

L'aventure de la modération - épisode 3

Maiwann, membre de l'association, a publié sur son blog une série de cinq articles sur la modération. Nous les reproduisons ici pour leur donner (encore) plus de visibilité.

Voici le troisième.

Les cas d'étude c'est bien sympa, mais cela ne donne qu'une vision partielle du fonctionnement. Aussi je vais détailler ma montée en compétence sur la modération, et celle du collectif parce que ça fait partie des solutions à mettre en place pour que se soit un boulot le moins délétère possible !

Les outils accessibles via Mastodon

Mastodon permet, pour les utilisatriceuses inscrites chez nous de :

- flouter toutes les images de ce compte (et celles qu'il publiera ensuite) (pratique pour les instances qui hébergent du porno)
- supprimer un message
- limiter la visibilité du compte à ses abonnés seulement
- suspendre le compte sur un temps restreint
- supprimer le compte définitivement
- contacter l'utilisatriceuse par message privé
- contacter l'utilisatriceuse par mail
- bloquer un Mastodon entier (!!) (aussi appelé instance).

Pour les utilisatriceuses qui ne sont pas chez nous, il est seulement possible de :

- flouter toutes les images de ce compte (et celles qu'il publiera ensuite) (pratique pour les instances qui hébergent du porno)
- limiter la visibilité du compte à ses abonnés
- bloquer le compte

— contacter l'utilisateurice par message privé.

Lorsqu'une personne effectue un signalement, nous recevons une notification par mail si :

- la personne faisant le signalement se trouve chez nous
- la personne signalée est chez nous ET que la personne ayant fait le signalement a demandé explicitement à ce que nous le recevions.

Les signalements les plus importants à gérer étant évidemment ceux concernant des personnes se trouvant sur notre (instance de) Mastodon et ayant un comportement problématique. Actuellement, la majorité des signalements pour lesquels nous sommes notifiés sont faits par une personne inscrite chez nous à propos de contenus qui se trouvent ailleurs, sans doute pour prévenir les modératrices de l'autre Mastodon qu'il y a quelque chose à regarder de leur côté.

Au fur et à mesure du temps, j'ai l'impression que si les signalements ne sont pas liés à du contenu chez nous, et ne sont pas explicitement problématiques, on peut laisser la responsabilité d'agir aux modératrices des autres instances.

L'important étant de bien s'occuper de ce qui se trouve chez nous. □



My Neighbor Mastodon - CC-BY David Revoy

Comment ça se passe ?

Donc une fois la notification par mail reçue, on se connecte à un compte avec les droits de modération (pour certains leur compte personnel, pour d'autres le compte de modération que nous avons créé à la sortie de la charte pour avoir un compte dédié à ce sujet). Nous avons accès à une interface récapitulant le compte concerné, les contenus signalés, le message du signalement expliquant pourquoi il a été effectué (extrêmement important car sinon on perd énormément de temps à comprendre ce qui est reproché) et quelques autres indicateurs (notamment le nombre de signalements déjà effectués, pratique pour repérer les récidivistes).

Démarre ensuite dans la grande majorité des cas un travail d'enquête afin de comprendre **le contexte**. C'est une action extrêmement chronophage mais qui nous permet de savoir si la personne a une façon de se comporter correcte, un peu pénible ou carrément problématique, ce qui va influencer le niveau de sévérité de l'action de modération.

Si nous avons de la chance, il suffit de parcourir une dizaine de contenus, mais parfois la personne a partagé des vidéos, et il est nécessaire de regarder l'entièreté de la vidéo pour se rendre compte de son caractère problématique... ou non. C'est comme cela que je me suis retrouvée à regarder vingt minutes d'une vidéo célébrant le Brexit où se trouvaient Asselineau, Nicolas Dupont-Aignan, Florian Philippot. Quel excellent moment.

C'est ce travail de contextualisation qui explique pourquoi il est extrêmement important, lorsque vous réalisez un signalement, de mettre le plus de détails possibles : quels sont les contenus incriminés ? Pourquoi pensez-vous que cela mérite une action de modération ? Trop souvent nous avons un signalement composé d'un seul pouet, hors contexte, qui nous oblige à nous y attarder bien plus que si on nous avait fourni des explications détaillées. (Merci pour nous !)

Une fois cette contextualisation réalisée, il est temps de se décider... est-ce qu'il y a quelque chose à faire ? Est-ce qu'on supprime les contenus incriminés ? Est-ce qu'on contacte la personne ?

