

INTELIGÊNCIA ARTIFICIAL “ARTIFICIAL”:
o trabalho oculto dos moderadores de conteúdo comercial

“ARTIFICIAL” ARTIFICIAL INTELLIGENCE:
the hidden work of commercial content moderators

Ney Maranhão*
Thiago Amaral Costa Savino**

RESUMO

Discute-se a definição e a dinâmica da atividade humana de moderação de conteúdo para plataformas de mídia social. Problematiza-se a centralidade desse tipo de trabalho para a própria viabilidade de um tal modelo de negócio, bem como a paradoxal invisibilidade técnica e social que, insistentemente, recai sobre tais trabalhadores. Conclui-se que as plataformas digitais devem fornecer maior transparência sobre as condições de trabalho de seus moderadores de conteúdo, havendo de garantir-lhes, ademais, plena sintonia com o vetor ético-jurídico do trabalho decente. A pesquisa é qualitativa, eminentemente bibliográfica, tendo sido utilizado o método hipotético dedutivo.

* Professor de Direito do Trabalho da Universidade Federal do Pará (UFPA). Professor Permanente do Programa de Pós-Graduação *Stricto Sensu* em Direito da Universidade Federal do Pará (PPGD/UFPA – Mestrado e Doutorado). Doutor em Direito do Trabalho pela Universidade de São Paulo (USP), com estágio de Doutorado-Sanduíche junto à Universidade de Massachusetts (Boston/EUA). Especialista em Direito Material e Processual do Trabalho pela Universidade de Roma – La Sapienza (Itália). Mestre em Direitos Humanos pela Universidade Federal do Pará (UFPA). Professor instrutor da Escola Nacional de Formação e Aperfeiçoamento de Magistrados do Trabalho (ENAMAT/TST). Professor convidado em diversas Escolas Judiciais de Tribunais Regionais do Trabalho. Professor Coordenador do Grupo de Pesquisa “Contemporaneidade e Trabalho” – GPCONTRAB (UFPA/CNPQ). Titular da Cadeira nº 30 da Academia Brasileira de Direito do Trabalho (ABDT). Titular da Cadeira nº 25 da Academia Paraense de Letras Jurídicas (APLJ). Juiz Titular de Vara da Justiça do Trabalho da 8ª Região (PA-AP). Lattes: <http://lattes.cnpq.br/5894619075517595> / ORCID: <https://orcid.org/0000-0002-8644-5902> / E-mail: ney.maranhao@gmail.com / Facebook: Ney Maranhão II / Instagram: @neymaranhao / Youtube: Prof. Ney Maranhão / Twitter: @ney_maranhao

** Acadêmico do Curso de Direito da Universidade Federal do Pará (UFPA). Bolsista PIBIC/CNPq. Membro do Grupo de Pesquisa “Contemporaneidade e Trabalho” – GPCONTRAB (UFPA/CNPQ). E-mail: thiagosavino1999@gmail.com.

PALAVRAS-CHAVE

Moderação de conteúdo comercial. Plataforma de mídia social. Risco ocupacional. Trabalho decente.

ABSTRACT

The definition and dynamics of the human activity of content moderation for social media platforms is discussed. This type of work for the viability of such a business model centralization, as well as the paradoxical technical and social invisibility that insistently falls on such workers, is questioned. It is concluded that digital platforms must provide greater transparency about the working conditions of their content moderators, in addition to guaranteeing their full alignment with the ethical-legal vector of decent work. The research is qualitative, eminently bibliographic, using the hypothetical deductive method.

KEYWORDS

Commercial content moderation. Social media platforms. Occupational risks. Decent work.

SUMÁRIO

- 1 Introdução;
- 2 O que é moderação de conteúdo comercial?;
 - 2.1 Moderação de Conteúdo: acessório ou elementar?;
- 3 Os humanos por trás da tela;
 - 3.1 Arranjos de trabalho em moderação de conteúdo comercial;
 - 3.2 Condições de trabalho e riscos ocupacionais;
- 4 Rumo ao trabalho decente;
- 5 Conclusão;
- Referências.

Data de submissão: 09/03/2021.

Data de aprovação: 02/05/2021.

1 INTRODUÇÃO

Nas últimas décadas, temos observado as plataformas digitais crescerem exponencialmente em tamanho, alcance e influência em todo o mundo. Eles se tornaram parte do nosso cotidiano, um espaço *online* no qual nos conectamos e compartilhamos com amigos, familiares e novas pessoas as fotos, experiências, informações e assim por diante.

Para manter uma certa ordem entre um número tão grande de pessoas, um conjunto de regras teve de ser criado para manter o que nós, como sociedade, consideramos inaceitável fora da rede – e, em consequência, um sistema para reforçar esses padrões. Assim, a Moderação de Conteúdo Comercial (MCC) foi projetada para atender a essa demanda e filtrar o pior que a internet tem a oferecer. O que a maioria das pessoas não sabe, no entanto, é que esse trabalho não é feito, exclusivamente, por inteligência artificial ou outros processos automatizados, mas por seres humanos muitas vezes em empregos precários, de baixa remuneração, sob o risco de *Burnout*, Transtorno de Estresse Pós-traumático, ou pior, como consequência da natureza de seu trabalho, em um processo **projetado** para fazê-lo parecer invisível e autômato – ou como inteligência artificial “artificial”.

As plataformas digitais se apresentam, publicamente, como espaços “abertos, imparciais, conectivos, progressivos, transformadores” (GILLESPIE, 2018, p. 13, tradução nossa),¹ e a moderação como periférica ao que fazem – como varrer o chão –, um fardo que **tiveram** não por escolha, mas por ter de suportar para tornar a rede um lugar melhor e mais seguro para todos. Essa percepção das plataformas e da MCC limitou nossa capacidade de fazer perguntas sobre como elas o fazem, quem faz esse trabalho, sob quais condições e a que custo.

A definição de plataforma que norteia esta pesquisa é a adotada por Gillespie (2018, p. 16-17, tradução nossa), no sentido de “sites e serviços *online* que hospedam, organizam e veiculam conteúdo compartilhado pelos usuários ou interações sociais para eles, sem ter produzido ou comissionado (o “grosso” d) esse conteúdo”, construído como uma infraestrutura invisível “para processamento de dados para atendimento ao cliente, publicidade

¹ [O]pen, impartial, connective, progressive, transformative (GILLESPIE, 2018, p.13).

e lucro”,² e que modera “o conteúdo e a atividade dos usuários, utilizando alguma logística de detecção, revisão e fiscalização”.³

Há poucas pesquisas, especialmente em Direito, que discutem as condições de trabalho e os riscos ocupacionais pelos quais os moderadores são submetidos. A maioria dos estudos existentes tem se concentrado nas consequências políticas e éticas da MCC e seu impacto sobre o discurso público, a liberdade de expressão e a transparência – por exemplo, em relação às *fake news*. Quase não há dados sobre os trabalhadores que fazem esse trabalho e é pouca a informação sobre o que fazem, por qual treinamento eles passam, como são suas condições de trabalho, que tipos de diretrizes recebem, e assim por diante (CARMI, 2019).

