



Por trás da for you

O dia a dia do moderador de conteúdo sob o controle do TikTok

Lúcio, Moderador de conteúdo anônimo



Suggested citation:

Lúcio. 2026. "Behind the For You - Content Moderators under TikTok's Control". [“Por trás da for you: O dia a dia do moderador de conteúdo sob o controle do TikTok”]. In: M. Miceli, A. Dinika, K. Kauffman, C. Salim Wagner, and L. Sachenbacher (eds.) *Data Workers' Inquiry*. Creative Commons BY 4.0. <https://data-workers.org/lucio/>



DATA

weizenbaum
institut

TU
Technische
Universität
Berlin

Por trás da for you: O dia a dia do moderador de conteúdo sob o controle do TikTok © 2026 by Lúcio is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>



Introdução

Meu nome é Lúcio, e trabalho na área de moderação de conteúdo há cerca de quatro anos. Pode parecer pouco tempo, mas dentro desse universo onde tudo se move em uma velocidade diferente do resto do mundo, tudo é para “ontem” e as regras mudam de uma hora para outra, essa experiência equivale a uma vida inteira. Baseado em minha experiência pessoal e diversas conversas com colegas ao longo dos anos, esse texto oferece uma análise sobre as formas como gerentes e empresas de tecnologia constroem um sistema para extrair cada vez mais produtividade de seus funcionários ao custo da saúde e bem-estar destes. Moderadores de conteúdo lidam com conteúdo sensível e traumatizante, que causa impactos profundos em nossa saúde, mas, além disso, enfrentamos desafios ligados a estratégias de gestão que priorizam números em detrimento do bem-estar humano. Esse relato expõe um sistema de métricas implacável, competição constante entre colegas e a falta de perspectiva de carreira.

Meu objetivo ao compartilhar esta experiência é duplo: primeiro, desvendar para os leigos os bastidores da moderação em uma rede social multinacional; segundo, mostrar para quem já conhece a área que nossas experiências e dores raramente são únicas. Eu descrevo o sistema de supervisão e gestão, mas sobretudo é um relato sobre como esse trabalho transforma vidas e molda mentes, explorando a realidade frequentemente ignorada que existe por trás da missão crucial de garantir um ambiente digital seguro e confiável ao usuário final. Deixando evidente como as grandes plataformas digitais constroem sistemas que, enquanto prometem segurança aos usuários, negligenciam a saúde mental e o desenvolvimento profissional daqueles que sustentam essa promessa nos bastidores. Este é um alerta sobre os verdadeiros custos humanos por trás da curadoria do que você vê — e do que é mantido longe dos seus olhos — na sua For You Page.

PARTE 1

Entrando No Mundo Da Moderação

Em 2022, iniciei o meu trabalho como moderador de conteúdo na Byte Dance Brasil, na área de short vídeos do TikTok. A Byte Dance é a empresa que é dona do TikTok e eu fui contratado como funcionário em tempo integral na Vila Olímpia, em São Paulo, um bairro conhecido como o centro tecnológico da cidade e sede das principais empresas de finanças, tecnologia e fundos de investimento. Durante mais de três anos testemunhei de perto o funcionamento de uma operação de Trust and Safety, departamento responsável pela segurança dos usuários e qualidade da moderação da plataforma, que começou pequena e cresceu exponencialmente.

Moderação de conteúdo garante que o que é consumido pelos usuários de uma plataforma esteja em conformidade com suas regras de uso e diretrizes da comunidade. Como moderadores de conteúdo, nosso papel é analisar mídias, fotos, vídeos, comentários e imagens criadas pelos usuários e postados na plataforma. A importância do moderador nesse processo é garantir que o que chega na sua timeline (no TikTok, a for you) esteja dentro das regras de comunidade criadas pela plataforma, mantendo um ambiente seguro para os usuários.

