L’aventure de la modération – épisode 2

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 deuxième.

Maintenant que je vous ai fait une introduction de six pages, c’est le moment de passer au concret : Comment est-ce que se déroule la modération ?

Les cas simples : les fachos, les mascus, l’extrême-droite

De façon surprenante, j’ai découvert que tout ce qu’il y a de plus délétère sur les réseaux sociaux capitalistes, c’est à dire en résumé : les comptes d’extrême-droite, étaient très simples à modérer.

Nous en avons eu un exemple lors de l’été 2018. Twitter a réalisé une vague de fermeture de comptes de mascus provenant du forum tristement connu de jeuxvideo.com, surnommé le 18-25. Comme souvent, ces personnes crient à la censure (…) et cherchent un réseau social alternatif dont la promesse serait de promouvoir une liberté d’expression leur permettant de dire les atrocités qu’ils souhaitent.

My Neighbor Mastodon – CC-BY David Revoy

Ils débarquent donc sur Mastodon, avec un schéma qui a été le suivant :

1° Première inscription sur le premier Mastodon trouvé

Les 18-25 étant majoritairement francophones, ils se sont retrouvés souvent sur mamot.fr et parfois chez nous en s'inscrivant sur framapiaf.org, parfois sur mastodon.social qui est le Mastodon proposé par le développeur principal.

Nous avons donc pu commencer une première vague de bannissement en suivant les profils des nouveaux inscrits, et leurs premiers contenus qui concourraient au prix de « qui dit les choses les plus atroces ».

Nous avons fermé les inscriptions momentanément pour endiguer le flot, et peut-être avons nous à un moment viré massivement les nouveaux comptes.

« Quoi ?! » vous entends-je dire, « mais c'est en opposition totale avec le point 3

de la charte » !

> *Nous différencions personnes et comportements. Nous modérerons les comportements enfreignant notre code de conduite comme indiqué au point 5, mais nous ne bannirons pas les personnes sous prétexte de leurs opinions sur d'autres médias.*

Ma réponse est donc « vous avez raison » suivie d'un « et nous sommes humains donc nous faisons ce que nous pouvons, en commençant par nous protéger ».

Je ne me souviens pas vraiment de notre façon de modérer à ce moment là : avons-nous attendu que chacun poste quelque chose de problématique avant de le virer ? Ou avons-nous agi de façon plus expéditive ? Même si nous étions dans le second cas, nous étions sur un grand nombre de compte qui venaient de façon groupée pour revendiquer une « liberté d'expression » leur permettant de dire des horreurs… Il n'y a à mon avis pas besoin d'attendre que chacune de ces personnes indique quel est sa façon de penser alors qu'elles viennent massivement d'un endroit (le forum de jeuxvideo.com) qui porte une vision de la société délétère.

Donc peut-être qu'on n'a pas respecté strictement la charte sur ce moment là, et encore une fois, je préfère protéger les membres de l'association et les utilisateurices en faisant preuve d'un peu d'arbitraire, plutôt que de suivre une ligne rigide.

2° S'inscrire sur un autre Mastodon

Après leur premier bannissement, des informations sur le fonctionnement décentralisé de Mastodon ont circulé. Les nouveaux arrivants se sont donc inscrits sur une autre instance, et ont plus ou moins rapidement été bannis une nouvelle fois.

C’est alors que des personnes du fédiverse ont expliqué à ces personnes qui se plaignaient de la fameuse « censure » qu’ils pouvaient monter leur propre Mastodon, et y appliquer leurs propres règles.

3° Monter une instance Mastodon rien que pour les 18-25

Sitôt dit, sitôt fait, les nouveaux arrivants s’organisent, montent leur Mastodon (soit grâce à leurs compétences techniques, soit grâce à un service qui propose de le faire) et s’y inscrivent tous.

Victoire ! Leur instance Mastodon, leurs règles, plus personne ne les bannira !

Quand à nous, nous n’avions plus qu’à appuyer sur le bouton « Bloquer l’instance » pour nous couper de ce Mastodon dédié au 18-25. Problème réglé.

Alors je ne voudrais pas vous faire croire que cette cohue s’est faite sans effet problématique. Déjà parce que je ne connais pas ses effets sur tous les utilisateurices de Mastodon, mais aussi parce que les 18-25 ont eu le temps de poster des contenus très violents, le point culminant étant une photo de la salle du Bataclan le soir du 13 novembre, remplie de cadavres. Je n’ai pas vu cette image mais un ami de l’association oui, et je maudis notre manque d’expérience pour ne pas avoir pu empêcher qu’il soit confronté à cette image horrible.

Mais de façon générale, tous les discours de haine sont vraiment aisés à modérer. Juste : Ça dégage. Fin de la discussion.