Diante disso, tencionamos problematizar a centralidade desse tipo de trabalho para a própria viabilidade do modelo de negócio via plataformas digitais, bem como a paradoxal e constante invisibilidade técnica e social que recai sobre tais trabalhadores, sobretudo para que se avalie sua sintonia com o vetor ético-jurídico do trabalho decente. A pesquisa é qualitativa, eminentemente bibliográfica, tendo sido utilizado o método hipotético dedutivo.

Na seção seguinte, será discutida a definição de moderação de conteúdo comercial, como é tratada pelas plataformas e como, de fato, elas são constitucionais e elementares para seu modelo de negócios. Em seguida, nosso foco será direcionado para expor e

² *[P]latforms are: online sites and services that host, organize, and circulate users' shared content or social interactions for them, without having produced or commissioned (the bulk of) that content, built on an infrastructure, beneath that circulation of information, for processing data for customer service, advertising, and profit* (GILLESPIE, 2018, p. 16).

³ *[M]oderate the content and activity of users, using some logistics of detection, review, and enforcement* (GILLESPIE, 2018, p. 17). Essa definição não inclui serviços de mensagens como *WhatsApp*, *Facebook Messenger*, *Google Hangout* e *Skype*, pois geralmente envolvem conversas privadas entre grupos fechados e específicos, sem o mesmo acesso público e visibilidade que outras plataformas oferecem, ainda que tenham suas próprias formas de moderação (GILLESPIE, 2018, p. 17).

destacar o que ser um moderador significa: seus arranjos de trabalho, organização e condições de trabalho e alguns dos riscos a que estão sujeitos. Por fim, consignaremos nossa conclusão e proporemos alguns cursos de ação e parâmetros mínimos a serem alcançados quanto à MCC. Também sustentamos que as plataformas devem reconhecer o verdadeiro custo de sua atividade e proporcionar um trabalho decente, com todas as condições que isso implica a seus trabalhadores.

2 O QUE É MODERAÇÃO DE CONTEÚDO COMERCIAL?

A moderação de conteúdo de espaços sociais e de informação *online* não é novidade. Desde o início da Internet, regras foram criadas para garantir padrões mínimos de interação nas comunidades *online*. A moderação nessas primeiras comunidades era feita pelos próprios usuários, de forma voluntária, geralmente indivíduos com valores semelhantes, unidos pelo interesse em vez da localização geográfica. Eles criaram suas políticas a partir da presunção de que o site seria usado, da mesma forma que fariam, por pessoas de boa-fé com definições semelhantes do que é, ou não, aceitável (GILLESPIE, 2018).

Na medida em que a *web* se desenvolveu e as plataformas de mídia social cresceram em tamanho e diversidade, essa abordagem, contudo, mostrou-se insuficiente. As plataformas agora têm de lidar com uma enorme quantidade de conteúdo e conciliar seus próprios interesses com os de usuários, governos e acionistas:

As políticas de moderação são, na melhor das hipóteses, compromissos razoáveis - entre usuários com valores e expectativas diferentes, bem como entre as demandas dos usuários e as demandas de lucro” (GILLESPIE, 2018, p. 13, tradução nossa).⁴

⁴ *Moderation policies are, at best, reasonable compromises—between users with different values and expectations, as well as between the demands of users and the demands of profit* (GILLESPIE, 2018, p. 13)

Assim, a moderação teve de ser organizada e executada em escala industrial, com profissionais pagos para fazer o trabalho em nome de grandes entidades comerciais. A moderação de conteúdo, portanto, tornou-se comercial (ROBERTS, 2019).

De acordo com Roberts (2019, p. 33, tradução nossa), “*moderação de conteúdo comercial [MCC] é a prática organizada de triagem de conteúdo gerado pelo usuário postado em sites, mídias sociais e outros meios de comunicação online*”⁵. Pode ocorrer antes ou depois do *upload* do material; neste caso, geralmente é:

[...] resultado de reclamações sobre o material por moderadores ou outros administradores do site, por partes externas (por exemplo, empresas alegando apropriação indevida de seu material), ou por outros usuários⁶ (ROBERTS, 2019, p. 33).

Com isso, **denunciaram** um determinado conteúdo devido à sua suposta inadequação às diretrizes da comunidade do site – também chamada de padrões, regras ou políticas, a depender de qual plataforma estamos falando.

Segundo sustenta Gillespie (2018), a MCC é uma tarefa que as plataformas assumem com relutância. A maioria delas tem um **mito fundador** semelhante: um pequeno grupo de pessoas unidas pelo sonho de um espaço verdadeiramente aberto, livre e democrático para todos, em qualquer lugar do mundo, se reunirem, interagirem e compartilharem em tempo real. A mera existência de um sistema invisível de moderadores decidindo o que pode, e não pode permanecer na *web* é um confronto direto dessa visão de uma comunidade livre, não mediada e imparcial de **iguais**.

⁵ [C]ommercial content moderation is the organized practice of screening user generated content posted to internet sites, social media, and other online outlets (ROBERTS, 2019, p. 33).

⁶ [...] result of complaints about material from site moderators or other site administrators, from external parties (for example, companies alleging misappropriation of material they own), or from other users [...] (ROBERTS, 2019, p. 33).

No entanto, as plataformas **devem**, em alguma medida, moderar. Não há nenhuma que não imponha qualquer tipo de regras. Parâmetros mínimos são necessárias, tanto para remover material ofensivo ou ilegal e proteger os usuários uns dos outros – de *bullying*, assédio, discriminação e violência extrema, por exemplo –, bem como em uma perspectiva econômica, para mantê-los usando o site e mostrar aos anunciantes e parceiros de negócios, novos usuários e ao público em geral que as plataformas são empresas seguras e confiáveis. Na verdade, “as plataformas são feitas pelo trabalho que entra na moderação de conteúdo, e não são plataformas sem ele” (GILLESPIE, 2018, p. 68, tradução nossa).⁷

2.1 Moderação de conteúdo: acessório ou elementar?

Como parte de sua imagem pública, as plataformas de mídia social gostam de se apresentar como meras intermediárias entre usuários que têm algo a dizer e usuários que possam querer ouvir e interagir com eles. Elas não produzem seu próprio conteúdo, mas exibem e divulgam aquele produzido por outros. A verdade, todavia, é que isso não é tudo o que as plataformas fazem.