Moderação sempre envolve decisões complexas, que exigem balancear a proteção dos usuários com sua liberdade de se expressar na plataforma. A partir desta descrição de moderação de conteúdo de redes sociais, é fácil de imaginar de maneira abstrata que se trata de ver tudo de ruim que há na internet. Mesmo antes de começar, eu já tinha ouvido relatos de que era muito ruim moderar as redes sociais, mas achava que minha experiência com moderação de comunidades, que envolveu vivenciar todos os tipos de ofensas possíveis, tinha me preparado um pouco. Ainda assim, sofri para me adaptar. Eu já tinha aprendido a lidar com afirmações preconceituosas, ofensas e racismo contra pessoas pretas ou latinas, mas isso não me preparou para o conteúdo sensível gráfico.

A operação na Byte Dance é estruturada em dois setores, ou rounds, compostos por várias equipes menores. A maior parte do meu tempo na Byte Dance, eu fui moderador de Round 2, ou seja, conteúdos que estão crescendo em engajamento na plataforma. Nossa time era responsável por moderar conteúdos com potencial de viralizar, e recebíamos tudo que alcançasse uma certa quantidade de visualizações. Geralmente esses conteúdos são parte de alguma trend, sempre recebemos muitas danças, vídeos de reação, conteúdo de bandas de k-pop, New United ou outras bandas jovens, ou vendendo acesso a plataformas de conteúdo sexual como *OnlyFans* e *Privacy*, muitas vezes também dentro da trend do momento. Ao longo dos anos a empresa expandiu nossas responsabilidades para outros produtos além de *short videos*, incluindo live streams e mensagens, imagens e áudios de *chats* (conversas privadas). O conteúdo é muito variado, mas para todos nós a importância do trabalho é clara, especialmente para proteger crianças. Eu vi em primeira mão a quantidade de conteúdo sexual disfarçado que poderia facilmente chegar na página de recomendações *for you*, assim como a velocidade assustadora com a qual os desafios ou *challenges* viralizam e se espalham. Lembro bem da época do perigosíssimo *skull break challenge*, no qual duas pessoas dão uma rasteira em uma terceira pessoa, que consequentemente cai de costas e bate a cabeça. A moderação de conteúdo tem um papel essencial de notificar essas trends e rastreá-las, além de remover os vídeos horripilantes e violentos que surgem em função delas.

A responsabilidade e a natureza perturbadora de alguns dos conteúdos que avaliamos pesam muito e deixam marcas profundas em quem faz esse trabalho. Não é por acaso que a equipe de moderação tem alta rotatividade e a maioria dos meus colegas está em tratamento psicológico ou psiquiátrico. Mas o baque maior é, e sempre foi, para os colegas que são moderadores de Round 1, que lidam com conteúdos considerados “sensíveis”. Em sua maioria, são materiais reportados por usuários por serem violentos, perigosos e/ou ilegais. E aqui estamos falando de conteúdos que são bem sensíveis *mesmo*: corpos, pessoas mutiladas, violência explícita e os piores tipos de abuso, também de crianças. As consequências psicológicas de trabalhar com conteúdo sensível e tóxico, não são o foco deste artigo, ainda assim, é importante deixar claro que todas as práticas de gerência documentadas aqui são implementadas no contexto de um trabalho extremamente demandante e potencialmente nocivo para a saúde mental e física dos trabalhadores¹.

O conteúdo sempre causou um impacto muito grande na saúde mental dos moderadores. Para quem entra, é uma surpresa e um baque muito forte. Após alguns meses na empresa, todos sentimos e observamos uma dessensibilização emocional: o conteúdo passa a ser somente mais um conteúdo, é só mais um vídeo, só mais uma foto. O impacto negativo na saúde mental é enorme. Mas esse texto vai além do conteúdo, focando na rotina do moderador e nas estratégias da gerência para aumentar constantemente a produtividade, criando regras e práticas focadas em alcançar metas cada vez mais altas e deixando moderadores exaustos e com níveis de estresse perigosos. Meu objetivo é mostrar que os desafios que os moderadores enfrentam não vem somente do conteúdo e sim, também, de um meio tóxico que usa o ser humano e a força de trabalho somente como números.