À Framasoft nous avons un tchat dédié à la modération, dans lequel nous partageons notre analyse du contexte, le signalement, et nous proposons l'action ou les actions qui nous semblent les plus appropriées. L'idée est de pré-mâcher le travail aux copaines tout en n'ayant pas à porter seul le poids de la décision de la « bonne » action de modération à réaliser.

Donc une fois que j'ai une idée de ce qui me semblerait « le mieux » à faire (ou pas), je demande leur avis aux copaines sur ce tchat. Il y a ensuite un petit temps d'attente nécessaire, pas problématique dans la plupart des cas mais très stressant quand la situation revêt un caractère « d'urgence », et que quelques heures comptent (par exemple si quelqu'un commence à se prendre une vague de harcèlement, plus vite on peut la protéger mieux c'est).

Si on n'est pas d'accord, on discute, jusqu'à tomber d'accord ou ne pas tomber d'accord et utiliser la règle du « c'est qui qui fait qui décide ». Donc si on se dit « cette personne il faudrait la contacter pour lui expliquer » mais que personne n'a l'énergie de rédiger quelque chose... ben on fait différemment en choisissant une action qui tient compte du niveau d'énergie des modératrices.

Et enfin, on réalise l'action de modération « finale » qui, si elle est technique, est hyper rapide (nonobstant un petit message pour prévenir si on a envie d'en faire un), mais qui si elle est dans la « communication » n'est que le début du chemin !

J'espère que cela vous permet de davantage mesurer l'effort **invisible** qu'est la modération : Recevoir la notification, Identifier le niveau d'urgence, Contextualiser, Synthétiser, Trouver une décision juste, la Soumettre au collectif, Attendre les retours, Discuter, Agir, peut-être Répondre... 9 à 10 étapes qui ne sont vues de personne, excepté si les contenus ou le compte sont purement et simplement supprimés. Il est donc très facile de dire que « rien » n'est fait alors que ça s'active en coulisses.

De plus, dans un monde où tout et n'importe quoi est « urgent » autant vous dire que ça n'est pas facile de prendre le temps de décider sereinement d'une action la plus juste possible alors que précisément, cela demande du temps de ne pas réagir sur le coup de l'émotion !

L'anonymat des modératrices

Lorsque j'ai débuté la modération, j'avais dans l'idée qu'il était important que les modératrices soient clairement identifiées pour qu'il soit plus simple de savoir

d'où iels parlent. Non seulement pour savoir si iels sont directement concerné·es par certaines oppressions systémiques, mais aussi pour pouvoir avoir un œil sur leur niveau de déconstruction (ce n'est pas parce qu'on est une meuf et donc qu'on subit le sexisme que cela nous donne un badge « a déconstruit le patriarcat », loin de là).

J'en suis complètement revenue, ou plutôt j'ai bougé sur ma position : ce ne sont pas les individus qu'il faut connaître, mais la position d'un collectif, et savoir si on a envie de faire confiance à ce collectif... ou non !

Pourquoi ce changement ? Déjà parce que lors de la vague de harcèlement que nous avons subie à la publication de la charte, la violence était très intense alors que je n'étais pas directement citée : c'était le compte de l'association (@Framasoft@framapiaf.org) qui l'était, et je ne voyais que ce qui tombait dans mon fil ou ce que j'allais voir de moi-même (enfin en étant sous le coup de la sidération, moment où l'on a pas toujours les bonnes pratiques : maintenant je ferme l'ordinateur pour m'éloigner du flot).

Si j'étais frustrée de ne pas pouvoir mettre de visage « humain » sur les personnes qui avaient écrit cette charte, j'ai surtout été très protégée par mon anonymat du moment. Et j'ai pu choisir, en mes termes et dans un timing que j'ai choisi, de partager mon ressenti sur la situation (avec l'article intitulé Dramasoft du nom du hashtag utilisé pour l'occasion).

Et je me souviens que je savais très clairement que je prenais un grand risque en communiquant sur ma participation à cette charte (et j'ai fait attention à ne pas parler des actions de modération en tant que telle dans un espoir de limiter l'impact). Le fait d'en contrôler la totalité de la publication, quand, sur quel support, auprès de qui, quand je me suis sentie prête, m'a énormément protégée.

