Comment la plus grande communauté du monde reste saine ?

De nos jours, internet et les réseaux sociaux sont aux centres des utilisations et des discussions. En ne vivant qu'à travers ces réseaux, une utopie se créait, en pensant qu'il y a moins de négatifs que l'on ne pense. Mais une grande partie des utilisateurs oublie que le bon fonctionnement des forums, sites, blogs ou réseaux sociaux s'opèrent grâce à de la modération pour maintenir un équilibre entre toutes les communautés. Cette modération est un point essentiel et particulier pour les réseaux sociaux, qui ont un statut « publique », fonctionnant comme une communauté (ne défendant par l'image d'une marque par exemple). Pour revenir sur la modération, elle est invisible pour tous ces utilisateurs.

On parle bien de modération de contenu, pour rafraîchir la mémoire de certains, en voici une définition. C'est simplement le fait de contrôler et filtrer les contenus que peuvent poster les utilisateurs sur les différentes applications proposant le partage d'informations (réseaux sociaux, forum, presse digitale, site, blog, etc.), à travers des publications ou des commentaires (sur toutes types de publications ou commentaires, textuels, audios, photographies ou vidéos).

« Modération : évitez les extrêmes, pardonnez les injures, autant que vous pensez qu'elles méritent le pardon... »

Benjamin Franklin

A savoir, on retrouve des chiffres délirants sur le réseau social [Instagram](#) d'un point de vue mondial, avec en 2020, **1 milliards d'utilisateurs actifs mensuels, 500 millions de comptes actifs chaque jour, 53 minutes en moyenne par utilisateurs tous les jours et une grande diversité avec l'âge des utilisateurs.**

Une question de liberté d'expression se pose !

En ce qui concerne les réseaux sociaux ou internet en général, on entend souvent dire que beaucoup de gens se sentent pousser des ailes derrière leur écran, et ne réfléchissent pas aux poids des mots qu'ils utilisent pour juger, commenter la personne ou son travail. Le maintien de l'ordre a un rapport direct avec la liberté d'expression qui est souvent lié à la liberté d'opinion et de la presse, qui a vu sa naissance dans la déclaration des droits de l'homme et du citoyen en 1789.

« Tout citoyen peut parler, écrire, imprimer librement. »

Mais que ce soit dans la vie réelle ou sur internet, pour garantir la paix sociale, l'état a mis en place des bornes à cette liberté d'expression. Voici les catégories principales des limites de la liberté d'expression.

1. Diffamation
2. Injure
3. Provocation à la haine
4. Apologie du terrorisme

Ces bornes permettent de cibler des axes précis, pour une modération facilitée et claire. Un document de référence se nommant « **Charte de modération** » qui se trouve indispensable (pas obligatoire) aujourd'hui lors de la création d'un site ou d'une page communautaire. Son utilité principale est l'affichage aux utilisateurs et modérateurs des règles à respecter concernant les différentes interactions, ce qui amène vers une protection légale lorsque du contenu est supprimé.

Pour ce qui est de sa composition, il se doit d'être court pour que tous utilisateurs aient la foi de le lire. Puis on y retrouve essentiellement les règles de bonnes conduites, ainsi que les risques encourus en cas de non-respect.

Mais comment on modère ?

Avant de parler des processus de modération, il est nécessaire de donner un exemple en replongeant dans les chiffres *d'Instagram*, pour mettre en évidence le volume démentiel de contenu publié, qu'il faut modérer. **100 millions de photos et de vidéos publiées quotidiennement** et **285 commentaires en moyenne par publications**.

Parlons stratégie maintenant, avec le choix de la temporalité de modération des contenus, soit « **a priori** » qui signifie avant la mise en ligne de la publication avec une vérification, on peut retrouver ça hors des réseaux sociaux, avec par exemple la mise en ligne d'annonce sur [Leboncoin](#), ou « **a posteriori** », qui logiquement vérifie après la mise en ligne, qui revient à la forme habituelle utilisée par les réseaux sociaux.

En vient la stratégie fondamentale, avec la décision d'une modération « **Automatique** » ou « **Manuelle** ».