Conforme argumenta Gillespie (2018, p. 28), existem, tradicionalmente, dois tipos de modelos de negócios de mídia e comunicação: “condutores de informação interpessoal”, como empresas de telefonia e correios, que são confiados para transportar informações em troca do pagamento de uma taxa, sem ter conhecimento ou interferir em seu conteúdo em qualquer forma; e “produtores de conteúdo midiático”, como rádio, cinema, televisão e jornais, cuja mercadoria é o material e entretenimento que produzem. As plataformas de mídia social, por outro lado, possuem características de ambos os modelos. Elas são responsáveis por conectar usuários e transportar informações para um público – uma pessoa, uma lista de amigos, ou de forma genérica, para todos os usuários que possam estar interessados nesse conteúdo

⁷ *Platforms are made by the work that goes into content moderation, and they are not platforms without it* (GILLESPIE, 2018, p. 68).

–. Como também prometem tornar a internet um lugar melhor ao organizar esse conteúdo, destacando, algoritmicamente, algumas postagens sobre outras, realizando curadoria, selecionando sugestões personalizadas para o usuário, facilitando a busca através de ferramentas de pesquisa, *hashtags* e assim por diante. Os usuários confiam suas informações às plataformas, que as utilizam com o fim de produzir um fluxo envolvente e constantemente atualizado para nos entreter e fazer com que continuemos a interagir com o *site* e produzamos mais dados; elas passaram de meras condutoras para constituir o conteúdo em si para a pessoa que o acessa.

Com efeito, a moderação é crucial para manter esse sistema; na verdade, ele não seria possível sem esse auxílio. Para manter as pessoas usando o site e os anunciantes pagando pelos anúncios, é necessário manter a tortura animal, decapitações e violência sexual fora da tela. A maioria dos usuários não se sentiriam seguros e dispostos a acessar a plataforma sem uma garantia mínima de que não encontrará conteúdo psicologicamente perturbador. Da mesma forma, as empresas também não ficariam felizes se seus anúncios fossem exibidos antes de uma execução terrorista. Além disso, exigências legais devem ser obedecidas. As pessoas esperam que as plataformas moderem, e continuam a utilizar o site com base nessa suposição, mesmo que não conscientemente. Portanto, a moderação é, em muitos aspectos, a própria mercadoria que as plataformas oferecem:

[...] se a Uber lucra conectando motoristas independentes com passageiros interessados, coordenando e assegurando sua interação, e cobrando uma taxa em troca, o Twitter faz o mesmo: reúne produtores de conteúdo independentes com ouvintes interessados, coordena sua interação e cobra uma taxa em troca - na forma de dados valiosos do usuário (GILLESPIE, 2018, p. 18, tradução nossa).

A moderação de conteúdo comercial é essencial para todas as plataformas. Não é uma atividade secundária ou acessória, mas

uma parte elementar e constitucional de seu modelo de negócios, necessária para proteger tanto a marca da empresa, quanto para garantir o cumprimento das leis existentes e para manter os usuários acessando e carregando conteúdo ativamente. Nesse sentido, se esse é um serviço tão importante para as comunidades *online*, por que a maioria das pessoas nunca ouviu falar dele? A resposta é, fundamentalmente, simples: porque as plataformas não querem que o façamos.

O fato de o conteúdo gerado pelos usuários ser moderado vai contra a visão utópica que as plataformas projetam de si: livres, imparciais e democráticas, uma zona de liberdade de expressão. Nossa experiência *online* foi projetada para termos a sensação de que é imediata, orgânica, não regulamentada, e o conteúdo existente para parecer um resultado natural, que simplesmente está lá (CARMI, 2019). Assim, a ideia de que o que é postado e permanece na rede não é resultado apenas das ações de usuários, mas o produto de um processo de filtragem executado por entidades comerciais e guiado por razões privadas e econômicas, e isso muda, substancialmente, nossa percepção sobre essa indústria. O que fica na rede é o resultado de uma equação meticulosamente elaborada, que leva em conta não apenas as **diretrizes da comunidade** do site, mas também o potencial para que um determinado conteúdo se torne viral e para que as pessoas o acessem e gerem mais dados. Além deles, são considerados sua recepção por uma audiência imaginada e os danos que poderia causar à imagem da empresa (GILLESPIE, 2018; ROBERTS, 2019).

Além disso, moderadores – as pessoas que fazem esse trabalho – não se encaixam na imagem construída do Vale do Silício como o sonho futurista de uma sociedade que se movimenta em torno da tecnologia e está na vanguarda da inovação, na qual cada pessoa – como programadores, cientistas, engenheiros, gurus do *marketing*, etc. – tem um trabalho ideal e desfruta de um alto padrão de vida. Como um ex-executivo do *Google* disse em certa ocasião: “Não há lugar nessa cultura para moderação

de conteúdo. É muito *nitty-gritty*⁸ (BARRET, 2020, p. 8, tradução nossa).

Nós, como usuários, não pensamos muito enquanto navegamos na *web* ou reportamos conteúdo que consideramos ofensivo ou inapropriado. Se o fizermos, tendemos a pensar na moderação como um serviço feito por inteligência artificial ou qualquer outro tipo de processo automatizado. Essas tecnologias, entretanto, não são capazes de lidar com uma porcentagem considerável do conteúdo carregado pelos usuários, especialmente imagens, vídeos e materiais escritos mais sutis. Na realidade, há humanos fazendo esse trabalho, responsáveis por manter o pior que a *web* tem a oferecer longe do público. Mais do que isso, são “seres humanos mal pagos que correm risco de *burnout*, dessensibilização e pior por causa da natureza de seu trabalho” (ROBERTS, 2019, p. 25, tradução nossa).⁹

As condições em que trabalham são deliberadamente escondidas pelas empresas de mídia social, com pouca informação tornada pública, protegidas por cláusulas de confidencialidade. De fato, o conjunto completo de regras adotadas pelas plataformas e os detalhes de como a moderação é feita são mantidos escondidos, sob o argumento de que a transparência total poderia permitir que usuários inescrupulosos burlem essas normas e que empresas rivais tenham acesso a informações confidenciais sobre práticas comerciais privadas (ROBERTS, 2019). É o momento que, como sociedade, começemos a fazer perguntas sobre como exatamente esse trabalho é feito.

⁸ *There's no place in that culture for content moderation. It's just too nitty-gritty* (BARRET, 2020, p.8). *Nitty-gritty* significa os elementos básicos de uma situação ou atividade, o âmago da questão, os pormenores. Foi utilizada nesse contexto para indicar uma atividade considerada muito **simples** ou **básica** dentro do Vale do Silício, que exige um tipo de trabalho **manual** e específico aquém do esperado por alguém que trabalha na indústria de tecnologia.

⁹ *[P]oorly paid human beings who risk burnout, desensitization, and worse because of the nature of their work* (ROBERTS, 2019, p. 25).