Je reviens d'ailleurs sur une partie de ce que j'ai énoncé au-dessus, l'importance de

savoir si iels [les modérateurices] sont directement concerné·es par certaines oppressions systémiques

Je pense à présent qu'il est très problématique de penser pouvoir regarder un·e individu pour se permettre de juger de si iel ferait un·e bon·ne modérateurice : ce sont les actions et la direction dans laquelle se dirige le collectif qui compte. Nous ne pouvons pas déterminer, de par quelques pouets ou contenus partagés, si une personne est « suffisamment » déconstruite ou non.

Pire, cela nous encourage à faire des raccourcis, qui peuvent être d'une violence inouïe. Aussi, il nous a par exemple été reproché de concevoir une charte pour des enjeux de modération dont nous n'avions pas conscience, notamment parce que nous n'avions pas de personnes concerné·es par les oppressions systémiques dans l'équipe (qui seraient composés de mecs cis, het, blancs...).

Déjà c'est faux, car il y a des femmes qui subissent le sexisme qui en font partie (dont moi par exemple !) mais surtout parce qu'il n'est pas possible de savoir quel est le contexte et l'identité de chacun·e des membres, et notamment si certain·es seraient par exemple dans le placard, cachant soigneusement leur identité ou orientation sexuelle.

Rendez-vous compte de la violence, pour quelqu'un qui n'a pas fait son ou ses *coming-out*, de recevoir de la part d'adelphes qui sont dans le même groupe oppressé qu'elle un message sous entendant qu'elles ne sont pas légitimes puisqu'elles ne sont pas *out*. La violence est infinie, et je crois que c'est ce qui m'a mise le plus en colère dans cette histoire, car je ne pouvais pas vérifier si des membres étaient touché·es par ces accusations (puisque n'ayant pas fait leur *coming-out*... vous m'avez comprise).

Aussi, à nouveau, je pense que la solution repose sur le collectif, pour pouvoir à la

fois s'adresser à quelqu'un tout en protégeant les individus.

À Framasoft, nous avons publié cette charte et ouvert un compte modération derrière lequel chaque membre peut lire les messages, et nous agissons pour des personnes qui nous font confiance. Plus récemment encore, nous avons re-partagé notre vision du monde dans notre manifeste, sans compter les nombreuses conférences et articles de blog qui expliquent comment nous nous positionnons. Si cela ne suffit pas aux utilisatrices, ils peuvent nous bloquer ou aller ailleurs, sans animosité de notre part : nous faisons avec nos petits moyens.

Le temps passé

Une question qui m'a été souvent posée c'est : Combien de temps vous y passez, à faire de la modération ?

Alors déjà il y a une grosse différence de temps passé entre le moment de mise en place de la charte de modération, et maintenant.

Lors de la mise en place de la charte et environ un an ensuite, nous étions sur quelques heures par semaine. Actuellement, c'est plutôt quelques heures par mois, et pour des cas moins complexes à gérer (même si on n'est pas à l'abri d'une vague de harcèlement !). Et c'est à mettre au regard du nombre : 1600 utilisateurs actifs chez nous, 11 000 inscrit·es (donc beaucoup de comptes en veille) et 6 millions d'utilisateurs au total sur tout le réseau.

Cette immense différence s'explique par deux raisons :

- Nous avons viré tous les relous.
- Nous avons fermé les inscriptions, et donc il n'y a plus de nouveaux.

Les utilisatrices qui restent sur framapiaf sont celles qui ne nous demandent pas trop de boulot, et donc on s'en sort bien !

Mais je pense que même sur une instance dont les inscriptions restent ouvertes, on doit gagner en tranquillité grâce aux blocages d'instances déjà effectués et à la montée en compétence de l'équipe de modération (si l'équilibre est entretenu pour que les modératrices restent, ça veut dire que le collectif en prend suffisamment soin !).

Note : Si les chiffres d'Olivier Tesquet sont bons (je n'ai pas vérifié), il y a en fait sur Mastodon une bien meilleure façon de modérer que sur Twitter parce que les modératrices seraient juste... plus nombreuses ?

Mastodon compte aujourd'hui 3615 instances, c'est à dire 3615 politiques de modération potentiellement différentes.

Au passage, ramené au nombre d'utilisateurs actifs, ça fait au moins un modérateur pour 250 mastonautes. Sur Twitter, c'est un pour... 200 000.

Et après ?

C'est l'heure de faire un petit bilan de cette triplète d'articles ! Ce sera dans le prochain épisode !