Débutons par la modération « Manuelle », qui comme son nom peut le faire sous-entendre, est menée par des « modérateurs humains » ou également appelé « Éboueurs du Web », avec l'aide ou non des signalements des utilisateurs. Les entreprises peuvent prendre la décision « **d'internaliser** » cette modération, qui est utilisée pour des volumes peu important, ou de « **l'externaliser** » qui au contraire adoptée pour des volumes importants, ce qui consiste à se tourner vers un prestataire spécialisé ([Competence Call Center](#), [CPL Ressources](#), [Task Us](#), [Micro Sourcing](#), etc.), cette dernière est privilégiée par les grosses entreprises avec notamment [Facebook](#), Instagram, [YouTube](#) qui vont chercher une certaine protection judiciaire en cas de problème lors de la modération sur leurs plateformes.

A propos de la modération manuelle, beaucoup de scandale éclatent, les modérateurs sortent du silence, en revendiquant un ras le bol des conditions de travail qu'ils peuvent vivre, créant chez beaucoup de la dépression ou des tendances suicidaires (Contrat stipulant l'interdiction de parler de leur travail, interdiction de communiquer avec ses collègues, temps de pauses chronométrées, un salaire de misère, aucun soutien psychologique avec la visionnage de négatifs et d'horreurs quotidiens, etc.).

Dans l'enfer des modérateurs Facebook - Invisibles #3 - YouTube

Quant à la modération automatique, elle se concentre seulement sur les commentaires des publications. Cette fonctionnalité est mise à disposition aux auteurs des publications, il s'agit d'un filtrage automatique des commentaires, en précisant des mots insultants personnalisés.

A ce stade, les humains sont privilégiés, car contrairement aux robots ils sont capables de comprendre le contexte, en prenant en compte les subtilités des propos de chacun. Mais peut être qu'un changement arrivent...

Le changement c'est maintenant...enfin bientôt

L'évolution des technologies, commence à rentrer de plus en plus dans notre quotidien, c'est ce que nous pouvons apercevoir avec **l'Intelligence Artificielle**. Dans le cas de la modération automatique

des commentaires textuels on aperçoit des applications qui émergent, elles vont permettre de rivaliser avec l'humain, sur l'analyse du contexte, en anticipant les fautes de frappe ou d'orthographe, ainsi que les smileys, tout en adaptant son jugement par rapport au profil de l'auteur des contenus.

On retrouve deux applications phares, « [Smart Moderation](#) », qui permet de modérer du contenu anglais intelligemment, actuellement en version bêta et proposant aux utilisateurs de l'implémenter sur YouTube, Instagram ou Facebook. Et enfin, le meilleur pour la fin...une application Française, COCORICO ! Sous le nom de « [BodyGuard](#) », elle a été créée par Charles Cohen, développeur de 22 ans en 2018, cette application propose la modération intelligente de contenu en temps réel en français cette fois-ci, avec la possibilité de l'ajouter sur YouTube, [Twitter](#), Instagram ou la plateforme de streaming [Twitch](#). Ces applications se composent de quelques modérateurs humains pour prendre la décision sur des erreurs que l'IA peut effectuer. Au sujet du modèle économique, t à l'application anglophone se voit payante, contrairement à Bodyguard qui est encore gratuit.

Petit focus sur cette application, depuis 2020, elle **comptabilise 50 000 utilisateurs**, analyse chaque **mois 20 millions de commentaires et c'est 2 % d'erreurs**. Pour en arriver à un perfectionnement d'une IA, il faut utiliser de la donnée, pour lui apprendre les commentaires qui sont haineux ou non. Concernant le processus d'analyse suivi par l'IA de Bodyguard, il effectue une analyse sur le vocabulaire en recherchant les mots haineux, ou insulte, en prenant en compte tous les types d'orthographe ou émoji, puis vient l'analyse du contexte, pour enfin prendre une décision sur le commentaire s'il est haineux ou non.

PRÉSENTATION DE BODYGUARD - YouTube

Conclusion

Cette communauté sur internet cesse de grandir d'année en année, ce qui augmente par la même occasion le volume des informations, demandant un travail encore plus intense pour la modération.

Pour supporter cette évolution et continuer de garder un savoir vivre et l'innocence des plus fragiles sur les réseaux, la modération humaine ne peut être la solution miracle, mais qu'elle solution choisir ?

La modération entièrement délivrée aux IA ? Un hybride entre humain et IA ?