

Unil

UNIL | Université de Lausanne
IDHEAP
Institut de hautes études
en administration publique
Droit public



**NORMES PRIVÉES ET DÉCISIONS ALGORITHMIQUES
LA MODÉRATION DES CONTENUS PAR LES RÉSEAUX SOCIAUX ET SES ENJEUX POUR LA DÉMOCRATIE**

**CYCLE DE CONFÉRENCES SUR 'DROITS HUMAINS ET RÉSEAUX SOCIAUX' ORGANISÉ PAR LA SECTION SUISSE DE LA
COMMISSION INTERNATIONALE DES JURISTES**

UNIVERSITÉ DE LAUSANNE – 2 MARS 2023

**PROFESSEURE SOPHIE WEERTS
MATHIEU FASEL, ASSISTANT-DIPLÔMÉ**

Introduction

- Remarques liminaires

- Intervention centrée sur la modération des contenus par les réseaux sociaux, une des activités des plateformes des réseaux sociaux et non sur les effets de la mise en ligne de contenu sur la démocratie.
- La notion de démocratie est comprise comme intimement liée à la notion d'Etat de droit, pour lequel le principe de la légalité est central.

- Structure de l'intervention

- Contexte
- Gouvernance de la modération des contenus
- Enjeux
- Perspectives de solutions

Contexte

- La modération des contenus par les entreprises de réseaux sociaux est devenue un enjeu de société dont se sont saisi diverses entités civiles et étatiques pour encourager ces entreprises à développer une activité plus conforme à l'état de droit et aux droits humains.
- Parmi les plus marquantes, nous pouvons citer :
 - En 2018, une coalition d'ONG publie les principes de Santa Clara, lesquels ont été complétés d'une seconde version en 2021.
 - Organisations internationales
 - UE : dès 2016, l'UE produit un code de conduite en matière de désinformation auquel ont adhéré les plus grandes entreprises de réseaux sociaux ; en 2022, l'UE adopte le Digital Services Act, législation contraignante fixant toute une série d'obligations aux plateformes de réseaux sociaux
 - Le Conseil de l'Europe publie en 2021 un guide sur la modération des contenus.
 - Les Nations Unies mettent en place le Forum sur la Gouvernance d'Internet. L'organisation appelle aux respecter principes en matière de Business & Human Rights

Conditions générales

Art. 14 DSA

1. Les fournisseurs de services intermédiaires incluent dans leurs conditions générales des renseignements relatifs aux éventuelles restrictions qu'ils imposent en ce qui concerne l'utilisation de leur service vis-à-vis des informations fournies par les destinataires du service. Ces renseignements comprennent des informations sur les politiques, procédures, mesures et outils utilisés à des fins de modération des contenus, y compris la prise de décision fondée sur des algorithmes et le réexamen par un être humain, ainsi que sur le règlement intérieur de leur système interne de traitement des réclamations. Ils sont énoncés dans un langage clair, simple, intelligible, aisément abordable et dépourvu d'ambiguïté, et sont mis à la disposition du public dans un format facilement accessible et lisible par une machine.
2. Les fournisseurs de services intermédiaires informent les destinataires du service de toute modification importante des conditions générales.
3. Lorsqu'un service intermédiaire s'adresse principalement à des mineurs ou est utilisé de manière prédominante par des mineurs, le fournisseur de ce service intermédiaire explique les conditions et les éventuelles restrictions relatives à l'utilisation du service d'une manière compréhensible pour les mineurs.
4. Lorsqu'ils appliquent et font respecter les restrictions visées au paragraphe 1, les fournisseurs de services intermédiaires agissent de manière diligente, objective et proportionnée en tenant dûment compte des droits et des intérêts légitimes de toutes les parties impliquées, et notamment des droits fondamentaux des destinataires du service, tels que la liberté d'expression, la liberté et le pluralisme des médias et d'autres libertés et droits fondamentaux tels qu'ils sont consacrés dans la Charte.
5. Les fournisseurs de très grandes plateformes en ligne et de très grands moteurs de recherche en ligne fournissent aux destinataires des services un résumé des conditions générales, y compris des mécanismes de recours et de réparation disponibles, concis, facilement accessible et lisible par une machine, dans un langage clair et dépourvu d'ambiguïté.
6. Les très grandes plateformes en ligne et les très grands moteurs de recherche en ligne au sens de l'article 33 publient leurs conditions générales dans les langues officielles de tous les États membres dans lesquels ils proposent leurs services.

