



Réseaux sociaux et modération, quel est le souci ?

Les réseaux sociaux apparaissent aujourd'hui comme bien plus que de simples plateformes de discussion ou de divertissement. Ils sont aussi des vecteurs d'informations auxquels chacun peut avoir accès librement. Chez Flint, nous réfléchissons aussi bien à la manière dont est produite l'information qu'à la façon dont nous la recevons, la consommons. Ce newskit est le premier d'une série qui s'attache à observer la manière dont les réseaux sociaux régulent l'information et les discussions. Comment fonctionne la modération des grandes plateformes ? Comment a-t-elle évolué au fil du temps ? Qui la critique ? Pour quelles raisons ?



Pourquoi c'est intéressant ? Parce que cela permet de mieux comprendre comment chacun de nous reçoit l'information sur les réseaux sociaux, et comment la modération qui en est faite influe sur ce que l'on voit ou non.

En bref (clique sur les liens pour avoir accès à la partie concernée)

- ① Les réseaux sociaux ont **transformé l'accès à l'information et nos possibilités de pouvoir discuter en public**. Ce que nous y recevons et la manière dont nous le recevons est régi par des algorithmes, qui s'adaptent à nos comportements.
- ② **Si les États-Unis ont une perception très large de la liberté d'expression, en France, cette dernière n'est pas absolue**. Or, assurée par des humains et des robots, les **décisions de modération** entrent quelquefois en contradiction avec des discours de défense de la liberté d'expression.
- ③ C'est le **paradoxe de la gestion des contenus** sur les réseaux. Si les outils et les algorithmes de modération propres à chaque plateforme se développent depuis plusieurs années, des critiques subsistent : gestion insuffisante des fausses informations, des contenus clivants et haineux, notamment envers les femmes et les minorités, persistance de contenus pédocriminels, **accusations de biais** politiques ou discriminants...



Les réseaux sociaux, la liberté d'expression et la juste information

Internet et les réseaux sociaux ont chamboulé du même coup l'accès à l'information et les possibilités de discussion publique. Avant leur développement, l'accès aux médias, à une certaine forme de porte-voix était géré par des *gatekeepers*, des "portiers" de l'information : les journalistes, les curateurs, les éditeurs, etc. Sur les réseaux sociaux, ce ne sont plus ces personnes qui trient, hiérarchisent et restituent l'information : ce sont des algorithmes qui s'en chargent en remontant telle discussion, tel élément, tel groupe dans tes fils d'actualité (sur l'évolution du *gatekeeping* la sphère publique, lire par exemple les spécialistes du numérique [danah boyd et Laura Reed](#)

Dans [Culture numérique](#), le sociologue **Dominique Cardon** détaille la **spécificité des algorithmes** et des métriques (likes, partage, commentaires) régissant les réseaux sociaux : comme ces dernières sont visibles, nous adaptons nos comportements pour les faire grimper. Ce qui veut dire que si nous constatons qu'un type d'information ou une manière de la présenter fonctionne particulièrement bien avec l'algorithme d'une plateforme, nous avons tendance à nous y adapter et à poster de nouvelles publications qui y correspondent.

En résumé, nous utilisons les réseaux sociaux pour passer le temps, discuter (la promesse de Facebook est de rester en contact avec nos proches) **mais aussi nous informer** (celle de Twitter est de savoir "ce qu'il se passe"). Tout ce que nous recevons et la manière dont nous le recevons est trié par des algorithmes auxquels nous nous adaptons ([danah boyd](#), [Le Point](#)). Effet pervers : ces lieux numériques créent des effets de bulle et sont le théâtre de l'expansion de théories complotistes, du partage de fausses informations, de cyberviolence et de grandes difficultés à débattre.

Pour [Le Monde](#), la philosophe Monique Canto-Sperber analyse que *"la parole déchaînée qui prospère [sur ces plateformes] et la multiplication des velléités de censure sont un symptôme manifeste du fait que quelque chose ne va plus avec notre concept de liberté d'expression. La liberté de parler s'est perdue dans la viralité."*



Liberté d'expression, une notion mais plusieurs cadres

[L'article 19](#) de la Déclaration universelle des droits de l'Homme, signé par les Nations Unies en 1948, garantit à tous *"la liberté d'opinion et d'expression, ce qui implique le droit de ne pas être inquiété pour ses opinions et celui de chercher, de recevoir et de répandre, sans considérations de frontières, les informations et les idées par quelque moyen d'expression que ce soit"*.