3 OS HUMANOS POR TRÁS DA TELA

Os moderadores de conteúdo comercial são “profissionais pagos para examinar o conteúdo carregado nos *sites* de mídia social em nome das empresas que solicitam a participação do usuário”.¹⁰ Eles são responsáveis por analisar a adequação do conteúdo *online* gerado pelos usuários, avaliar se ele viola as regras das plataformas e decidir se pode ficar ou se deve ser filtrado – um trabalho que deve ser feito em alguns segundos e centenas de vezes por dia (ROBERTS, 2019, p. 1, tradução nossa).

O que sabemos sobre esses trabalhadores é, principalmente, resultado de entrevistas concedidas a jornalistas e acadêmicos (CHEN, 2012, 2014; ROBERTS, 2019; NEWTON, 2019a, 2019b; KOEBLER; COX, 2018; BARRET, 2020) por alguns trabalhadores corajosos o suficiente para quebrar seus contratos de confidencialidade, ou do mínimo que as plataformas estiveram dispostas a revelar depois de muita e intensa pressão pública. Eles representam milhares de trabalhadores em todo o mundo: *Facebook* tem apenas 15.000 pessoas contratadas para fazer esse trabalho, a esmagadora maioria delas empregadas por empresas terceirizadas; *Google* conta com 10.000 moderadores para lidar com o material carregado no *YouTube* e outros dos seus produtos e cerca de 1.500 moderadores trabalham para o *Twitter*, uma empresa muito menor (BARRET, 2020). Os números, provavelmente, são muito maiores, considerando que essas empresas contratam *crowdworkers* de *sites* de microtrabalho para ajudar a processar a enorme quantidade de conteúdo carregado todos os dias – a um custo mínimo (ROBERTS, 2019).

Considerando-se que as plataformas não são obrigadas a seguir padrões mínimos ou diretrizes externas de funcionamento, bem como a MCC tratada como segredo empresarial, não há uma

¹⁰ *Commercial content moderators are professional people paid to screen content uploaded to the internet’s social media sites on behalf of the firms that solicit user participation* (ROBERTS, 2019, p. 1).

estrutura comum para esse processo, de forma que cada plataforma cria e usa seu próprio sistema. No entanto, alguns modos de operação comuns podem ser identificados. Segundo Chen (2014), é possível dizer que existem pelo menos dois tipos de moderadores de conteúdo: **moderadores ativos**, que filtram postagens em tempo real; e **moderadores reativos**, que filtram o conteúdo denunciado pelos usuários como inapropriado, sendo os mais comuns entre as maiores plataformas de mídia social.

A maioria das plataformas geralmente organizam sua cadeia de moderação em um modelo piramidal, no qual o conteúdo é, primeiramente, analisado por um grande grupo de moderadores e, posteriormente, é revisado por supervisores internos antes de ser repassado para moderadores da própria empresa de mídia social. Eles podem ser contratados diretamente pela plataforma, para trabalhar *in-house*, mas a maioria é contratada por empresas terceirizadas, seja como empregados, prestadores de serviço autônomos ou como mão de obra sob demanda, solicitada por meio de plataformas de *crowdwork*, como a *Mechanical Turk* da *Amazon* e *Upwork* (GILLESPIE, 2018; ROBERTS, 2019). No topo, uma equipe de políticas composta por advogados, profissionais de relações públicas e especialistas em gestão de crises criam as regras a serem reforçadas pelos moderadores (KOEHLER; COX, 2018).

No caso do *Facebook*, Chen (2012) aponta que há uma grande equipe na **linha de frente** – geralmente composta por *crowdworkers* e trabalhadores terceirizados – filtrando postagens, vídeos e imagens denunciados pelos usuários e tomando decisões rápidas sobre responder. Os moderadores não só têm que tomar a decisão **certa**, mas também indicar a razão correta, de acordo com as diretrizes do *Facebook*. Eles podem **confirmar** a denúncia (e excluir o conteúdo), **não confirmar** (deixar a postagem na rede) ou **escaloná-lo**, passando-o para os funcionários do *Facebook* para uma revisão posterior. Por razões de economia de tempo e privacidade, o conteúdo analisado, geralmente, não é acompanhado da identidade do usuário e pouco contexto é fornecido.

Pelo menos até o ano de 2019 (NEWTON, 2019b), a produtividade dos moderadores era medida por meio de uma pontuação de precisão (*accuracy score*), que deveria ser mantida em torno de 98%. De cada empresa terceirizada, o *Facebook* selecionaria, aleatoriamente, uma pequena quantidade das decisões semanais tomadas por um moderador para auditar. Inicialmente, essas decisões seriam revisadas por um supervisor, ou *quality assurance worker* (QA). Em seguida, os funcionários do *Facebook* auditariam um conjunto de decisões do QA e, a partir dessas deliberações coletivas, uma pontuação de precisão é gerada. De acordo com os moderadores, contudo, as notas dadas pelos QA não são totalmente confiáveis, pois são produto de um julgamento altamente subjetivo. Eles, frequentemente, conversam informalmente com o QA para rever determinada decisão e são capazes de obter muitos **pontos** de volta, fato que indica a inadequação desse sistema como a única maneira de avaliar a qualidade de seu trabalho (BARRET, 2020, p. 17).

3.1 Arranjos de trabalho em moderação de conteúdo comercial

Os moderadores de conteúdo comercial podem ser encontrados sob uma variedade de locais e condições de trabalho, a depender da necessidade da plataforma que busca esse serviço e, se for o caso, da empresa que o oferece. Eles constituem uma força de trabalho globalmente dispersa, que trabalha 24 horas por dia, sete dias por semana, cuja extensão pode ser regulada de acordo com a quantidade de conteúdo a ser filtrado. A partir dessas considerações, Roberts (2019) propõe uma taxonomia dos arranjos de trabalho referentes à MCC, de acordo com quatro contextos laborais básicos: interno (ou *in-house*), *boutique*, *call-center* e *sites* de microtrabalho.

A MCC feita **internamente** (ou *in-house*) é caracterizada pela prestação do serviço na sede física da empresa. Os moderadores podem ser contratados diretamente pela empresa de mídia social, na qualidade de empregados em tempo integral ou sob a forma de

contratos de prazo limitado ou de meio período; ou por empresas terceirizadas, mesmo que o trabalho seja prestado em local pertencente à tomadora do serviço. Normalmente, são oferecidos os melhores salários e condições de trabalho entre todos os arranjos (ROBERTS, 2019, p. 44).

As boutiques são

[...] empresas especializadas que oferecem gerenciamento da marca em mídias sociais, de forma geral, ou moderação de conteúdo, em específico, para outras empresas (ROBERTS, 2019, p. 45, tradução nossa).¹¹

Essas empresas são contratadas para gerenciar a presença *online* de seus clientes, que, geralmente, não são empresas de mídia digital ou tecnologia, mas que exigem certos tipos de moderação – como para suas contas em redes sociais, seções de comentários ou para imagens compartilhadas por usuários. Os trabalhadores variam de empregados permanentes em tempo integral a autônomos contratados para prestar um serviço delimitado.