Gouvernance de la modération des contenus

1. Les algorithmes

- La modération des contenus se fait par le biais de processus automatisés en raison de la gigantesque quantité de contenus publiés sur les réseaux sociaux chaque jour
- Ces algorithmes sont principalement de deux types :
 - des modèles de correspondance (*matching models*), qui contrôlent que les contenus publiés ne correspondent pas en tout point à un autre contenu interdit ou protégé.
 - des modèles de prédiction (*predictive models*), qui vont chercher à voir si un contenu posté en ligne est similaire à une classe d'autres textes, images ou vidéos qui sont interdits. Si le contenu atteint un seuil de similarité suffisant, alors il est enlevé.

Gouvernance de la modération des contenus

2. Les modérateurs humains

- Les décisions de modération de contenus peuvent également être prises par des modérateurs humains, en «back-up» des algorithmes
- Cette activité est souvent externalisée et s'effectue dans des conditions de travail difficiles

Gouvernance de la modération des contenus

3. Les règles

- Les règles qui guident les décisions de modération sont regroupées sous ce que les entreprises de réseaux sociaux appellent «Community Guidelines» ou «Community Standards»
- Ces règles sont utilisées pour justifier les décisions de modération auprès des utilisatrices et utilisateurs des réseaux sociaux



Your Post Goes Against Our Community Guidelines

We removed your post because it goes against our **Community Guidelines** on hate speech or symbols. If you post something that goes against our guidelines again, your account may be deleted, including your posts, archive, messages and followers.

 Post removed for hate speech or symbols
Posted April 1 at 1:03 PM 

Hate speech guidelines

When hate speech is being shared to challenge it or to raise awareness, we may allow it. In those instances, we ask that you express your intent

Next



Your Account May Be Deleted

Some of your previous posts didn't follow our **Community Guidelines**. If you post something that goes against our guidelines again, your account may be deleted, including your posts, archive, messages and followers.

-  Post removed for hate speech or symbols
Posted April 1 at 1:03 PM 
-  Comment removed for bullying or harassment
Posted April 1 at 10:03 AM
-  Story removed for nudity or sexual activity
Posted March 10 at 8:05 PM 

OK

Enjeux

- La question de la qualité rédactionnelle des Community Standards de Facebook se pose avec d'autant plus d'acuité que ceux-ci jouent le rôle de base légale restreignant la liberté d'expression des individus.
- La sécurité juridique est un concept essentiel en droit, notamment en droit international des droits de l'homme. Elle est aussi un enjeu pour les utilisateurs et utilisatrices des réseaux sociaux.
- Il y a une tension majeure entre, d'un côté, l'exigence de sécurité juridique sur les réseaux sociaux qui est en train d'émerger (notamment par les art. 14 et 16 du DSA de l'UE) et, de l'autre, le recours massif à des processus de prise de décision automatisés.

Perspectives de solution

- Le statu quo: ne pas prendre de mesure, et se fier à la bonne volonté des entreprises de réseaux sociaux de prendre des mesures pour lutter contre la désinformation et les discours de haine → solution en vigueur qui ne protège manifestement pas les systèmes démocratiques
- La plus radicale ('tirer la prise') : on estime que les réseaux sociaux ne peuvent pas être utilisés de façon adéquate avec la préservation de nos démocraties → solution improbable étant donné l'imbrication des entreprises de plateformes de réseaux sociaux dans l'économie
- Réglementer : pour la Suisse, l'adoption par l'Union Européenne du Digital Services Act est particulièrement intéressante. Le règlement européen propose des solutions novatrices, tant du point de vue des exigences de transparence que par les nouveaux instruments adoptés (audits, risk assessments, etc.)