

O Lado Negro do TikTok: A Perspetiva da Moderação

The Dark Side of TikTok: The Moderation Perspective

El Lado Oscuro de TikTok: La Perspectiva de la Moderación


Daniela Tavares • Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal • tavaresdanielasousa@gmail.com

 <https://orcid.org/0009-0007-4073-0664>

Inês Saraiva • Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal • saraivaines44@gmail.com

 <https://orcid.org/0009-0009-5087-5140>

José Leão • Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal • zemleao.535@gmail.com

 <https://orcid.org/0009-0009-4054-4460>

Resumo Este artigo analisa os mecanismos de moderação de conteúdo no TikTok e as perceções dos utilizadores portugueses sobre esses processos. Numa primeira parte, revê-se a literatura sobre redes sociais online e moderação, distinguindo-se a intervenção humana da moderação algorítmica e discutindo-se as respetivas vantagens, limitações e impactos psicossociais. Na componente empírica, aplicou-se um questionário online a 85 participantes, maioritariamente jovens adultos. Os resultados mostram um conhecimento limitado sobre o trabalho dos moderadores e o funcionamento dos algoritmos. Conclui-se que a eficácia da moderação no TikTok exige um equilíbrio entre sistemas automáticos e revisão humana. Sugere-se que, numa investigação futura mais aprofundada, sejam incluídos métodos qualitativos, para clarificar o impacto da moderação e melhorar práticas de governação de plataformas digitais.

Palavras-chave: TikTok, Moderação de Conteúdos, Redes Sociais Online, Algoritmos.

Abstract This article analyses the content moderation mechanisms on TikTok and Portuguese users' perceptions of these processes. The first part reviews the literature on online social networks and moderation, distinguishing human intervention from algorithmic moderation and discussing the respective advantages, limitations and psychosocial impacts. In the empirical component, an online questionnaire was administered to 85 participants, mostly young adults. The results show limited knowledge of the moderators' work and how the algorithms work. It is concluded that effective moderation on TikTok requires a balance between automatic systems and human review. It is suggested that further research should include qualitative methods to clarify the impact of moderation and improve governance practices on digital platforms.

Keywords: TikTok, Content Moderation, Online Social Networks, Algorithms.

Daniela Tavares é licenciada em Ciências da Comunicação pela Universidade de Trás-os-Montes e Alto Douro desde 2023. Realizou um estágio curricular no jornal *Público* na secção do Local, no âmbito do Mestrado em Ciências da Comunicação na Universidade de Trás-os-Montes e Alto Douro, no ramo de Jornalismo.

Inês Saraiva é licenciada em Línguas e Relações Empresariais pela Universidade de Trás-os-Montes e Alto Douro, desde 2022. Estagiária nos gabinetes de comunicação e de eventos na Câmara Municipal de Vila Pouca de Aguiar, é estudante do Mestrado em Ciências

da Comunicação da Universidade de Trás-os-Montes e Alto Douro, no ramo de Relações Públicas e Publicidade.

José Leão é licenciado em Ciências da Comunicação pela Universidade de Trás-os-Montes e Alto Douro, desde 2023. Gestor de redes sociais na agência de comunicação Grandes Planos, é estudante do Mestrado em Ciências da Comunicação da Universidade de Trás-os-Montes e Alto Douro, no ramo de Publicidade e Relações Públicas.

SOPCOM
ASSOCIAÇÃO PORTUGUESA DE
CIÊNCIAS DA COMUNICAÇÃO

**Rev
Com**

Revista Comunicando

vol. 14, núm. Special, Esp. e025008,
2025

Associação Portuguesa de Ciências da
Comunicação,
Portugal

ISSN-E: 2182-4037

ISSN-L: 2182-4037

directorcomunicando@sopcom.pt

Recepción: 30 Julio 2024

Aprobación: 01 Febrero 2025

Publicación: 17 Septiembre 2025

Resumen Este artículo analiza los mecanismos de moderación de contenidos en TikTok y las percepciones de los usuarios portugueses sobre estos procesos. La primera parte revisa la literatura sobre redes sociales en línea y moderación, distinguiendo la intervención humana de la moderación algorítmica y discutiendo las respectivas ventajas, limitaciones e impactos psicosociales. En el componente empírico, se administró un cuestionario en línea a 85 participantes, en su mayoría adultos jóvenes. Los resultados muestran un conocimiento limitado del trabajo de los moderadores y del funcionamiento de los algoritmos. Se concluye que una moderación eficaz en TikTok requiere un equilibrio entre los sistemas automáticos y la revisión humana. Se sugiere que la investigación futura incluya métodos cualitativos para aclarar el impacto de la moderación y mejorar las prácticas de gobernanza en las plataformas digitales.

Palabras clave: TikTok, Moderación de Contenidos, Redes Sociales en Línea, Algoritmos.