Eu trabalhei na Byte Dance por aproximadamente três anos, e hoje estou em uma empresa de serviços terceirizados ou BPO (na sigla em inglês para business process outsourcing), moderando conteúdo para outra rede social e percebo muito do mesmo tratamento. Reconheço os mesmos padrões do TikTok, e isso mostra que não é um sistema diferente de empresa para empresa e sim, de um setor que explora trabalhadores. Além do conteúdo, há cobrança intensa por métricas a serem atingidas e se cria uma competição entre os funcionários para produzir cada vez mais.

¹ Outras contribuições ao Data Workers' Inquiry exploram em profundidade os impactos psicológicos deste trabalho. Particularmente: <https://data-workers.org/kauna/> <https://data-workers.org/fasica/>

PARTE 2

Métricas E Mais Métricas

Logo no início ficou claro que a rotina na empresa é muito regrada e as métricas essenciais. A vida dos moderadores gira em torno de três métricas: tempo de moderação, velocidade e qualidade.

Tempo de moderação se refere ao tempo que de fato estamos logados no sistema avaliando conteúdo. Todos os moderadores têm que entregar 6 horas de moderação no dia, o que parece fácil a princípio, mas na realidade é muito desafiador se pensarmos em todas as atividades que fazem parte do trabalho. O time tem uma série de reuniões semanais para atualizar e esclarecer policies que não contam para esse tempo, apesar de serem obrigatórias e essenciais. Pausas para banheiro ou para esticar as pernas não contam para esse tempo, assim como as sessões obrigatórias de bem-estar, chamadas de wellbeing. Ou seja, atingir essa meta exige muita coordenação do dia-a-dia e fatores como reuniões inesperadas, ou o seu dia começar atrasado por ter que ir até o presencial da empresa atrapalham o andamento do seu dia e torram manter as métricas por entrega mais complexas.

A segunda métrica é a média de tempo que o moderador leva por vídeo ou task. O objetivo a ser atingido era de um vídeo a cada dez segundos. O que por si só já é difícil, considerando que a maior parte dos conteúdos dura pelo menos 30 segundos. Mas o desafio maior é manter uma boa velocidade e, ao mesmo tempo, manter a média de qualidade em 97%, ou seja, tomar a decisão correta na maioria esmagadora dos casos.

A qualidade é medida por um time de qualidade que re-modera uma pequena quantidade de casos. Se há divergência entre a decisão dos moderadores e da qualidade, tanto na ação apropriada a ser tomada, quanto na categorização de qual policy o conteúdo viola, isso é computado como um erro do moderador e baixa sua média. O sistema é estruturado de forma bastante desfavorável aos moderadores, pois o time de qualidade tem tempo para avaliar cada conteúdo criteriosamente enquanto os moderadores trabalham sob pressão. Além disso, a qualidade só re-modera uma pequena quantidade de casos, tornando essa métrica muito imprevisível. Se eu fazia em torno de sete mil casos durante a semana, a qualidade re-moderava em torno de vinte, aleatórios, e minha métrica de qualidade se baseava somente nestes. Se você tem mais casos sendo revisados pela qualidade, tem mais chance de terem erros, baixando a sua média. Porém, uma pessoa que tem menos, o impacto de cada um na média é muito maior, gerando um resultado distorcido que pode ser muito maior ou menor do que sua taxa real de acertos.

A loteria da qualidade

Como não há uma proporção fixa para estabelecer quantos casos a qualidade checa, a nota de qualidade sempre é imprevisível.

O exemplo mais extremo é quando apenas um dos seus casos passa pela qualidade, se estiver certo sua nota é 100% naquela semana, se estiver errado é 0%. Uma nota dessa derruba a sua nota pelo mês inteiro, e não é possível voltar para dentro da meta.