O terceiro arranjo é definido pelo ambiente de *call center*, no sentido de

[...] empresas terceirizadas que oferecem um conjunto de serviços (frequentemente referidos como *Business Process Outsourcing*, ou BPO), entre os quais a moderação do conteúdo gerado pelo usuário é apenas um serviço, muitas vezes secundário (ROBERTS, 2019, p. 45, tradução nossa).¹²

É comum que essas empresas tenham uma variedade de clientes internacionais e estejam localizadas ao redor do mundo. Os trabalhadores, geralmente, são empregados pagos pela hora

¹¹ [...] *specialized firms that offer social media brand management, generally, or content moderation specifically, for other firms* (ROBERTS, 2019, p. 45).

¹² [...] *third- party companies offering a suite of services (frequently referred to as business process outsourcing, or BPO), among which moderation of user-generated content is but one service, and is often secondary* (ROBERTS, 2019, p. 45).

– um valor relativamente baixo, embora mais alto do que a maioria dos demais trabalhos em *call centers*–, e deles são exigidas competências linguísticas e culturais correspondentes aos dos usuários que produziram o conteúdo, “mas que muitas vezes estão em desacordo com sua própria auto expressão cultural e linguística local” (ROBERTS, 2019, p. 46, tradução nossa).¹³ Espera-se que o treinamento seja fornecido, embora a condução efetiva possa não ser suficiente para preparar moderadores para a verdadeira natureza de seu trabalho. A maioria dos relatos de más condições de trabalho feitas pelos moderadores referem-se a esses locais (CHEN, 2012; NEWTON, 2019a, 2019b; ROBERTS, 2019).

Por fim, a moderação pode ser executada através de **sites de microtrabalho**, como *Amazon MTurk* e *Upwork*, dependendo apenas do acesso a um computador e conexão com a internet. Os trabalhadores podem trabalhar de qualquer lugar, incluindo suas próprias casas, a qualquer momento. Sua relação com o tomador de serviço, cuja identidade é provavelmente desconhecida para eles, é por tarefa: eles são pagos por **peça** – ou por item visto ou exibido – muitas vezes recebendo na ordem de centavos por cada uma, não têm benefícios ou garantias e são incentivados a fazer o maior número possível de tarefas. Dentre todos os arranjos em MCC, este é aquele em que os trabalhadores estão mais vulneráveis (ROBERTS, 2019, p. 47).

3.2 Condições de trabalho e riscos ocupacionais

É esperado que os moderadores analisem centenas de peças de conteúdo todos os dias e tomem decisões rápidas a seu respeito. Muitas vezes, isso implica o contato com postagens psicologicamente perturbadoras, envolvendo racismo, abuso sexual, tortura animal e violência extrema. Levando-se em conta documentos internos vazados e relatos dos próprios moderadores, é possível ter uma visão do que isso significa na prática.

¹³ [...] *but which are often at odds with their own local cultural and linguistic self-expression* (ROBERTS, 2019, p. 46).

O conjunto de regras que norteiam as decisões relativas à moderação de conteúdo é composto por dois documentos públicos principais: os **termos de serviço**, um contrato que estabelece os termos, obrigações e condições sob os quais o usuário e a plataforma interagem; e as **diretrizes da comunidade**, que definem a visão da plataforma sobre o que é apropriado e o que não é, bem como seus princípios, proibições e justificativas.

Gillespie (2018) acredita que são

[...] declarações de política e princípio - em relação às quais os operadores da plataforma se debatem em alguns momentos e as ignoram em outros, implantadas quando são úteis e desviadas quando constroem¹⁴ (GILLESPIE, 2018, p. 31, tradução nossa).

Afinal, a MCC é dirigida por interesses econômicos e privados das plataformas de mídia social, em um processo intencionalmente obscurecido para evitar responsabilização pública (ROBERTS, 2019). Há, ainda, outro conjunto de diretrizes internas que explica em termos mais específicos e com base em situações concretas como os moderadores devem agir. Elas são escondidas do público e só são conhecidas por conta de documentos vazados.

Em uma reportagem de 2012, Chen trouxe à tona algumas destas diretrizes: moderadores devem confirmar uma **denúncia** sempre que for “retratada a mutilação de pessoas, de animais ou humanos decapitados, desmembrados, carbonizados ou queimando”.¹⁵ Discursos violentos, como “adoro ouvir crânios rachar”¹⁶ também devem ser apagados. “Cabeças esmagadas, membros, etc. são ok desde que interiores não estejam expostos”;¹⁷ “feridas

¹⁴ *[S]tatements of both policy and principle—struggled over by the platform operators at some moments and ignored at others, deployed when they are helpful and sidestepped when they are constraining* (GILLESPIE, 2018, p. 31).

¹⁵ *Depicting the mutilation of people or animals, or decapitated, dismembered, charred or burning humans* (CHEN, 2012).

¹⁶ *I love hearing skulls crack* (CHEN, 2012).

¹⁷ *Crushed heads, limbs etc are OK as long as no insides are showing* (CHEN, 2012).

profundas podem ser mostradas”, assim como “sangue excessivo” (CHEN, 2012, tradução nossa).¹⁸

Newton (2019b, tradução nossa) também compartilhou as experiências pessoais que alguns moderadores tiveram enquanto trabalharam na indústria: pessoas colocando “fogos de artifício acesos na boca de cachorros”, outras “mutilando os genitais de um rato vivo, e cortando o rosto de um gato com um machado”.¹⁹ “Pessoas autistas devem ser esterilizadas”,²⁰ embora pareça ofensivo, permanece na rede devido ao fato de que o autismo não é uma “característica protegida”, como raça e gênero, não violando a política da plataforma (NEWTON, 2019a, tradução nossa).

Essa constatação é verdadeira para os trabalhadores de sites de microtrabalho. Em um artigo que destacou a realidade das pessoas que trabalham na *Amazon Mechanical Turk (MTurk)*, Reese e Heath apontaram que os *turkers* não têm conhecimento do conteúdo da imagem de antemão. É possível optar por sair depois de exibido o conteúdo, mas nenhum pagamento é recebido, tendo o tempo perdido no processo. Um dos moderadores declarou: “Havia uma cesta cheia de cabeças. Foi o que vi há alguns meses. O cara pegando fogo, eu tive que marcar aquele vídeo. Foi tipo 10 centavos por foto” (REESE; HEATH, 2016, tradução nossa).²¹

É como um ex-diretor de assuntos públicos e proteção à criança da *Microsoft* declarou: “É um trabalho nojento”²² (ROBERTS,

¹⁸ Deep flesh wounds are ok to show; excessive blood is ok to show (CHEN, 2012)

¹⁹ [...] *p*ut lit fireworks in dogs' mouths. [He watched people] mutilate the genitals of a live mouse, and chop off a cat's face with a hatchet (CHEN, 2012).