[The Conversation](#) rappelle qu'aux États-Unis, **la liberté d'expression est garantie par le Premier amendement de la Constitution**, auquel chaque Américain accorde *"une vénération quasi religieuse"*. Cet amendement protège également tout discours de haine (sauf quelques exceptions, comme par exemple si les propos traduisent [un danger clair et présent](#)).

En France, **la liberté d'expression est garantie par la Déclaration des droits de l'Homme et du citoyen** ([articles 10-11](#), 1789) mais est aussi encadrée par plusieurs textes, comme le souligne [Vice](#) : la Déclaration universelle des droits de l'homme et la convention européenne des droits de l'homme ([article 10](#), 1953) entre autres. Contrairement aux textes américains, les lois françaises définissent des abus de cette liberté d'expression : la diffamation, l'injure, l'incitation à la haine et à la discrimination, l'atteinte à la vie privée et l'apologie du terrorisme sont punis par la loi.

[La charte des droits fondamentaux de l'Union européenne](#) encadre également cette liberté d'expression, en garantissant à tout citoyen européen une liberté d'opinion et une liberté de recevoir ou communiquer des informations *"sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontières"*.



Qui est responsable ?

Chaque réseau social est responsable de ses propres règles de modération

(inscrites dans ses Conditions générales d'utilisation), mais celles-ci diffèrent selon le pays où elles s'appliquent. Aux États-Unis, [la section 230 du Communication Decency Act](#)^{🇺🇸}, inscrit dans le code civil américain, autorise les entreprises du numérique à effectuer librement une modération de leurs contenus selon leurs critères (à lire dans [Wrap](#)^{🇺🇸}, ou sur [MSN](#)). Cela signifie que ces plateformes ne peuvent pas être poursuivies en justice pour avoir suspendu, supprimé ou laissé visible un contenu de leur plateforme (à mettre en parallèle du Premier amendement, cité plus haut). Cet été, des sénateurs démocrates ont proposé une loi, *The Health Misinformation Act*, qui rendrait responsables les plateformes en cas de publication de fausses informations concernant la santé ([New York Post](#)^{🇺🇸}), et qui créerait donc une exception à cette section 230.

En France, explique Nicolas Arpagian, expert en cybersécurité, à [franceinfo](#), **les réseaux sociaux ont le statut d'hébergeur, et non d'éditeur** (statut des médias). Ainsi, s'ils ne sont pas responsables de la publication des contenus, ils sont tenus par [la loi](#) d'agir en cas de signalement d'un contenu jugé problématique par un utilisateur. Il est toutefois précisé que les conditions de retrait d'une plateforme "*peuvent ne pas correspondre aux lois applicables en France et varient d'un hébergeur à l'autre*". [La loi pour la confiance dans l'économie numérique](#) (LCEN, 2004) oblige les réseaux sociaux à rendre accessibles tous les documents relatifs à leur modération indique [Siècle Digital](#), ce que tous ne respectent pas forcément de bonne grâce.

Ces plateformes ont aussi signé un "code de bonne conduite" avec la Commission européenne en 2016, explique [Nextinpact](#), dans lequel elles s'engagent à "*supprimer ou rendre inaccessible*" tout contenu haineux.

Dans une lettre ouverte de 2019, adressée à plusieurs ministères et institutions juridiques, la [Ligue des Droits de l'Homme](#) rappelle que de son point de vue, "*seule la décision de justice, issue du pouvoir de l'Etat institué démocratiquement, est acceptable quand il s'agit de censurer un propos*". Confier cette responsabilité aux opérateurs des réseaux sociaux fait courir un risque de "*privatisation des fonctions judiciaires et de mettre à mal les garde-fous démocratiques pour nos citoyens*".



Aux manettes : des robots, mais surtout des humains ?

La plupart des réseaux sociaux s'appuient sur des modérateurs humains, qui ont pour mission d'analyser chaque publication signalée par un utilisateur, et juger si elle viole les règles d'utilisation du réseau social. Selon [Technology Review](#)^{🇺🇸}, en 2020, Facebook comptait sur 15 000 modérateurs pour assurer ces opérations, dont une grande majorité était employée par des sociétés tierces. Selon [Mashable](#)^{🇺🇸}, média spécialisé dans l'univers numérique et les technologies, Youtube employait 10 000 modérateurs en 2020. Pour Twitter, Michaël Szadkowski, journaliste [au Monde](#), souligne le peu d'information dont on dispose : on ne connaît rien du profil des modérateurs, de leur formation, de leur manière de fonctionner, ni de leur nombre (en 2021, [Capital](#) estime à 1 500 le nombre de modérateurs Twitter). Dans [Le Monde](#), une source interne au forum Jeuxvideo.com évoque une dizaine de modérateurs pour le site.

