

# Liberté d'expression, modération et fake news : des défis stratégiques pour les plateformes numériques

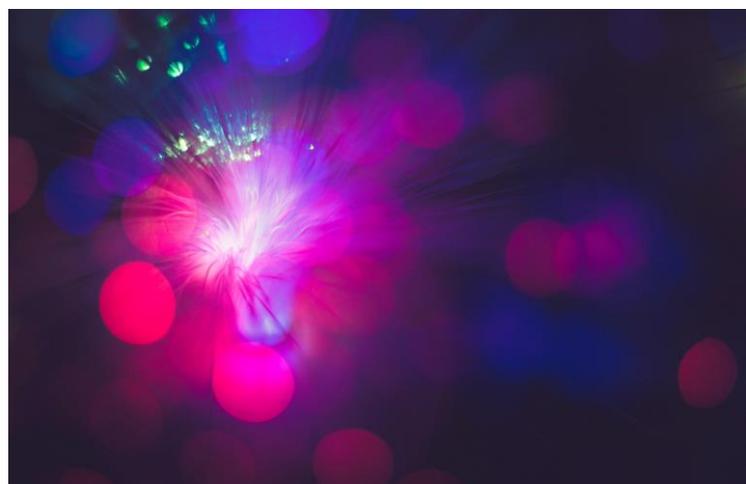
*La liberté d'expression, c'est pouvoir s'exprimer sans subir d'interférences arbitraires, notamment de la part des gouvernements ou des grandes entreprises. Ce droit fondamental peut pourtant parfois interférer avec les principes de vie privée ou de propriété intellectuelle, ou encore engendrer des dérives telles que l'intimidation ou le harcèlement. Les évolutions numériques créent les conditions d'un discours parfois difficilement contrôlable tant elles modifient la portée des propos qui peuvent rapidement devenir viraux.*

## Liberté d'expression : de nouvelles problématiques liées au numérique

**Le contrôle privé de la sphère publique** : un phénomène lié aux positions dominantes des Big Tech. En effet, les plateformes numériques que sont Google et Facebook notamment prennent chaque jour davantage de décisions liées à la liberté d'expression de leurs utilisateurs que n'en a prises la Cour Suprême des États-Unis dans son histoire.

**Le *jawboning*** : ce terme fait référence à l'action des gouvernements qui demandent aux plateformes de réguler les « lawful but awful speech » (discours légaux mais affreux) alors que leur propre législation n'est pas en mesure d'y faire face, et autorise encore la surveillance, la propagande et la censure sur Internet.

**La portée internationale des propos par rapport aux lois nationales** : les plateformes américaines fixent les principes d'expression pour le monde entier, même lorsqu'elles diffèrent des règles européennes. La politique de Facebook quant aux périodes électorales, par exemple, vise avant tout à ne pas interférer avec les discours des politiciens. Ce principe sous-entend de ne pas modérer les contenus et de s'affranchir de la notion d'égalité du temps de parole, pourtant primordiale en France.



**Le veto du perturbateur** : la modération permet de supprimer de manière unilatérale certains propos partagés sur les plateformes, et ce, parfois de manière abusive. Les communautés vulnérables – les femmes, les minorités, les transsexuels ou les dissidents – sont souvent les plus concernés par la modération des contenus.

**Liberté d'expression VS liberté d'accès** : existe-t-il un droit à l'amplification ? Les plateformes ont-elles l'obligation de donner la priorité à certains propos par rapport à d'autres et de limiter la visibilité de certaines informations ?

# Les plateformes numériques face à la multiplication des fake news

Selon M. Berriche & S. Altay, dans *Fact check : trois infox sur les infox* (2020), la population mondiale dans son ensemble est en fait assez peu exposée aux fake news, mais une petite partie d'entre elle y est beaucoup exposée, et ce, particulièrement à l'**extrême du spectre politique** (surtout à l'extrême droite).

Par ailleurs, ceux qui diffusent des fake news publient en moyenne 22 tweets par jour, contre 1 par jour pour les autres utilisateurs. Ils ont aussi tendance à utiliser davantage de pseudos, à partager leurs opinions politiques et davantage de contenus partisans, ainsi qu'à exprimer plus d'émotions négatives. En outre, les médias sociaux ont tendance à produire une **image déformée de la réalité**, les propos et contenus y étant souvent **plus extrêmes et violents**.

