



O recrutamento de voluntários é como a moderação começou na Internet, muito antes de os gigantes das mídias sociais perceberem que os esforços centralizados eram necessários. E a moderação voluntária pode ser bem-sucedida, permitindo o desenvolvimento de regulamentos sob medida alinhados com as necessidades de comunidades específicas. Mas sem comprometimento e supervisão significativos da Meta, esse sistema não pode enfrentar quanto conteúdo é compartilhado nas plataformas da empresa e com que rapidez. De fato, o júri ainda está em como funciona bem em X, que é usado por 21% dos americanos (meta são significativamente mais populares - apenas o Facebook é usado por 70% dos americanos, de acordo com Pew).

As notas da comunidade, iniciadas em 2021 como Birdwatch, são um sistema de moderação orientado à comunidade em X que permite que os usuários se inscrevessem no programa adicionem contexto às postagens. Ter usuários regulares fornecendo verificação de fatos públicos é relativamente nova e até agora os resultados são misturados. Por exemplo, os pesquisadores descobriram que os participantes têm maior probabilidade de desafiar o conteúdo dos quais discordam politicamente e que a sinalização do conteúdo como falso não reduz o [engajamento](#), mas também descobriram que as notas são normalmente precisas e podem ajudar a reduzir a propagação de posts enganosos.

Sou um moderador comunitário que pesquisa a moderação da comunidade. Aqui está o que aprendi sobre as limitações de confiar em voluntários para moderação - e o que a meta precisa fazer para ter sucesso:

## **1. O sistema perderá falsidades e pode ampliar conteúdo odioso**

Existe um risco real sob esse estilo de moderação que apenas publica sobre coisas que muitas pessoas sabem que serão sinalizadas em tempo hábil - ou de alguma forma. Considere como um post com uma foto de um cogumelo de tampa de morte e a legenda "saborosa" podem ser tratados sob as notas da comunidade - moderação no estilo. Se um especialista em micologia não vê o post, ou o vê somente depois de ser amplamente compartilhado, ele pode não ser sinalizado como "venenoso, não coma" - pelo menos até que seja tarde demais. As áreas de [tópicos](#) mais esotéricas serão submoderadas. Isso pode ter sérios impactos em ambos os indivíduos (que podem comer um cogumelo venenoso) e a sociedade (se uma falsidade se espalhar amplamente).

Fundamentalmente, as notas da comunidade de X não são visíveis para os leitores quando são adicionados pela primeira vez. Uma nota se torna visível para a base de usuários mais ampla somente quando colaboradores suficientes concordam que é preciso votando nela. E nem todos os votos contam. Se uma nota for classificada apenas por pessoas que tendem a concordar entre si, ela não aparecerá. X não coloca uma nota visível até que haja um acordo de pessoas que discordaram nas classificações anteriores. Esta é uma tentativa de reduzir o viés, mas não é à prova de falhas. Ainda depende das opiniões das pessoas sobre uma nota e não de fatos reais. Muitas vezes, o que é necessário é a experiência.



Eu moderado uma comunidade no Reddit chamada R/AskHistorians. É um site de história pública com mais de 2 milhões de membros e é muito estritamente moderada. Vemos as pessoas errarem os fatos o tempo todo. Às vezes, esses são erros diretos. Mas às vezes há conteúdo odioso que leva especialistas a reconhecer. Uma vez, uma pergunta que contém um apito de cães do Holocausto escapou da [revisão](#) por horas e acabou acumulando centenas de votos antes de ser pego por um especialista em nossa equipe. Centenas de pessoas – provavelmente com padrões de votação muito diferentes e opiniões muito diferentes sobre muitos tópicos – não perdiam apenas a natureza problemática do conteúdo, mas optaram por promovê-lo através de votos. Isso acontece com respostas a perguntas também. Pessoas que não são especialistas em história votarão respostas desatualizadas e com som verdade que não estão realmente corretas. Por outro lado, eles voarão boas respostas se refletirem pontos de vista difíceis de engolir.

R/AskHistorians trabalha porque a maioria de seus moderadores é historiadores especializados. Se a Meta deseja que suas notas da comunidade – programa de estilo funcione, deve garantir que as pessoas com o conhecimento para fazer avaliações vejam as postagens e que a experiência seja contabilizada na votação, especialmente quando há um desalinhamento entre entendimento comum e conhecimento especializado.

## **2. Não funcionará sem voluntários bem apoiados**

Os moderadores de conteúdo pago da Meta revisam o pior dos piores – incluindo sangue, abuso e exploração sexual e violência. Como resultado, muitos sofreram traumas graves, levando a ações judiciais e esforços de sindicalização. Quando a Meta reduz os recursos de seus esforços centralizados de moderação, ele estará cada vez mais até voluntários não pagos para manter a plataforma em segurança.

Os moderadores comunitários não têm um trabalho fácil. Além da exposição ao conteúdo horrível, como membros identificáveis de suas comunidades, eles também estão frequentemente sujeitos a assédio e abuso – algo que experimentamos diariamente em R/AskHistorians. No entanto, os moderadores comunitários moderam apenas o que podem lidar. Por exemplo, embora eu gerencie rotineiramente o discurso de ódio e a linguagem violenta, como moderadora de uma comunidade baseada em texto, raramente sou exposta a imagens violentas. Os moderadores comunitários também trabalham em equipe. Se eu for exposto a algo que acho perturbador ou se alguém está sendo abusivo, meus colegas assumem o controle e fornecem apoio emocional. Eu também me preocupo profundamente com a comunidade que moderei. Cuidado com a comunidade, colegas de apoio e auto-seleção ajudam a manter o moral dos moderadores voluntários (ISH).

(tagstotranslate) meta