²⁰ *Autistic people should be sterilized* (CHEN, 2012).

²¹ *There was a basket full of heads.' That's what I saw just a few months ago. The guy on fire, I had to tag that video. It was like 10 cents a photo* (REESE; HEATH, 2016).

²² Do original *It's a yucky job*.

2019, p. 38, tradução nossa). De fato, tais condições de trabalho representam um desafio em termos de saúde e segurança ocupacional em muitas formas, com um impacto particularmente intenso sobre a saúde mental dos moderadores. A própria natureza do trabalho de MCC é perigosa, especialmente em termos psicossociais. O contato com esse tipo de conteúdo perturbador e os problemas relativos à organização do trabalho podem ser uma fonte de estresse, ansiedade, *Burnout* e, em casos extremos, Transtorno de Estresse Pós-Traumático – TEPT; tanto é que o *Facebook* concordou, recentemente, em pagar US\$ 52 milhões a moderadores atuais e antigos para compensá-los por problemas de saúde mental desenvolvidos no trabalho, reconhecendo tacitamente a natureza perigosa da atividade (BARRET, 2020, p.12).

Vale ressaltar que os moderadores que trabalham em *sites* de microtrabalho, como *crowdworkers*, estão, além disso, sujeitos a uma série de outros riscos, como consequência da precariedade desse arranjo de trabalho na *gig economy* (ORGANIZAÇÃO, 2018). Restringiremos nossa análise para focar em alguns dos principais fatores de risco ocupacional, especificamente relacionados ao trabalho com MCC.

De acordo com Carmi (2019), o trabalho de MCC é organizado para se assemelhar a uma linha de produção, na qual os moderadores devem processar centenas de postagens por dia, o mais rápido possível, e manter altas metas de precisão, sustentando um ritmo constante durante longos turnos. Em algumas empresas, os trabalhadores são notificados a cada poucos minutos das suas **notas de qualidade**, as quais mostram o número de erros que cometeram e determinam se os contratos de curto prazo serão rescindidos ou prorrogados (MURGIA, 2020). Essa intensa carga de trabalho é reflexo da **alta demanda por velocidade e precisão** e pode ter seu impacto sobre a saúde mental e física dos trabalhadores, aumentando o estresse e a ansiedade relacionados ao trabalho.

Os moderadores também lutam para manter as metas de precisão devido ao fato de que “eles têm várias fontes de verdade a

considerar” (NEWTON, 2019a, tradução nossa).²³ Como já discutido, o conjunto de regras que regem a moderação é composto por documentos públicos e privados; aqueles escondidos do público oferecem detalhes mais específicos sobre questões complexas e são atualizados com maior frequência. Essas diretrizes em constante mudança podem ser confusas para os moderadores e resultar em diferentes decisões sobre um determinado tema feitas na mesma semana (ou dia), o que é ainda mais prejudicado pelo fato de que as diretrizes não oficiais, guiadas pelos interesses econômicos das plataformas, podem se mostrar contraditórias, quando comparadas às públicas, gerando *falta de clareza e incerteza sobre seu trabalho*.

Além disso, *conflitos morais, éticos e culturais* podem surgir entre as crenças pessoais dos moderadores e as decisões que são obrigados a tomar em relação à moderação, especialmente se considerarmos que muitas empresas terceirizadas estão localizadas em países com culturas tão diferentes quanto os Estados Unidos, as Filipinas e a Índia. Para muitos trabalhadores, a MCC normalmente exige

[...] assumir ou incorporar um conjunto de valores, e fazer julgamentos que podem variar dos seus próprios códigos morais e valores pessoais e culturais (ROBERTS, 2016, p. 1, tradução nossa).²⁴

A **insegurança no emprego** é outro fator de risco a ser considerado: um grande número de trabalhadores é contratado sob a forma de contratos de meio período ou temporários de baixa remuneração, alternando entre momentos de emprego e desemprego. Os que trabalham em sites de microtrabalho, em particular, recebem por tarefa (na base de centavos) e precisam filtrar uma

²³ [T]hey have several sources of truth to consider (NEWTON, 2019a).

²⁴ [T]aking on or embodying a set of values, and making judgments that may vary from their own moral codes and personal and cultural values (ROBERTS, 2016, p. 1).

grande quantidade de conteúdo na incerteza se serão capazes de receber o mínimo para viver. Mesmo a maioria dos empregados formais e em tempo integral não duram mais de um ano ou dois no trabalho (BARRET, 2020, p. 18). Exige-se dos moderadores que revejam o conteúdo psicologicamente perturbador e mantenham uma alta pontuação de precisão de 98%; em muitos casos, se não se demitirem, eles serão despedidos sob o pretexto de não atender à taxa exigida.

A **exposição contínua a conteúdo emocionalmente perturbador**, como violência, pornografia e discurso de ódio, também pode levar ao aumento do estresse, ansiedade e, em casos mais extremos, a TSPT. A dessensibilização emocional é outra consequência possível desse trabalho, bem como a adoção das visões conspiratórias e discriminatórias sustentadas no conteúdo que os moderadores devem filtrar, como a negação do Holocausto, devido à exposição excessiva e contínua (NEWTON, 2019a).

Além disso, o **isolamento**, tanto **social** quanto **geográfico**, tem impactos sobre a saúde e a organização coletiva dos trabalhadores. Os acordos de confidencialidade e o tratamento de informações relacionadas ao que fazem como segredo de empresa impedem que eles desabafem e se abram, mesmo com entes queridos, levando a um aumento dos sentimentos de isolamento e ansiedade. Além disso, o distanciamento do trabalho do resto dos trabalhadores da empresa, literal e figurativamente (enquanto trabalhadores de baixo *status* na indústria de tecnologia), e a variedade de arranjos de trabalho e títulos sob os quais se encontram dificultam a identificação deles como um grupo e a sua organização coletiva com o fim de exigir e lutar por melhores condições de trabalho.

Por fim, a **disponibilidade de acompanhamento psicológico e tratamento por profissionais no local de trabalho** ou, no mínimo, apoio financeiro para que os trabalhadores o busquem em outro lugar é uma necessidade. Para muitas pessoas envolvidas nessa atividade, no entanto, o acompanhamento não é

adequadamente prestado por profissionais preparados e qualificados, ou simplesmente não é fornecido, especialmente no caso de *crowdworkers*. Mesmo quando é disponibilizado, os moderadores geralmente não possuem garantia da continuidade do apoio psicológico após o fim de seus contratos, sendo obrigados a suportar as consequências psíquicas derivadas do seu trabalho por conta própria (ROBERTS, 2019).