Les plateformes numériques disposent déjà de moyens pour gérer la diffusion de fake news : supprimer le post, **avertir les utilisateurs** (un drapeau sur le post, par exemple, contraint l'utilisateur à un second clic pour accéder au contenu – une méthode qui a d'ailleurs prouvé son efficacité lors de la pandémie, puisque 90% des utilisateurs n'ont pas consulté les 50 millions de contenus qui s'étaient vu imposer ce drapeau), ou **réduire la visibilité** des contenus dont la véracité pose question.

Imposer des règles contraignant la diffusion des contenus en ligne c'est aussi, pour les plateformes, y adosser un « **coût de réputation** », en obligeant par exemple l'utilisateur à se demander « que penserait ma famille de ce contenu ? ».

**Les plateformes numériques ont un rôle majeur à jouer pour limiter la désinformation et doivent indéniablement être poussées à signaler les contenus qui sont faux, partiellement faux ou incomplets, et à bannir les récidivistes.**

## La législation évolue pour responsabiliser les plateformes

La législation sur les services numériques est fondamentale. Grand chantier numérique de l'Union Européenne, le **règlement DSA** (Digital Services Act) sera bientôt applicable pour mieux protéger les utilisateurs européens. Il vise à encadrer les activités des plateformes numériques, en particulier celles des GAFAM, en mettant en pratique le principe selon lequel **ce qui est illégal hors ligne est illégal en ligne**.

Cette nouvelle législation fixe ainsi un ensemble de règles pour responsabiliser les plateformes numériques et **lutter contre la diffusion de contenus illicites ou préjudiciables**. Voté par le Parlement Européen en juillet 2022 et approuvé par le Conseil de l'Union Européenne le 4 octobre dernier, le DSA sera applicable d'ici février 2024, sauf les plateformes ou moteurs de recherche comptant plus de 45 millions d'utilisateurs (soit 10 % de la population européenne) qui seront concernés **dès 2023**.



# La modération : des enjeux stratégiques face à la complexité de la tâche

L'**automatisation de la modération** s'est déployée de façon spectaculaire au cours des cinq dernières années. Les algorithmes de Meta, par exemple, ont été développés à grande échelle depuis 2017, **entraînés sur les contenus signalés** par les utilisateurs ou sanctionnés par les modérateurs, et étendus depuis 2019 à la modération de contenus discriminatoires et racistes. Cela a permis à Facebook et Instagram de **passer de 25 à 95% de contenus haineux modérés automatiquement**. Pour autant, Facebook ne communique pas le nombre absolu de contenus supprimés.



Entraînés sur les contenus signalés ou dont la fiabilité est discutée par les utilisateurs en commentaire, ces algorithmes se montrent encore d'une **efficacité limitée pour modérer automatiquement les fake news**. Détecter des fake news nécessite en effet d'entraîner constamment les algorithmes sur des données labélisées – comme véridiques ou non – pour qu'ils puissent les repérer. Lorsqu'un sujet est inédit ou formulé en d'autres termes, il échappe encore à la modération. Facebook a par exemple dû entraîner spécifiquement ses algorithmes de modération sur des données de l'Organisation Mondiale de la Santé pour que la plateforme soit en mesure de modérer les contenus liés au Covid.

La gestion des fake news peut néanmoins conduire à une **modération abusive**. Facebook autorise donc des procédures d'appel pour les publications supprimées. Les 13 millions de posts supprimés de la plateforme Facebook ont donné lieu à **1,4 million de procédures d'appel**, et 30% des messages ont ainsi été republiés. En moyenne, 20 à 40 % des appels conduisent à une republication des posts sur YouTube, Facebook et Instagram.

Rachel Griffin, auteure de *The Sanitised Platform* (2022), souligne néanmoins que les processus de modération des plateformes impactent davantage les personnes issues de milieux marginalisés. Il semble en effet que ces dernières suppriment de manière disproportionnée les **points de vue « non traditionnels »**, alors qu'elles privilégient la visibilité des contenus commerciaux.

