

La justice française ordonne à Twitter de détailler ses moyens de lutte contre la haine en ligne

Le réseau social devra communiquer sous deux mois des documents détaillant le fonctionnement de son service de modération.

Par **Chloé Woitier** et AFP agence

Publié il y a 2 heures,

Mis à jour il y a 11 minutes



Twitter a été attaqué en mai 2020 par six associations déplorant l'opacité de ses services de modération. *Mike Blake / REUTERS*

La justice française a ordonné mardi à la société Twitter de communiquer à des associations de lutte contre les discriminations, dans un délai de deux mois, des documents détaillant ses moyens de lutte contre la haine en ligne.

Pour les associations plaignantes (L'UEJF, SOS Racisme, la Licra, J'accuse, SOS Homophobie et le Mrap), « *Twitter va devoir enfin se responsabiliser, arrêter de louvoyer et penser éthique plutôt que profit et expansion internationale* ».

Ces six associations avaient assigné le réseau social à Paris en mai 2020, estimant que l'entreprise manquait de façon « *ancienne et persistante* » à ses obligations de modération. Elles ont demandé en référé qu'une expertise soit ordonnée dans le but d'engager, par la suite, un procès sur le fond.

Dévoiler les coulisses de la modération pour la France

Dans sa décision, le tribunal a ordonné à Twitter International de transmettre *«tout document administratif, contractuel, technique ou commercial relatif aux moyens matériels et humains mis en oeuvre dans le cadre du service Twitter pour lutter contre la diffusion des infractions d'apologie de crimes contre l'humanité, l'incitation à la haine raciale, à la haine à l'égard de personnes à raison de leur sexe, de leur orientation ou identité sexuelle, l'incitation à la violence, notamment l'incitation aux violences sexuelles et sexistes, ainsi que des atteintes à la dignité humaine»*.

La société devra aussi, ajoute le tribunal, détailler *«le nombre, la localisation, la nationalité, la langue des personnes affectées au traitement des signalements provenant des utilisateurs de la plate-forme française»*, mais aussi *«le nombre de signalements»* issus de France en matière d'apologie des crimes contre l'humanité et d'incitation à la haine raciale, et *«les critères et le nombre des retraits subséquents»*.

Le juge a par ailleurs ordonné à Twitter de communiquer *«le nombre d'informations transmises aux autorités publiques compétentes, en particulier au parquet»*.

Boîte noire

Les services de modération sont la boîte noire des réseaux sociaux. Si ces derniers communiquent volontiers sur le nombre de modérateurs dédiés au marché européen, les plateformes refusent de détailler ces chiffres pays par pays. Les régulateurs et les élus se plaignent de ne pas pouvoir savoir combien de modérateurs sont assignés au marché français, si le français est bien leur langue maternelle, et quelles sont leurs conditions de formation et de travail.

Autre défi : la coopération des plateformes dans les enquêtes de police et de justice. Si certains réseaux sont bons élèves, d'autres sont pointés du doigt pour la lenteur avec laquelle ils réagissent aux réquisitions judiciaires. Certains ne donnent pas suite à ces demandes d'information. Au premier semestre 2020, Twitter n'a transmis des informations que pour 71% des réquisitions reçues de la part des autorités françaises. Cette situation a conduit trois victimes françaises de cyberharcèlement à attaquer le réseau social en justice. Ce dernier n'avait pas répondu aux réquisitions pour identifier les harceleurs, ce qui avait entraîné le classement sans suite des plaintes des victimes.

«Face à la propagation de la haine sur les réseaux sociaux, la responsabilité est double. Celles des individus qui se pensent anonymes et celle des GAFAs qui pensent que leur système juridique fait force de loi. Par sa décision, la justice française démontre que les GAFAs ne peuvent imposer leur propre loi. Twitter ne pourra plus impunément laisser la haine se déverser sur sa plateforme», commente l'Union des étudiants juifs de France dans un communiqué.

Dans leur plainte, les associations avaient expliqué avoir signalé au réseau social 1.110 tweets haineux, principalement des insultes homophobes, racistes ou antisémites sans équivoque, et constaté que seuls 12% d'entre eux avaient été supprimés dans «*une période raisonnable allant de 3 à 5 jours*». «*Ces résultats sont intolérables. (...) Ce que montre ce 'testing', c'est une inaction massive de la part d'une plateforme qui refuse manifestement de mettre les moyens humains nécessaires à la modération des contenus que son activité génère*», avait déclaré en mai 2020 le président de SOS Racisme Dominique Sopo.