Você pode ser muito sortudo e ter números altos de qualidade e entregar poucos casos. Se em uma semana você entregar três mil casos, enquanto meus colegas entregaram sete mil, eles provavelmente teriam mais casos re-analisados e você pode se dar bem no processo.

Mas de fato, a nota não é um número que expressa realmente o quanto esse moderador realmente acertou ou não, é um jogo de sorte, sem margem de erro.

A vida dos moderadores gira em torno dessas métricas de desempenho por um motivo muito simples: notas baixas quase sempre significa demissão logo no mês seguinte e notas altas garantem pagamento de bônus. As três métricas eram a base para a nota mensal de desempenho, que por sua vez eram avaliadas duas vezes ao ano. Quase como o boletim no fim do semestre escolar. As *performance reviews* acontecem em agosto e em março, e apesar de nossa vida ser regida por números, o processo não é muito claro. O que sabemos é que logo após as reviews sempre há uma série de demissões e notas baixas quase sempre significam demissão. Enquanto notas altas garantiam bônus generosos por boa *performance*, chegando a até três salários para os melhores, e ainda poderiam alavancar uma promoção. Fica fácil de imaginar o estresse que toma conta dos times de moderação na época das reviews.

PARTE 3

Conflitos Com A Qualidade

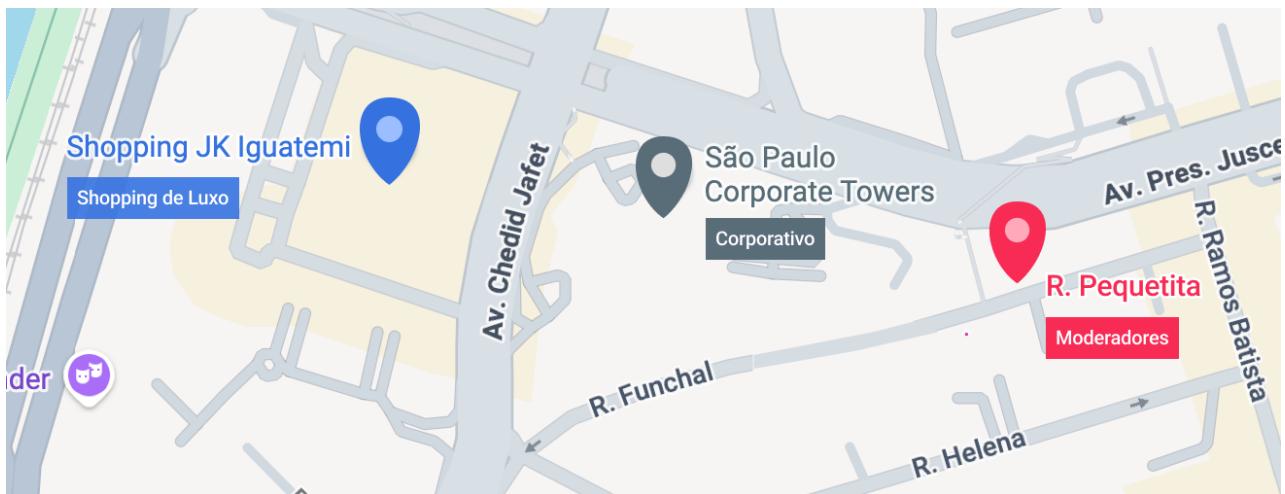
Quando comecei, me dediquei muito nos primeiros meses tentando aprimorar, me dedicando no estudo das policies que mudam constantemente e trabalhando focado. Porém, sempre foi muito difícil alcançar todas as métricas e logo percebi que eu não era o único, que o sistema de métricas não é feito com nosso sucesso em mente. Todos tínhamos semanas em que a quantidade de casos que passavam pela qualidade era baixa e a nota de qualidade se tornava totalmente imprevisível. Atritos entre moderação e qualidade por conta das métricas são uma constante.



