

PV de l'audition de Frances Haugen

Frances Haugen est une lanceuse d'alerte. Elle a travaillé chez Google, Yelp et Facebook, en tant que cheffe de produit. Ses compétences en ingénierie algorithmique lui ont fait réaliser les perturbations que les médias sociaux entraînent sur la vie démocratique. En 2021, elle est à l'origine des « Facebook Files ».

Comment encourager Facebook à développer des algorithmes plus vertueux ?

Pour Frances Haugen, le problème est de réussir à **responsabiliser Facebook sur le long terme**, car il sera impossible de résoudre définitivement les dangers que l'entreprise pose à la santé des démocraties. Elle propose donc une structure d'évaluation des risques avec une approche en 3 temps :

- Une évaluation menée par l'entreprise
- Une évaluation menée par la communauté
- Une responsabilisation rendue possible par l'accès aux données permettant de juger de la qualité de l'action de régulation de Facebook.

Les structures d'évaluation du risque existantes sont insuffisantes. En effet :

- Elles ne concernent que les contenus illégaux, alors que les menaces proviennent de contenus qui appartiennent à une **zone plus grise**.
- Elles sont uniquement de la responsabilité de l'entreprise, qui souffre d'un **manque de diversité** sociale, culturelle, géographique et linguistique, ce qui limite sa capacité à identifier certains problèmes. L'introduction d'un « auditeur », suggéré par le DSA, ne résout pas entièrement cette difficulté car il viendrait probablement du « monde de la tech ».

Il faut donc associer l'entreprise et la communauté à un régulateur capable de consulter d'autres parties prenantes. Il faut bien prendre en considération **qu'il n'existe pas de Facebook unique, mais une multiplicité de Facebook différents selon les utilisateurs** : un effort d'intégration de perspectives différentes est nécessaire.

A cet égard, il est important de se départir de l'illusion selon laquelle une IA magique pourrait retirer tous les « mauvais contenus » de la plateforme : la grande diversité linguistique mondiale, et même au sein des nations, rend cette perspective impraticable, à moins de condamner certains espaces linguistiques à une absence totale de modération.

La nécessité de s'attaquer au modèle actuel de Facebook

Il faut obliger Facebook à **se regarder dans le miroir**.

Pour l'instant, l'entreprise encourage une positivité naïve, et cherche à décourager ceux qui pointent les problèmes. Nous devons imposer à l'entreprise une structure qui l'oblige à **reconnaître ses défauts, et à formuler publiquement des stratégies visant à y remédier**. Pour chaque problème identifié, il faut pouvoir la tenir responsable des actions entreprises, ce qui nécessite un **accès à une partie des données internes**.

Pour cela, il serait stratégique de réunir et de préciser les demandes de données des chercheurs sous la forme d'une liste commune. La crédibilité dans les négociations sera renforcée par le caractère concret et collectif de la demande.

Cette responsabilisation doit avoir pour but de faire en sorte que l'incitation première de Facebook ne soit pas la satisfaction des actionnaires, mais la sécurisation de leurs produits et de sa communauté. De nombreuses solutions au problème des contenus « nuisibles », qui pourraient potentiellement fonctionner n'importe où dans le monde, ne sont pas implémentées par Facebook car elles réduiraient largement ses marges de profits.

C'est le modèle économique de Facebook qui constitue le fond du problème : il encourage à maximiser le temps passé par les utilisateurs sur la plateforme, et leur engagement avec le contenu. On sait maintenant que cela favorise la propagation des contenus les plus extrêmes et nuisibles. En 2018, Zuckerberg avait lui-même reconnu que les algorithmes de classement fondés sur l'engagement pouvaient être dangereux.

L'introduction de la **modification MSI** (Meaningful Social Interactions) devait permettre de corriger la survalorisation du contenu le plus viral, en introduisant un critère de « santé du réseau » mesurant la probabilité qu'un contenu soit aimé ou repartagé. Or, cette modification qui devait favoriser les interactions avec le cercle social restreint a eu pour effet pervers de **mettre en avant les contenus les plus extrêmes**.