La profession de modérateur n'est pas sans risque pour la santé mentale de ces personnes, soulignent les articles et documentaires de [The Verge](#)^{🇬🇧}, [TechCrunch](#)^{🇬🇧}, du [Washington Post](#)^{🇬🇧}, de [La Revue des Médias](#), d'[Arte](#) ou de [Programme B](#). Depuis plusieurs années, **les réseaux sociaux ont automatisé une partie de leur modération** ([Le Monde](#) sur Twitter, [The Guardian](#)^{🇬🇧} sur Facebook). Ces algorithmes sont capables d'analyser le contenu des publications et de les supprimer en cas de violation des règles, sans qu'elles n'aient été signalées par les utilisateurs. En mars 2020, pour pallier les difficultés de modération liées aux restrictions sanitaires, Facebook, Twitter et Youtube ont accentué leur modération algorithmique, rapporte [Le Figaro](#).

? Les paradoxes de la modération

[Au Monde](#)^{🇫🇷}, la philosophe Monique Canto-Sperber signale qu'aujourd'hui, les algorithmes des réseaux sociaux créent "*une distorsion assez rapide du débat, faussé d'avance*". Le but des entreprises est d'augmenter le temps passé par les utilisateurs sur leurs plateformes et ainsi "*la poursuite de la croissance du cours de l'entreprise en Bourse*", selon l'économiste Joëlle Toledano. Pour y parvenir, leurs algorithmes favorisent les contenus qui polarisent le plus les discussions.

Depuis quelques années, explique l'hebdomadaire belge [Le Vif](#), **ces algorithmes sont accusés de favoriser la diffusion de fausses informations** en les rendant plus visibles que les informations jugées fiables que ce soit sur Youtube ([franceinfo](#)), Facebook ([Socialnetlink](#), [L'ADN](#)), Twitter ([Science](#)^{🇬🇧}, [Slate](#)), ou TikTok ([Femmes Actuelles](#)). L'ONG [Avaaz](#)^{🇬🇧} qualifie même l'algorithme de Facebook de "*menace pour la santé publique*".

La viralité des contenus participe à la visibilité d'informations fausses, quand bien même leur nombre est relativement faible - 0,15% du contenu médiatique global selon [Science Advance](#)^{🇬🇧} (lire aussi [The Conversation](#)). Certains chercheurs préconisent donc au public de... ne rien faire face à une fausse info ([The Conversation](#)^{🇬🇧}) : la commenter, même pour la corriger, ne reviendrait qu'à la faire grimper dans les métriques algorithmiques, donc à lui donner de la visibilité ([The Conversation](#)^{🇬🇧}) (pour les journalistes, c'est plus compliqué à trancher, car "une rumeur fausse est un fait vrai", rappelle le philosophe Luc de Brabandère [dans Usbek & Rica](#)).

Les réseaux sociaux sont régulièrement accusés d'être le "terreau de radicalisation" de terroristes, comme le mentionne [France Culture](#) ou [Le Temps](#), en laissant se diffuser des messages de haine et des vidéos violentes. Cela représente un danger notamment pour les plus jeunes, très présents sur ces plateformes relate [NBC News digital](#)^{🇬🇧}.

La modération de contenus et comptes pédocriminels reste un fort enjeu pour les plateformes, notamment pour Youtube dont l'algorithme a été accusé en 2019 de favoriser la visibilité de ces contenus ([Daily Geek Show](#)^{🇬🇧}, [Télérama](#)) ou encore Twitter qui laisserait se développer des communautés pédocriminelles en toute impunité ([Numerama](#)).

Le harcèlement est une autre problématique importante, puisqu'il se déroule majoritairement sur les réseaux sociaux, selon le [Pew Research Center](#)^{🇬🇧}. En 2019,

montre [Statista](#), 81% des 18-30 ans victimes de cyber-violences les ont vécues sur Facebook. Le confinement de l'an passé a entraîné une augmentation des messages haineux et discriminatoires sur les réseaux sociaux, ajoute [20minutes](#). En mars 2020, 14% des messages postés sur l'ensemble des réseaux pouvaient être qualifiés de "haineux".