Par ailleurs, les utilisateurs se lançant dans des procédures d'appel sont principalement issus de catégories privilégiées vivant dans des régions plus développées du monde, ayant un accès régulier à internet, et sont principalement des hommes. Il est donc essentiel de **mieux informer l'ensemble des utilisateurs** quant à la possibilité d'avoir recours à des procédures d'appel lorsque la modération semble abusive.

Dans la même logique, les politiques de modération des plateformes doivent être auditées afin de **prévenir l'opacité stratégique**, car elles peuvent amener les plateformes numériques à faciliter l'accès à des contenus spécifiques qui servent leurs propres intérêts.

**Le règlement DSA imposera ainsi un audit annuel des processus de modération des plateformes.**

# Soutenir l'émergence d'une Tech responsable

Afin de contribuer à l'émergence de **meilleures pratiques dans le secteur** et accompagner les acteurs qui adoptent un comportement responsable en la matière, Sycomore AM a élaboré **un cadre d'investissement** permettant d'évaluer les **impacts et externalités** d'un projet ou d'une entreprise de la Tech.

Ainsi, selon notre analyse, une entreprise de la Tech peut être considérée responsable si son activité, ses pratiques et sa gouvernance en matière de technologie la font rentrer dans **au moins deux des catégories suivantes** :

## TECH FOR GOOD

Les produits et services proposés par l'entreprise sont-ils **destinés à avoir un impact social et environnemental positif**, dès leur conception ?

## GOOD IN TECH

L'entreprise utilise-t-elle la technologie de manière responsable pour **réduire les externalités négatives** sur les personnes et l'environnement ?

## LEVIERS D'AMÉLIORATION

Quelle est la volonté et la capacité du management d'améliorer les performances de l'entreprise sur ces dimensions **dans un avenir proche** ?

## Intégrer les aspects de l'investissement responsable dans nos décisions de gestion

Les considérations relatives à la fois à la liberté d'expression et à la lutte contre la diffusion de contenus illicites ou préjudiciables sont principalement traitées dans la dimension Good in Tech de notre approche.

Notre analyse vise notamment à s'assurer d'une utilisation éthique de la technologie, à déterminer les moyens humains et financiers que l'entreprise dédie à la **cybersécurité**, à vérifier le respect de **RPGD**, que les données personnelles des utilisateurs finaux sont chiffrées et **anonymisées**, et que l'entreprise atténue les externalités négatives liées à l'utilisation de sa technologie, telles que l'**addiction**, l'**exposition aux écrans**, ou les **troubles de l'attention**.

Cette analyse nous amène à exclure de l'univers d'investissement de notre fonds **Sycomore Sustainable Tech** des entreprises proposant des produits ou services de **surveillance massive**, ne respectant pas la **vie privée** des utilisateurs ou plus généralement les droits humains, tels que le **droit du travail**, la santé et la sécurité des collaborateurs.

## Dialoguer avec les entreprises sur la gestion de leurs impacts et externalités

Nous nous attachons par ailleurs à sensibiliser les entreprises dans lesquelles nous sommes investis à l'impact négatif que la Tech peut avoir sur l'ensemble de leurs parties prenantes, et les incitons à **adopter des standards de responsabilité**.

Récemment, nous avons ainsi encouragé **PayPal** à annoncer publiquement son engagement en faveur d'une intelligence artificielle (IA) éthique. Grâce au travail de spécialistes expérimentés sur le sujet, PayPal a déjà formalisé une **politique interne sur l'IA éthique**, et souhaite communiquer sur ses actions et engagements en la matière dans son prochain rapport RSE.

Dans la même logique, suite à la récente fuite de données subie par l'opérateur mobile américain **T-Mobile**, nous avons initié un dialogue avec l'entreprise afin de l'encourager à publier plusieurs indicateurs relatifs à la **cybersécurité**, tels que les effectifs dédiés et la durée moyenne d'interruption de ses systèmes. T-Mobile juge néanmoins que les données relatives au nombre annuel d'atteintes à la protection sont trop sensibles pour être publiées.