Partenariat » entre les plateformes et les acteurs sociaux (« Nous appelons les géants du Web à un sursaut décisif pour le droit à l'information fiable », Le Monde, 02/05/2020. Signée entre autres par Joseph Stiglitz, Christophe Deloire et Shirin Ebadi).

40 En ce sens, v. Créer un cadre français de responsabilisation des réseaux sociaux : agir en France avec une ambition européenne, rapport préc.

ANNEXES

Personnes auditionnées

- **Serge Abiteboul,** membre de la mission de régulation des réseaux (2019) et membre du collège de l'ARCEP
- **Lucien Castex**, Secrétaire général d'Internet Society France
- **Guillaume Champeau,** directeur du service Éthique et Affaires juridiques, **Leonard Cox,** vice-président des Affaires publiques et RSE, **Jean-Claude Ghinozzi,** Président-directeur général **et Sébastien Ménard**, Conseiller en stratégie, QWANT
- **Guillaume Goubert,** directeur du journal La Croix
- **Béatrice Oeuvrard,** chargée des affaires publiques, Facebook
- **Audrey Herblin-Stoop,** chargée des affaires publiques, Twitter
- **Jonathan Parienté**, journaliste au Monde, chef du service des Décodeurs
- **Ramón Ruti**, co-fondateur et CTO de Storyzy

Composition du groupe de travail ayant contribué à l'élaboration de ce document

Laurence Devillers	Claude Kirchner
Emmanuel Didier*	Jérôme Perrin
Karine Dognin-Sauze	Catherine Tessier
Christine Froidevaux	Serena Villata*
Eric Germain	Célia Zolynski*
Alexei Grinbaum	
Jeany Jean-Baptiste	**corapporteurs*

Les membres du Comité national pilote d'éthique du numérique

Gilles Adda	Christine Froidevaux	Christophe Lazaro
Raja Chatila	Jean-Gabriel Ganascia	Gwendal Le Grand
Theodore Christakis	Eric Germain	Claire Levallois-Barth
Laure Coulombel	Alexei Grinbaum	Caroline Martin
Jean-François Delfraissy	David Gruson	Tristan Nitot
Laurence Devillers	Emmanuel Hirsch	Jérôme Perrin
Karine Dognin-Sauze	Jeany Jean-Baptiste	Catherine Tessier
Gilles Dowek	Claude Kirchner - directeur	Serena Villata
Valeria Faure-Muntian	Augustin Landler	Célia Zolynski

La publication de ce bulletin a été validée le 8 juillet 2020 lors de l'assemblée plénière incluant Emmanuel Didier (membre du CCNE) en tant qu'invité avec 16 voix pour, 1 voix contre et 2 abstentions.

Contact presse : communication@comite-ethique.fr