Là où ça devient difficile, se sont sur les cas qui jouent avec les limites.

Cas limites : quand une personne est… pénible

Je pense que c'est ce type de cas qui nous a valu une réputation de « mauvaise modération » : certaines personnes qui ne posaient pas de problème légaux, n'étaient pas agressifs au sens strict du terme ou ne partageaient pas de contenu violent, mais qui étaient… relous.

Un peu de *mansplanning* par ci, une discussion sur un sujet d'actualité avec une position moralisante par là… Bref, tout ce qui fait plaisir à lire (non) sans réellement mériter de couperet de la part de la modération.

Nous sommes arrivés à un point où les salariés qui faisaient de la modération avaient masqué le compte depuis leur profil personnel, mais ne se sentaient pas légitimes à le virer de notre Mastodon puisque rien de suffisamment répréhensible n'avait été fait.

Maintenant que j'ai un peu de recul, je peux dire que nous avons deux postures possibles dans ce cas, chacune nécessitant un travail de contextualisation lié au compte : quelle description du compte ? Quels contenus partagés ? Dans quelle ambiance est la personne ? Ça n'est pas la même chose si la personne écrit dans sa biographie « Mec cis-het et je vous emmerde » que si il est écrit « Faisons attention les un·es aux autres ».

Bien sûr, rien n'est strictement « éliminatoire », mais chaque élément constitue un faisceau d'indices qui nous pousse nous-même à passer plus ou moins de temps à prendre soin de la personne, de façon équivalente au soin que cette personne semble mettre dans ces échanges sur le média social.

Si la personne semble globalement prendre soin, mais a été signalée, nous pouvons décider de rentrer en discussion avec elle. C'est une option intensément chronophage. On pourrait se dire que pour trois messages de 500 caractères, pas besoin d'y passer trop de temps… au contraire !

Autant que possible, chaque message est co-écrit, relu, co-validé avant d'être envoyé, afin de s'assurer que le plus de personnes possibles de l'équipe de modération soient en accord avec ce qui y est écrit, le ton donné… et cela dans un temps le plus réduit possible histoire que l'action de modération ne date pas de quinze jours après le contenu (c'est parfois ce qui arrive). Or les membres de l'équipe sont pour beaucoup bénévoles, ou sinon salarié·es avec un emploi du temps très chargé.

Il faut aussi pas mal de relecteurices pour éviter LE truc qui fait perdre trois fois plus de temps : se faire mal comprendre par la personne à laquelle on écrit. Et là… c'est le drame !

Une autre option, ressemblant à celle-là mais avec moins de précautions de notre part est d'utiliser notre position d'autorité « Nous sommes la Modération » pour signaler aux personnes que nous comptons sévir en cas de récidive.

Cependant, si la personne ne semble pas intéressée par le fait de prendre soin des autres personnes, nous avons une astuce pour ne pas retomber dans le cycle infernal du « est-ce qu'on devrait faire quelque chose ou non ? » : le « ça prend trop d'énergie à l'équipe de modération ».

> *Cependant, nous bannirons toute personne ayant mobilisé de façon répétée nos équipes de modération : nous voulons échanger avec les gens capables de comprendre et respecter notre charte de modération.*

En effet, c'est un indicateur qui nous permet de ne pas « boucler » trop… si l'on est plusieurs à se prendre la tête sur un signalement, alors que notre énergie serait mieux ailleurs… alors on décide souvent que cela suffit pour mériter un

petit message disant en gros « Bonjour, vous prenez trop de temps à l'équipe de modération, merci d'aller vous installer ailleurs ». Cela nous a sorti de maintes situations délicates et c'est une félicité sans cesse renouvelée que de pouvoir citer cette partie de la charte.

Cas limites : quand la revendication est légitime

Autre cas compliqué à gérer : les différents moments où les personnes ont une revendication dont le fond est légitime, mais où la forme fait crisser les dents.

Par exemple, revenons sur la réputation de Framapiaf d'avoir une mauvaise modération (ou une absence de modération ce qui revient au même).

Clarifions un instant les reproches qui me semblent étayées de celles pour lesquelles je n'ai pas eu de « preuve » :

— Il a été dit que nous ne faisions pas de modération, voire que nous ne lisions pas les signalements : c'est faux.

— Par contre, comme expliqué dans le cas au-dessus, nous n'étions pas outillés et expérimentés pour gérer des cas « limites » comme ceux au-dessus qui ont entraine une pénibilité dans les interactions avec de nombreuxses utilisateurices.

Nous avions donc bien un manque dans notre façon de faire de la modération. Si vous avez déjà les explications du « pourquoi », je tiens à dire que cela ne nous empêche pas de nous rendre compte que cela a été pénible pour d'autres utilisateurices.