Co-working onde os times de moderação ficavam, as mesas eram rotativas e nunca havia espaço suficiente para todos

Enquanto moderação é cobrada por quantidades de casos diários, tempo e acertos durante o processo de moderação, a área da qualidade é cobrada para garantir que os processos estão sendo feitos e sendo seguidos de acordo com o que foi estipulado. Isso gera um atrito entre as duas áreas que buscam coisas diferentes. O time de qualidade tem sua performance afetada somente pelo trabalho deles, já o time de moderação tem os resultados afetados pelo trabalho desse outro departamento. Não é um atrito individual, mas há conflitos constantes que precisam ser discutidos, mas não são.

Essa estrutura gera um conflito entre a área de moderação e a área de qualidade, uma contra a outra e cada uma atrás do seu objetivo. Mas, na verdade, ambas são parte do processo de manter a segurança dos usuários na plataforma. O único motivo para os atritos constantes é a decisão da empresa de separar os processos de qualidade e de moderação, como se cada uma tivesse um objetivo diferente e esquecendo que os resultados de moderação dependem de qualidade. Isso também se reflete na separação espacial, pois a qualidade não fica nem no mesmo andar que nós. Isso faz-se esquecer que tá todo mundo junto no mesmo barco, não é uma guerra ou um conflito, o diálogo é essencial.



Os edifícios onde o corporativo e os moderadores estavam localizados eram fisicamente separados.

Durante meus anos na empresa, era claro que nos times nos quais a moderação tinha menos oportunidade de crescimento e de expor a sua perspectiva à qualidade, havia uma grande diferença entre as notas, piores desempenhos e mais cobrança para os dois lados. Enquanto nos times onde a moderação trabalhava em conjunto com a qualidade, aí sim, víamos uma melhora expressiva, o ambiente de trabalho era mais tranquilo e leve; quando havia dúvidas os canais de diálogo eram claros e havia mais alinhamento. Minha experiência deixa claro que essa separação desnecessária entre as áreas piora e muito a vida dos moderadores. No entanto, a gestão da Byte Dance, que inclusive escolheu alocar tanto moderação quanto qualidade em um prédio separado daquele ocupado pelo corporativo, escolhe manter os processos assim.



São Paulo Corporate Towers: a torre chiquérrima onde ficam os escritórios do corporativo

PARTE 4

Competição

Como já disse, a avaliação das nossas métricas afeta diretamente nossa remuneração e as dificuldades geradas pela estruturação dos processos de qualidade são apenas uma parte do problema. Além disso, há um sistema intransparente de notas mensais, no qual a performance é comparada entre todos os moderadores em comparativo ao time. Ou seja, aquele resultado imprevisível de qualidade, assim como as métricas de tempo de moderação e de velocidade, são comparados aos resultados de todos os moderadores. Durante as temidas *performance reviews*, cada um recebe uma nota que varia de A a D e reflete um agregado de todas as métricas e médias do semestre.

Para entender melhor a situação e como surgia a competição ferrenha, é preciso entender que há uma quantidade limitada de notas para distribuir durante essa avaliação. Por exemplo, em uma equipe de 20 pessoas, podem haver no máximo duas notas A, quatro notas B, cinco notas C e o restante de notas D. Mesmo quando minha média de qualidade é 97% e eu entreguei todas as horas de produção, cumprindo com tudo que a empresa pede, isso não garante que eu receba uma boa nota com A, ou B. Quando o time todo ou a grande maioria está dentro dessas métricas isso não é mais o suficiente: a META para alcançar vai subindo com base nas outras pessoas, se tem alguém com tudo 100%, você tem que ter essa nota também e fazer a mais que ele para conseguir a nota A.

A meta não é mais o que você precisava entregar para se manter no cargo, ela se torna o mínimo do mínimo para não receber uma nota negativa. Lembrando que tudo isso influencia o bônus anual e possíveis promoções de cargo e de salários.