4 RUMO AO TRABALHO DECENTE

A moderação de conteúdo comercial não é um trabalho fácil. É exigido dos moderadores que trabalhem como máquinas, mantendo um ritmo regular, rápido e constante. Esse tipo de trabalho é deliberadamente escondido do público e projetado para parecer um processo automatizado – ou como uma inteligência “artificial” (PRASSL, 2018).

Na realidade, os moderadores são submetidos a uma série de riscos ocupacionais, particularmente psicossociais, e as empresas que os contratam reconhecem isso. Em dezembro de 2019 e janeiro de 2020, uma das empresas terceirizadas contratadas pelo *Facebook* e outras plataformas de mídia social para MCC na Europa pediu aos seus funcionários que assinassem um formulário afirmando

Entendo que o conteúdo que será avaliado pode ser perturbador. É possível que a revisão desse conteúdo possa afetar minha saúde mental, e isso pode até levar a transtorno de estresse pós-traumático (TEPT)²⁵ (MURGIA, 2020).

O formulário também afirmava que os funcionários têm acesso a “*coaches* de bem-estar” para apoio à saúde mental, embora “o

²⁵ *I understand the content I will be reviewing may be disturbing. It is possible that reviewing such content may impact my mental health, and it could even lead to post-traumatic stress disorder (PTSD)* (MURGIA, 2020).

coach de bem-estar não seja médico e não possa diagnosticar ou tratar transtornos mentais” (MURGIA, 2020, tradução nossa).²⁶

Segundo Barret (2020), alguns terceirizados começaram a trabalhar, inicialmente, no escritório da própria plataforma e, posteriormente, foram transferidos para um dos escritórios da empresa que os contrataram. Eles relataram, entre outras coisas, que, no escritório da plataforma, tiveram acesso a moderadores mais experientes para consultar demandas razoáveis de trabalho, mais liberdade para gerenciar seus dias e mais tempo para tomar decisões seguras sobre o conteúdo exibido. Isso indica que os moderadores que trabalham *in-house* estão mais bem equipados para fazer seu trabalho, e que uma organização do trabalho que leve em conta sua saúde e segurança satisfatoriamente é possível.

Dessa forma, chegou o momento de as plataformas serem responsabilizadas por suas práticas e tratem a moderação de conteúdo comercial como parte elementar de seu modelo de negócio, suportando o custo real dessa atividade. Parâmetros mínimos devem ser adotados coletivamente, por todas as plataformas, a fim de assegurar condições de trabalho decentes. De fato, iniciativas conjuntas nessa indústria para proteger a saúde e a segurança dos moderadores não são sem precedentes. Um grupo chamado *The Technology Coalition*, criado para coordenar os esforços das grandes empresas de tecnologia para combater a exploração sexual infantil *online*, publicou, em janeiro de 2015, um documento chamado “Guia de Resiliência de Funcionários para o Manuseio de Imagens de Abuso Sexual Infantil”. Entre outras sugestões, o guia diz que “[I]mitar a quantidade de tempo que os empregados são expostos [à pornografia infantil] é fundamental” e recomenda que as empresas adquiram o “consentimento informado” de seus empregados, incluindo “fornecer um nível adequado de informação para que o empregado entenda o que essa função implica” (ARSHT; ETCOVITCH, 2018).

²⁶ [T]he wellness coach is not a medical doctor and cannot diagnose or treat mental disorders (MURGIA, 2020).

Assim, acreditamos que a MCC seria melhor executada, supervisionada e controlada se as plataformas acabassem com a terceirização – incluindo o uso de plataformas de microtrabalho –, gradualmente incorporando moderadores como empregados em tempo integral na plataforma para a qual agora trabalham indiretamente, com salários e suporte adequados. Três razões principais sustentam a ideia de que a terceirização deve ser abandonada: ela reforça uma visão periférica da moderação, prejudicando que investimentos e atenção devida sejam focados na MCC; ela marginaliza moderadores como trabalhadores de baixo *status* e remuneração em relação a outros trabalhadores das empresas de mídia social; e sustenta um ambiente caótico que impacta a tomada de decisão dos revisores (BARRET, 2020).

Não é necessário que as plataformas empreguem todos os seus moderadores em sua sede do Vale do Silício, sendo possível, e preferível, manter um sistema globalmente distribuído. Isso exigirá investimentos significativos, mas refletirá o verdadeiro custo de seus negócios. No mínimo, a moderação através de *sites* de microtrabalho deve ser abandonada, ou deve haver alguma garantia de que as tarefas de moderação postadas nesses *sites* tenham um conteúdo **mais leve**, a fim de proteger a saúde e a segurança dos trabalhadores.

Além disso, o aumento do número de moderadores permitiria um sistema de revezamento, possibilitando sua alternância entre atribuições com conteúdo menos perturbador. Isso também permitiria que os moderadores tenham tempo para consultar regularmente os psicólogos/psiquiatras da empresa, criando um histórico da sua saúde mental continuamente atualizado e permitindo a identificação precoce de sintomas de TSPT ou outros efeitos colaterais da revisão de conteúdo (BARRET, 2020).

Como vimos, nem todas as empresas de mídia social supervisionam diretamente seus moderadores terceirizados. Auditorias são executadas, mas sua abordagem atual não tem mantido resultados desejáveis. As plataformas devem, portanto, contratar

um supervisor de conteúdo para políticas e execução de moderação de conteúdo, a fim de centralizar as operações de MCC, permitindo simplificar processos-chave e ressaltando sua importância interna e externa (BARRET, 2020). Além disso, as plataformas devem patrocinar pesquisas sobre os riscos à saúde derivados do trabalho com moderação de conteúdo, que atualmente não são totalmente conhecidos.

É, ainda, de extrema importância fornecer a todos os moderadores cuidados médicos de alta qualidade no local de trabalho. Isso inclui a incorporação de padrões mínimos de aconselhamento psicológico no local de trabalho, nos contratos para serviços de moderação e o seu monitoramento eficaz, bem como o aumento dos programas de bem-estar, atualmente oferecidos pelas terceirizadas, disponibilizando psiquiatras bem qualificados para consulta. As empresas de mídia social também devem fornecer planos de saúde gratuitos, ou fortemente subsidiados, que cubram doenças mentais por algum período de tempo após o trabalhador deixar o emprego, ou até que ele obtenha cobertura por outro empregador.

Outrossim, as plataformas devem exigir que as empresas de moderação terceirizadas informem claramente aos candidatos a natureza do trabalho, requisitos e riscos inerentes a ele, bem como comprometer com mais transparência em suas atividades. Isso não quer dizer que a moderação de conteúdo deva ser completamente pública ou sem defeitos. Isso significa que as empresas de mídia social devem se responsabilizar pela forma como a MCC é organizada, por como esse trabalho é feito e por quem, pela maneira como pretendem tornar esse serviço eficaz e transparente, ao mesmo tempo em que zelam pelos direitos trabalhistas. Mais clareza significa melhor gestão, supervisão mais eficaz, mais estudos sendo feitos e um resultado mais confiável (GILLESPIE, 2018, p. 82). Finalmente, compromissos em todo o setor semelhantes aos da Coalizão Tecnológica são essenciais para atingir padrões mínimos e universais em MCC.