Cependant, la façon dont le problème nous a été remonté l'a été de façon souvent

agressive, parfois violente car insinuant des choses à propos de nos membres (qui ne seraient que des personnes non concernées par les oppressions… et c'est faux !).

Aussi, comment traiter ce sujet, qui porte des valeurs avec lesquelles nous sommes aligné·es (protéger les personnes des relous, facile d'être pour…) mais qui dans le même temps, nous agresse également ? Et comment faire cela sans tomber dans le *tone policing*, argument bien trop utilisé pour couper court à un débat en tablant sur la forme plutôt que le fond ?

> *Tone-policing : littéralement « modération du ton ». Comportement visant à policer une discussion ou un débat en restreignant ou en critiquant les messages agressifs ou empreints d'une forte charge émotionnelle.*

Voici mon point de vue : au vu de notre petite taille, nos faibles capacités humaines (des salariés qui font mille choses, des bénévoles), que notre structure est à but non lucratif (donc n'est pas là pour faire du bénéfice pour enrichir quelqu'un) et notre posture d'écoute que nous essayons d'équilibrer avec notre énergie disponible, il n'est pas OK de mal nous parler sauf cas exceptionnel.

Alors bien sûr les cas exceptionnels, ça se discute, ça se débat, mais en gros, à part si nous avons sérieusement blessé quelqu'un de par nos actions, il est possible de nous faire des critiques calmement, sans faire preuve de violence.

Et dans un souci d'équité, nous faisons l'effort de creuser les arguments de fond quand bien même la forme ne nous plait pas, afin de vérifier si les reproches que nos recevons sont des critiques avec de la valeur et de l'agressivité, ou seulement de l'agressivité.

Cela peut vous sembler étrange de passer du temps là-dessus, mais tant que vous

n’êtes pas dans l’association, vous ne vous rendez pas compte de la quantité de gens qui la critiquent. Chaque semaine, des personnes nous signalent par exemple qu’elles sont désespérées par notre utilisation de l’écriture inclusive alors que nous l’utilisons depuis plus de quatre ans. Ou régulièrement parce que nous sommes trop politiques. Alors oui, des reproches agressifs nous en avons régulièrement, mais des critiques constructives, c’est malheureusement bien plus rare.

Donc que faire face à cela ? Eh bien mettre le soin au centre pour les personnes qui reçoivent les agressions. Alors encore une fois, je ne parle pas de soutenir les méchants oppresseurs agressifs envers qui on serait en colère, il ne s’agit pas de ça. Il s’agit de réussir à voir chez une personne de notre archipel qu’elle a fait une erreur, tout en prenant soin de l’humain qui a fait cette erreur.

Nous voulons prendre soin : de nous, d’autrui, et des Communs qui nous relient

Parce que personnellement j’ai vu que se faire agresser alors que l’on fait honnêtement de son mieux n’entraine qu’un repli sur soi, coupant la personne de son désir de contribuer à un monde meilleur, parce qu’elle a été trop blessée.

Et bien sûr, l’équilibre entre protéger les personnes blessées et celles qui blessent **sans le vouloir** est extrêmement difficile à tenir. Il faut donc réussir à faire preuve de franchise : oui, là, tu ou on aurait pu faire mieux. À partir de maintenant, comment on fait pour que se soit possible de faire mieux ?

Inutile de dire que non seulement ça prend du temps, mais c’est aussi carrément risqué, car on peut se mettre à dos par effet domino tout le groupe dont les membres exprimaient directement cette agressivité, ou indirectement par du soutien aux contenus en questions. Que du bonheur !

Les autres cas…

Alors oui il y a des robots, mais ça c'est un peu pénible mais vite résolu.

Il y a aussi des spécificités liées à Mastodon : comment gérer des contenus qui sont sur un autre Mastodon, auquel vous êtes connectés, et qui sont illégaux dans votre pays mais légaux dans d'autres ? On prend souvent l'exemple dans ce cas du *lolicon* : les dessins « érotiques » représentant des mineures sont autorisés et courants au Japon, mais interdits en France. Il faut y penser !

Enfin un autre cas que vous avez en tête est peut-être le contenu pornographique. Nous n'en faisons pas grand cas ici : nous demandons à ces comptes de masquer par défaut leurs images (une fonctionnalité de Mastodon qui permet de publier la photo floutée et de la rendre nette sur simple clic) et idéalement de rajouter un #NSFW (*Not Safe for Work*) pour simplifier leur identification.

La suite

J'espère que ces cas un peu concrets vous permettent de mieux vous rendre compte des interstices qui complexifient le travail de modération. Mais du coté fonctionnement collectif, comment ça se passe ? Je vous raconte ça dans l'article numéro 3 la semaine prochaine.