Existe uma política muito clara de você não poder saber nada sobre os seus colegas, e também não poder perguntar. Há uma energia que paira, ninguém sabe o salário do próximo, qual grau de salário ele tem (promoção de salário) ou as notas mensais dentro do time. O que é mais frustrante ainda, já que somos comparados com o time sem saber os reais números dos melhores do seu time. É uma competição cega, na qual mesmo que você tivesse acima das métricas, não há garantias. Houve um momento que todo mundo estava entregando acima das métricas, fazendo com que as metas se tornassem o básico esperado.

Como moderadores, nós frequentemente não sabíamos nossas próprias métricas, muito menos as de nossos colegas ou como a liderança agrava cada um destes indicadores para estabelecer a comparação. A competição entre moderadores prospera nesse segredo e silêncio com consequências terríveis para nossa saúde mental, gerando estresse e ansiedade.

Claramente, só quem se beneficia desta estrutura de gerência é a empresa que consegue extrair mais produtividade de seus funcionários. Porém, ao custo da saúde e bem-estar destes, assim como o bom funcionamento do departamento de trust and safety que deixa de conseguir avaliar riscos cuidadosamente. Há um problema de gestão com líderes que não são capacitados para liderar e processos de qualidade separados da moderação ao invés de alinhados para melhorar a segurança da plataforma. Este controle ferrenho da produtividade do funcionário, que exige que

até mesmo pausas para ir ao banheiro ou buscar água sejam marcadas, é extremamente desnecessário, desumanizante e serve somente para uma micro gestão.

As dinâmicas da competição

Vamos usar como exemplo um time de moderadores que não moderam short video, como foi o meu caso durante uma época. Cada um deles tem as seguintes médias em um mês:

Lembrando que as metas estabelecidas são 6 horas de produção, 97% de qualidade e uma média de 10 segundos por caso.

Pessoa A	Pessoa B	Pessoa C
Modera chats	Modera chats	Modera live streams
Tempo de produção: 6h	Tempo de produção: 6h30	Tempo de produção: 6h
Ritmo: 9 seg/ caso	Ritmo: 9 seg/ caso	Ritmo: 25 seg/ caso
Qualidade: 98%	Qualidade: 96,5%	Qualidade: 97%
Casos por semana: 9.000	Casos por semana: 9.200	Casos por semana: 7.000

Nesse exemplo fictício, os moderadores A e B têm clara vantagem, pois estão lidando com conteúdos mais curtos e simples. Já a pessoa C, que lida com live streams muito mais longos, sabe que está em desvantagem. Mesmo tendo uma qualidade muito alta e próxima do time, seu ritmo é mais lento e provavelmente ficará com uma nota baixa. Ainda assim, a estrutura de incentivos provavelmente motiva a pessoa B a fazer horas extras e produzir mais, já que as métricas são analisadas para o time como um todo. De novo, destaco que nenhuma sabe os resultados das outras, o que incentiva ainda mais a competitividade.

Esta estrutura não é um sistema humano e não funciona. Suas consequências são alta rotatividade de funcionários, péssima saúde mental em toda a equipe de moderação, e qualidade e saúde corporativa decrescentes. Gestão ruim adoece todo o sistema de operação e para quem já está exausto, mais estresse é destruir o psicológico, e a saúde mental.

PARTE 5

Rua Sem Saída

A pressão para atingir e ultrapassar as métricas é máxima. Processar a quantidade de critérios para avaliação dos posts exige atenção constante e lidar com conteúdos sensíveis também cobra seu preço, emocional, físico e psicológico. É preciso deixar absolutamente claro que estamos falando um cargo de entrada, um cargo onde você, teoricamente, se desenvolveria em novas frentes. Porém, na Byte Dance não havia um plano de carreira estabelecido e estruturado para área de moderação. Nem de perto!

A progressão mais natural seria avançar para o cargo de analista de qualidade. Mas em 2021, quando houve uma grande expansão no time de qualidade, a empresa preferiu contratar pessoas de qualidade do mercado ao invés de aumentar as promoções internas. Ou seja, quem já era moderador e esperava aberturas de vagas para qualidade, que seria a promoção direta, acaba perdendo oportunidade, pois a vaga não era disponibilizada nem internamente. Nos poucos momentos em que houve esse tipo de abertura era uma quantidade ínfima, exemplo uma ou duas vagas para todo o time de moderação.