Ce cyberharcèlement touche particulièrement les femmes comme le montrent ce rapport de l'[Agence pour les droits fondamentaux de l'Union européenne](#) et celui de [la Commission des Nations Unies](#). Ces événements s'accompagnent d'impacts directs sur leur vie selon ce sondage d'[Amnesty International](#) (2017), les obligeant même parfois à quitter les réseaux sociaux, de manière temporaire ou définitive ([Marie Claire](#), [RTL](#), [The 19th](#)^{🇬🇧}, [Femmes Actuelles](#), [Arte](#)). En 2017, l'ONG a estimé à 1,1 million les tweets problématiques adressés à des femmes, soit un toutes les trente secondes, ajoute [Mashable](#)^{🇬🇧} (les femmes noires étaient à 84% plus concernées que les femmes blanches). En 2016, une [étude du Centre Hubertine Auclert](#), pour l'égalité femmes hommes montrait que chez les 12-15 ans, les filles étaient plus touchées par le cybersexisme que les garçons (1,5 à 2 fois plus).

Y-a-t-il des biais dans la modération ?

Selon une étude parue dans la revue [Intelligence artificielle, pratiques sociales et politiques publiques](#), l'intervention de l'IA dans la modération entraîne une discrimination accrue des minorités ethniques, sexuelles et de genre.

Aux États-Unis, les républicains accusent de longue date les principales plateformes américaines de "biais anti-conservateur" ([Le Monde](#)), ce qu'[une étude du NYU Stern](#)^{🇬🇧} Center démontre faux. Le [Pew research center](#)^{🇬🇧} montre que tous les bords politiques se sentent censurés.

Mais les plateformes ne sont pas accusées des mêmes maux dans leur manière d'organiser leur modération. Dans les prochains épisodes de cette série consacrée aux réseaux sociaux, nous aborderons le cas de Twitter, de Facebook, de Youtube, de TikTok et de Telegram. Notre but est d'analyser en détail leur politique de modération, quelles en sont les critiques et surtout de quelle façon elle joue sur l'information que nous recevons

→ Dites-nous si vous souhaitez voir une autre plateforme traitée !

*1er octobre
Par Tanguy Oudoire*

BROUILLON

Chiffre de fréquentations de différents services
<https://www.blogdumoderateur.com/une-minute-internet-2021>

Franceinfo > [Coronavirus : comment les réseaux sociaux ont tenté de soigner le mal de la désinformation](#)

- Quelques mois après avoir annoncé une plus grande automatisation de sa modération, Youtube a fait marche arrière après avoir constaté un bond du nombre de vidéos supprimées, dont beaucoup par erreur (+25% qu'en temps normal, [Le Monde](#)). Pour Neal Mohan, l'un des dirigeants de Youtube, les modérateurs humains "*prennent des décisions plus nuancées, notamment quand il s'agit de discours haineux, de désinformation médicale ou de harcèlement*". Ce risque de sur-censure avait été identifié par Monika Bickert, directrice de la politique des contenus de Facebook, en 2016 [au Figaro](#).

- En 2018, une enquête de [la Fédération internationale des journalistes](#) démontrait que les femmes journalistes étaient les plus gravement touchées par ces violences en ligne (64% des interrogées). La même année, le collectif "Prenons la une", qui milite pour la parité dans les médias et l'égalité dans les rédactions, réclamait dans [Le Monde](#) une plus grande protection des femmes journalistes. Constat confirmé en 2019 avec les révélations autour de la Ligue du LOL, évoquée ici par [La Tribune](#) ou [Novethic](#), ou en 2021 par ces témoignages sur [ABC News](#).

https://www.researchgate.net/publication/350965891_Gouvernes_par_les_algorithmes_sur_les_reseaux_sociaux_Analyse_des_resistances_aux_recommandations_par_les_jeunes

, comme l'explique [France Culture](#)

Mais pour la juriste Célia Zolynski "*compte tenu de la masse des publications, un système judiciaire ne peut pas être aussi opérationnel et efficace qu'un algorithme de modération*" ([Le Monde](#)).

- Ceux-ci nous permettent d'accéder aux contenus et aux discussions qui nous intéressent le plus, mais créent aussi des effets de bulles, laissent place à de fausses informations et participent à compliquer les débats d'idées.

Piste de solution : Wikipédia

<https://www.lefigaro.fr/secteur/high-tech/les-confidences-du-fondateur-de-wikipedia-au-figaro-les-fake-news-ne-sont-pas-un-probleme-insurmontable-20210516>