Communiqué de presse de la FRA
Vienne, le 29 novembre 2023

**EMBARGO : 29 novembre 2023 à 06:00 CET**

# Haine en ligne : nous devons améliorer la modération des contenus afin de lutter efficacement contre les discours de haine

**Les commentaires injurieux, le harcèlement et l'incitation à la violence passent facilement à travers les mailles des outils de modération des contenus des plateformes en ligne, selon un nouveau rapport de l'Agence des droits fondamentaux de l'Union européenne (FRA). Celui-ci montre que la plupart des discours de haine en ligne ciblent les femmes, mais que les personnes d'ascendance africaine, les Roms et les Juifs sont également touchés. Le manque d'accès aux données des plateformes et la mauvaise compréhension de ce qui constitue un discours de haine entravent les efforts visant à lutter contre la haine en ligne. La FRA appelle à davantage de transparence et d'orientations afin de garantir à tous un espace en ligne plus sûr.**

Le rapport de la FRA sur la modération des contenus en ligne examine les difficultés liées à la détection et à la suppression des discours de haine dans les médias sociaux.

Il souligne qu'il n'existe pas de définition communément admise du discours de haine en ligne. Les systèmes de modération des contenus en ligne ne sont pas non plus ouverts au contrôle des chercheurs. Il est donc difficile de se faire une idée précise de l'ampleur du discours de haine en ligne, ce qui entrave les efforts déployés pour le combattre.

L'analyse par la FRA des messages et des commentaires publiés sur les plateformes de médias sociaux entre janvier et juin 2022 indique ce qui suit :

- **Haine en ligne généralisée :** sur les 1 500 messages déjà évalués par les outils de modération de contenu, plus de la moitié (53 %) sont malgré tout considérés comme haineux par les codeurs humains.
- **Misogynie :** les femmes sont les principales cibles du discours de haine en ligne sur toutes les plateformes et dans tous les pays étudiés. La plupart des discours de haine à l'égard des femmes comprennent des propos injurieux, du harcèlement et une incitation à la violence sexuelle.
- **Stéréotypes négatifs :** les personnes d'ascendance africaine, les Roms et les Juifs sont le plus souvent la cible de stéréotypes négatifs.
- **Harcèlement :** près de la moitié (47 %) des messages haineux constituent du harcèlement direct.

Pour lutter contre la haine en ligne, l'UE et les plateformes en ligne devraient :

- **Fournir un espace en ligne plus sûr pour tous :** afin de prévenir la haine en ligne, les plateformes devraient accorder une attention particulière aux caractéristiques protégées telles que le genre et l'origine ethnique dans leurs

efforts de modération et de suivi des contenus. Les très grandes plateformes en ligne, telles que X (anciennement Twitter) ou YouTube, devraient inclure la misogynie dans leurs mesures d'évaluation et d'atténuation des risques au titre du règlement sur les services numériques (DSA). Tous les États membres de l'UE devraient également ratifier la Convention d'Istanbul afin de mieux protéger les femmes en ligne.

- **Fournir davantage d'orientations :** il n'est pas toujours facile de distinguer ce qui est considéré comme un discours de haine et ce qui est protégé au titre de la liberté d'expression. Les régulateurs nationaux et européens devraient fournir davantage d'orientations sur l'identification de la haine en ligne illégale.

- **Saisir toutes les formes de haine en ligne :** afin de veiller à ce que les différents types de haine en ligne soient détectés, la Commission européenne et les gouvernements nationaux devraient créer et financer un réseau de signaleurs de confiance, en associant la société civile. La police, les modérateurs de contenu et les signaleurs devraient être correctement formés, afin de veiller à ce que les plateformes ne passent pas à côté de contenus ou ne suppriment pas trop de contenus.
- **Tester la technologie pour y déceler les préjugés :** les fournisseurs et les utilisateurs d'outils automatisés de modération de contenu devraient tester leur technologie pour y déceler les préjugés afin de protéger les personnes de la discrimination, comme l'a également souligné le précédent rapport de la FRA sur les préjugés dans l'IA.
- **Garantir l'accès aux données pour la recherche indépendante :** la Commission européenne devrait veiller à ce que les évaluations des risques réalisées par les plateformes elles-mêmes dans le cadre de la législation sur les services numériques soient complétées par des recherches indépendantes. Seuls des tests et approches diversifiés permettront de dresser un tableau complet des types de haine qui ne sont pas correctement identifiés et supprimés, ainsi que de l'incidence sur les droits fondamentaux des personnes.

Le rapport couvre quatre plateformes de médias sociaux (Reddit, Telegram, X, et YouTube) en Bulgarie, en Allemagne, en Italie et en Suède. La FRA n'a pas été en mesure d'accéder aux données de Facebook et d'Instagram pour cette recherche.

Entre janvier et juin 2022, la FRA a recueilli près de 350 000 messages et commentaires sur la base de mots clés spécifiques. Des codeurs humains ont évalué de manière aléatoire environ 400 messages de chaque pays pour déterminer s'ils étaient haineux. 40 postes sélectionnés de manière aléatoire ont ensuite été évalués plus en détail par des codeurs et des experts juridiques. Le rapport montre les différents types de discours haineux observés dans les pays, parmi les groupes cibles et sur les plateformes concernés.

**Selon Michael O'Flaherty, directeur de la FRA :**

*« L'ampleur du volume de contenus haineux que nous avons identifiés sur les médias sociaux montre clairement que l'UE, ses États membres et les plateformes en ligne peuvent intensifier leurs efforts pour créer un espace en ligne plus sûr pour tous, dans le respect des droits humains, y compris de la liberté d'expression.*

*Il est inacceptable d'attaquer des personnes en ligne uniquement en raison de leur genre, de leur couleur de peau ou de leur religion. »*

Pour plus d'informations, veuillez contacter : media@fra.europa.eu / Tél. : +43 1 580 30 653