E não importa quantos anos você tenha de moderação, em toda empresa que você vai entrar, a moderação será o cargo de entrada e a falta do crescimento é a regra. Você precisa se especializar em outras funções para conseguir receber a promoção. Somente o seu trabalho não é o suficiente, tem que além do seu trabalho, saber lidar com gestão de pessoas, processos e desenvolvimento. Se cobra muito mais de quem está dentro do que de pessoas que vem do mercado.

Durante todo o período que eu estava lá, não houve uma perspectiva de crescimento, uma possibilidade de ascender. Muitos colegas excelentes, com capacidade, entraram ali como moderador e saiam dali como moderador. Não havia um plano de carreira sólido e estipulado, então, bater meta era somente uma forma de não ser demitido. Não havia ali um desenvolvimento para que todos pudessem crescer na empresa, para que todo aquele conhecimento de moderação, de interação da rede com o usuário, pudesse se tornar algo para beneficiar o moderador e torná-lo uma pessoa melhor preparada para dialogar com outras áreas do *trust and safety*.

Cada vez mais o debate da responsabilidade das redes sociais se torna um tema forte e necessário tanto no debate público quanto dentro das empresas. A moderação é a parte fundamental desse processo, mas não tem uma grande participação de quem está moderando nesse debate. Somente somos uma forma de números, o conteúdo deve ser analisado, e somente isso que importa.

IA salvadora?

Hoje empresas estão treinando IA para substituir a moderação de conteúdo humana. O Tiktok inclusive já começou layoffs na moderação, causando demissões em massa no Brasil², Malásia, Alemanha³ e ao redor do mundo⁴. O problema é que sempre há usuários mal-intencionados testando a plataforma para tentar se promover. Somente moderadores entendem o contexto, as dinâmicas da plataforma e onde o usuário está testando os limites das policies. É algo comum e rotineiro para moderadores sentir que o usuário está tentando ou até mesmo conseguindo burlar as regras e já começar a desenvolver planos de ação.

Sistemas de IA, em contraste, nunca vão conseguir identificar novas táticas e formas de apresentar conteúdos sensíveis ou perigosos a tempo de prevenir que estes cheguem aos usuários. IA é boa para quantidade de casos para serem moderados, não para qualidade desses casos. O processo de aprendizagem de máquina depende da disponibilidade de uma grande quantidade de dados e o trabalho humano vai ser sempre indispensável para manter a IA atualizada. As novas formas de burlar o sistema tem que ser identificadas ou comunicadas por usuários ao time responsável, que a partir disso elaborar um plano de ação, coletar e moderar os dados necessários e colocar a IA aprender de novo.

Sem o moderador, alguém que coloque contexto e cultura, história e diversidade na área, a IA vai ser só mais um amontoado de 0 e 1 repetindo o que foi ensinado antes, não adicionando nenhuma informação, nenhuma diversidade e assim deixando de proteger os usuários.

² "TikTok demite centenas, enquanto amplia uso de IA para moderação de conteúdo". October 11, 2024. Época Negocios. Fabiana Rolfini. <https://epocanegocios.globo.com/empresas/noticia/2024/10/tiktok-demite-centenas-enquanto-amplia-uso-de-ia-para-moderacao-de-conteudo.ghtml>

"Em ano eleitoral, TikTok faz demissões no Brasil em sua equipe de moderação e segurança". March 19, 2024. Estadão. Henrique Sampaio. https://www.estadao.com.br/link/empresas/ano-eleitoral-tiktok-demissoes-seguranca-brasil-nprei/?srsltid=AfmBOoomwgyZFpd3dreAttPk1SwHxsgUXa_ihBc1qgSKzJ2siNEX4qFT