Selon Frances Haugen, cela s'explique par le fait que les gens sont généralement peu sensibles aux repartages de contenu, sauf s'il vient de leurs 5 amis les plus proches, ou si le contenu est suffisamment extrême pour attirer leur attention.

Même une fois cet effet pervers établi, Zuckerberg a refusé de désactiver MSI dans des pays à risque de violence politique, faisant par-là preuve d'un « manque certain de leadership ».

Il faut bien comprendre que ce ne sont pas simplement les algorithmes qui causent un danger, mais leur **fonctionnement en concert avec des choix de produits**. Il faut développer un écosystème de sécurité, en partie indépendant de l'entreprise, pour analyser les données et prévoir les effets pervers éventuels de l'implémentation de nouveaux produits.

En particulier, il est vital que la structure collective d'évaluation des risques dispose d'un **cadre de travail flexible et dynamique**. Sans cela, il faudra attendre de nombreuses années entre l'implémentation d'un nouveau produit et la prise en compte réglementaire de ses effets pervers. Il faut pouvoir anticiper et s'adapter rapidement aux évolutions technologiques (Metaverse...)

La situation interne à Facebook quant à la question de sa nuisance sociale

De nombreux employés ont soutenu l'initiative de Frances Haugen. A son avis, de nouveaux lanceurs d'alertes sont appelés à émerger dans les prochaines années. Cependant, trop peu de personnes travaillent sur les problèmes de sécurité du réseau social, et leur action est immobilisée et découragée par l'impératif de profit.

La question se pose également de savoir si Facebook dispose de métriques internes capables de mesurer les éventuels dommages infligés. Il faudrait que l'entreprise soit obligée de rendre publique ses systèmes de détection des dommages, pour qu'il soit possible de lui reprocher de ne pas porter attention à tel ou tel problème.

Pour l'instant, **la qualité de ces métriques laisse à désirer** :

- Concernant les infox, il semble que chez Facebook, aucun contenu ne soit considéré « faux » avant d'avoir été « fact-checké » par un organisme indépendant. Dans certains contextes, notamment les pays où il existe peu ou pas de « fact-checking », ou les cas d'infox non-virales, cela rend la collecte et le traitement des données impossible.
- Facebook est au courant que certains publics sont particulièrement vulnérables et exposés aux contenus « nuisibles ». C'est notamment le cas des veuves, et plus généralement des populations en situation d'isolement social. Aucune métrique permettant de comptabiliser le nombre de personnes surexposées à ces contenus n'a pourtant été développée.

Les enfants sont particulièrement victimes des effets pervers de l'utilisation de Facebook

- Les enfants décrivent leur usage d'Instagram comme les rendant **malheureux et addictes**. Ils sont également conscients d'une forme de pression sociale qui rend l'arrêt de l'utilisation impossible, car elle est sanctionnée par un **ostracisme social**. De plus, leurs parents ne comprennent pas leur situation et tendent à les rendre entièrement responsables de leur consommation.

- De nombreux adolescents utilisent le « scrolling » comme moyen d'atténuer la dépression, la tristesse : il s'agit là d'un cercle vicieux si le contenu regardé est à l'origine, ou renforce ces sentiments.
- Le problème du « bullying » est, selon Frances Haugen, probablement sous-estimé.
- Il y a un risque important que le Metaverse soit destructeur pour l'image de soi des adolescents.

Quelques pistes de réflexion

Facebook a évolué dans les deux dernières années vers un fonctionnement par « groupes de partage ». Ces espaces sont très vulnérables à la diffusion de contenus « nuisibles ». Frances Haugen suggère que les groupes de plus de 1000 membres devraient être forcés de recruter leurs propres modérateurs humains. Cela aurait un double avantage :

- La modération ralentirait le flux d'infos, et pourrait contrebalancer la prime aux contenus extrêmes.
- La nécessité de recruter des modérateurs ferait friction : ce serait plus simple pour les groupes où se produit une véritable discussion que dans ceux qui ne font que faire circuler des liens.

Frances Haugen souligne que le problème n'est pas l'existence de mauvaises personnes ou de mauvais contenus, mais d'un système qui les amplifie systématiquement dans une optique de rentabilité économique.