5 CONCLUSÃO

A demanda pelas plataformas de mídias sociais está aumentando, assim como pela moderação de conteúdo. Como vimos, as empresas buscam maximizar seus lucros, terceirizando essa atividade, mas tal abordagem tem se mostrado inadequada. Mesmo o uso de inteligência artificial, responsável por sinalizar uma grande parte do conteúdo gerado pelos usuários, na maioria das vezes não dispensa a necessidade de moderadores humanos para confirmarem a **denúncia** (BARRET, 2020).

Como sociedade, demos às empresas privadas o poder de tomar decisões sobre moderação sem a devida responsabilização e transparência. Conforme argumenta Roberts (2019), as tecnologias não são neutras

[...] e portanto não são «naturalmente» benignas ou sem impacto. Pelo contrário, são, por sua natureza, construções sociotécnicas, ao mesmo tempo em que refletem seus criadores e também são criadas a serviço de algo ou alguém²⁷ (ROBERTS, 2019, p. 30, tradução nossa).

Padrões mínimos devem ser criados, adotados e reforçados por todas as empresas de mídia social. As plataformas devem ter a obrigação de fornecer mais transparência sobre suas práticas de moderação e sobre as condições sob as quais o trabalho dos moderadores é desempenhado. Nesse sentido, especialmente porque os arranjos de trabalho aplicados pelas empresas de mídia social são intencionalmente obscurecidos e fragmentados, com o intuito de evitar maior responsabilidade quanto ao tema saúde e segurança no trabalho, por exemplo. Além disso, a natureza da MCC não é compatível com o uso de *sites* de microtrabalho e outras práticas de terceirização. As plataformas devem,

²⁷ [...] and therefore are not “naturally” benign or without impact. On the contrary, they are, by their nature as sociotechnical constructions, at once reflective of their creators while also being created in the service of something or someone (ROBERTS, 2019, p. 30).

pois, assumir a responsabilidade sobre seu modelo de negócio e ser demandadas por usuários, anunciantes, legisladores e a sociedade civil como um todo, com o intuito de assegurar condições de trabalho decentes e um modelo de moderação de conteúdo comercial centrado na pessoa humana.

Concluimos, assim, pela importância nuclear desse tipo de trabalho para a própria viabilidade do modelo de negócio de plataformas de mídia social, nada obstante a insistente invisibilidade técnica e social que tem recaído sobre os trabalhadores que realizam moderação de conteúdo digital. Temos, desse modo, em resumo e sempre na ótica dos direitos humanos, que plataformas digitais devem fornecer maior transparência sobre as condições de trabalho de seus moderadores de conteúdo, havendo de garantir-lhes, igualmente, plena sintonia com o vetor ético-jurídico do trabalho decente.

Só assim teremos uma revolução digital, verdadeiramente guiada pela dignidade do ser humano e não pela técnica como simples fator de indução de lucro.

REFERÊNCIAS

ARSHT, Andrew; ETCOVITCH, Daniel. The human cost of online content Moderation. **JOLT, Journal of Law and Technology Digest**, Cambridge, Harvard Law School, 2 mar. 2018. Disponível em: <https://jolt.law.harvard.edu/digest/the-human-cost-of-online-content-moderation>. Acesso em: 24 jun. 2021.

BARRET, Paul M. Who moderates the social media giants? A call to end outsourcing. **Center for Business and Human Rights**, New York University, jun. 2020. Disponível em: <https://www.stern.nyu.edu/experience-stern/faculty-research/who-moderates-social-media-giants-call-end-outsourcing>. Acesso em: 2 jul. 2021.

CARMI, Elinor. The hidden listeners: regulating the line from telephone operators to content moderators. **International**

Journal of Communication, Los Angeles, v. 13, p. 440-458, 2019. Disponível em: <https://ijoc.org/index.php/ijoc/article/view/8588>. Acesso em: 2 jul. 2021.

CHEN, Adrian. Inside Facebook's outsourced anti-porn and gore brigade, where 'camel toes' are more offensive than crushed heads. **Gawker**, 16 fev. 2012. Disponível em: <https://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads>. Acesso em: 24 jun. 2021.

CHEN, Adrian. The laborers who keep dick pics and beheadings out of your Facebook feed. **Wired**, 23 out. 2014. Disponível em: <https://www.wired.com/2014/10/content-moderation/>. Acesso em: 24 jun. 2021.

GILLESPIE, Tarleton. **Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media**. New Haven: Yale University Press, 2018. *E-book*.

KOEBLER, Jason; COX, Joseph. The impossible job: inside Facebook's struggle to moderate two billion people. **Vice**, 23 ago. 2018. Disponível em: <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works>. Acesso em: 24 jun. 2021.

MURGIA, Madhumita. Facebook content moderators required to sign PTSD forms. **Financial Times**, Londres, 26 jan. 2020. Disponível em: <https://www.ft.com/content/98aad2f0-3ec9-11ea-a01a-bae547046735>. Acesso em: 24 jun. 2021.

NEWTON, Casey. The trauma floor: the secret lives of Facebook moderators in America. **The Verge**, 25 fev. 2019a. Disponível em: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>. Acesso em: 24 jun. 2021.

NEWTON, Casey. Bodies in seats: at Facebook's worst-performing content moderation site in North America, one

contractor has died, and others say they fear for their lives. **The Verge**, 19 jun. 2019b. Disponível em: <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa>. Acesso em: 24 jun. 2021.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO (OIT). **Digital labour platforms and the future of work: towards decent work in the online world**. Genebra, ILO, 2018. Disponível em: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_645337.pdf. Acesso em: 2 jul. 2021.

PRASSL, Jeremias. **Humans as a service: the promise and perils of work in the gig economy**. Oxford: Oxford University Press, 2018.

REESE, Hope; HEATH, Nick. Inside Amazon's clickworker platform: how half a million people are being paid pennies to train AI. **Tech Republic**, 16 dez. 2016, Disponível em: <https://www.techrepublic.com/article/inside-amazons-clickworker-platform-how-half-a-million-people-are-training-ai-for-pennies-per-task/>. Acesso em: 24 jun. 2021.

ROBERTS, Sarah T. **Behind the screen: content moderation in the shadows of social media**. New Haven: Yale University Press, 2019. *E-book*.

ROBERTS, Sarah T. Commercial content moderation: digital laborers' dirty work. **Media Studies Publications**, n. 12, 2016. Disponível em: <https://ir.lib.uwo.ca/commpub/12/>. Acesso em: 2 jul. 2021.