³"TikTok: Gemeinsam gegen Massenentlassungen". November 18, 2025. Ver.di. <https://www.verdi.de/themen/arbeit/++co++19d30bc8-622c-11f0-937c-b715127af524>

⁴"ByteDance's TikTok cuts hundreds of jobs in shift towards AI content moderation. October 11, 2024. Reuters. Rozanna Latiff. <https://www.reuters.com/technology/bytedance-cuts-over-700-jobs-malaysia-shift-towards-ai-moderation-sources-say-2024-10-11/>

"TikTok to lay off hundreds of UK moderators as it shifts to AI". Aug 22, 2025. Financial Times. Anna Gross and Tim Bradshaw. <https://www.ft.com/content/d277c456-4518-4994-8afd-c45a399db342>

Conclusão

A experiência como moderador de conteúdo revela uma contradição fundamental no coração das grandes plataformas digitais: a mesma indústria que se gaba de criar espaços seguros e conectados constrói seus alicerces sobre a exploração sistemática de trabalhadores invisíveis. Este relato demonstra que os desafios da moderação vão muito além do conteúdo gráfico e perturbador: eles estão enraizados em um modelo de negócio que trata seres humanos como recursos descartáveis.

No fim do dia, tá todo mundo fazendo o seu trabalho, tentando pagar as contas do mês e a empresa como organização se aproveita desse estado de necessidade do trabalhador para conseguir tirar mais e mais de cada função possível. A rixa entre áreas e times é competitiva e dá mais produtividade, mas ao custo dos funcionários. Esse texto explora como as grandes empresas de tecnologia empregam métricas, no caso, tempo de moderação, velocidade e qualidade, não apenas para acompanhar o bom funcionamento do sistema de *Trust and Safety*, mas como instrumentos para aumentar a produtividade sem aumentar salários ou o número de funcionários.

A separação artificial entre moderação e qualidade exemplifica como as empresas fragmentam processos que deveriam ser colaborativos, criando conflitos desnecessários que prejudicam tanto os trabalhadores quanto a eficácia do sistema de segurança como um todo. Enquanto a competição forçada por notas em um sistema de curva de distribuição gerava um ambiente tóxico de rivalidade e ansiedade. Esta não era simplesmente uma gestão ineficiente, mas uma arquitetura deliberada para maximizar a produtividade a qualquer custo, mesmo que esse custo seja a saúde mental dos trabalhadores.

Hoje estando em outra empresa de tecnologia, entendo que não é algo unicamente da ByteDance. Parte da surpresa e quebra de expectativa sobre as condições de trabalho que enfrentamos é por ser um nome atrelado a uma *BIG TECH*. Mas a realidade que encontrei é estrutural, nessa nova empresa há os mesmos problemas. A falta de interesse em trazer moderadores para o centro de um debate sobre *policy* e a falta de formas para especializar a aproveitar o conhecimento que os moderadores têm sobre o conteúdo. Além de uma falta generalizada de perspectivas de carreira que condena moderadores talentosos a uma estagnação profissional, desperdiçando conhecimento valioso que poderia enriquecer outras áreas de *trust and safety*.

Tudo isso é um sistema, independente de ser contratado diretamente ou por uma empresa terceirizada, que presta o serviço para a big tech. Não é sobre empresas de tecnologia, é sobre um sistema que usa e explora a nossa força de trabalho, a nossa saúde física e mental, sem querer que sejamos mais do que somente uma massa de trabalho. Além do controle que se estende até mesmo a nossas pausas mais básicas, a situação é agravada pelo silêncio imposto, não se pode discutir salários, métricas ou condições, criando uma cultura de isolamento que beneficia somente o empregador. Este testemunho serve como um alerta urgente: não podemos continuar ignorando as condições dos trabalhadores que sustentam nossa experiência digital. A moderação de conteúdo é fundamental para a segurança online, mas precisa ser reconhecida como um trabalho especializado que merece condições dignas, perspectivas de carreira e proteções adequadas contra os danos psicológicos inerentes à função.