1. Introdução^[1]

Em Portugal, em 2023, já eram mais de três milhões de pessoas a utilizar o TikTok (TikTok, 2024). Entre os utilizadores, destacavam-se sobretudo aqueles entre os 18 e os 24 anos. E, em 2024, “quase metade dos jovens portugueses entre os 18 e os 24 anos” estava nesta rede (Anjos, 2024, para. 1). Estes dados evidenciam o papel central do TikTok no ecossistema digital juvenil, reforçando a pertinência de estudar os mecanismos de moderação nesta plataforma e as perceções dos utilizadores. Será que os jovens conhecem o que está por detrás da rede social que ocupa os lugares cimeiros das descargas da App Store ou Play Store? Sabem como é feita a moderação dos seus conteúdos e o que essa moderação acarreta? É precisamente em torno destas questões que se desenvolve este artigo.

Inspirados por uma reportagem realizada pela jornalista da Renascença Daniela Espírito Santo (2022), intitulada como *Ser Moderador do TikTok É Ver o “Pior da Humanidade” por 900€*, partimos para a revisão de literatura. Deste modo, este artigo apresenta uma componente teórica onde se analisa e discute o conceito de redes sociais online, fazendo uma contextualização sobre o TikTok. Após este esclarecimento, entramos no foco da questão: perceber os vários aspetos da moderação — a sua definição, os vários tipos de moderação existentes, as vantagens e desvantagens, de que forma são aplicados e a perspetiva dos criadores de conteúdo e dos utilizadores.

No que diz respeito à parte empírica, as questões de investigação passam por saber se os utilizadores conhecem e estão familiarizados com o conceito de moderação. De forma a esclarecer este tema, aplicamos um questionário a uma amostra por conveniência com um total de 85 respostas de jovens universitários. A escolha por um público jovem universitário alinha-se com os objetivos da investigação, que visam compreender as perceções e o grau de conhecimento sobre os processos de moderação numa faixa etária particularmente ativa nas redes sociais online. Este segmento representa também o grupo mais suscetível à influência de algoritmos, criadores de conteúdo e dinâmicas de viralização, tornando-se assim um foco relevante para estudar a eficácia e os limites da moderação no TikTok.

O objetivo deste estudo é duplo: por um lado, compreender a perceção dos utilizadores do TikTok sobre os mecanismos de moderação de conteúdo; por outro, caracterizar os seus hábitos de utilização das redes sociais online, explorando se esses padrões podem estar relacionados com o grau de conhecimento ou opinião sobre a moderação. Esta abordagem visa fornecer um enquadramento mais completo da experiência dos utilizadores com as plataformas digitais.

Pretende-se, de uma forma inicial, fornecer um contributo para a área das Ciências da Comunicação com um tema ainda não muito divulgado.

2. Revisão de Literatura

2.1. Redes Sociais Online: Origem e Conceitos

As redes sociais online ocupam, hoje, uma centralidade sem precedentes nas sociedades desenvolvidas, conforme defendem os coordenadores do livro *Redes Sociais — Para uma Compreensão Multidisciplinar da Sociedade* (Fialho et al., 2018). São uma das componentes mais importantes na estrutura das relações sociais entre pessoas, organizações e grupos, através das quais se partilham valores, expectativas, interesses e uma imensidão de fluxos que as colocam num nível de complexidade elevado (Fialho et al., 2018). Assim, o próprio conceito de redes sociais online não é consensual, mas Fialho et al. (2018) apontam-no como um conjunto de relações pessoais definidas por organizações e sistemas lógicos de relacionamento social.

A própria rede social TikTok define-se como o “principal destino para vídeos curtos em dispositivos móveis”, referindo que a sua “missão é inspirar a criatividade e trazer alegria” (TikTok, s.d., para. 1). Em 2020, a rede contava com 500.000 utilizadores portugueses (Barroso, 2020). E, em 2023, já eram 3,3 milhões os utilizadores (TikTok, 2024), o que mostra o seu crescimento.

O TikTok é uma aplicação gratuita, que possibilita aos seus utilizadores assistir, criar e partilhar vídeos curtos (com a duração máxima de um minuto), normalmente com músicas que estão na moda como pano de fundo (Barroso, 2020).

Por ser um dos temas da ordem do dia, vários investigadores centram a sua pesquisa nesta área. Por exemplo, os cientistas da Universidade Zhejiang, na China, foram tentar perceber o porquê desta aplicação viciar tanto os seus utilizadores. Como explica a revista *Exame*, os pesquisadores perceberam que áreas do cérebro ligadas ao sistema de recompensa são ativadas pelos vídeos da rede, produzindo de forma rápida uma sensação de prazer e satisfação no organismo (Agência O Globo, 2022).

2.2. Moderação de Conteúdo

Karabulut et al. (2022) defendem que as plataformas digitais, nomeadamente as redes sociais online, comportam milhões de utilizadores que produzem e consomem bilhões de conteúdos todos os dias. Desta forma, existe a necessidade de filtrar e moderar o conteúdo que é apresentado nestas plataformas.

Por se tratar de um processo complexo, são várias as possíveis abordagens que procuram conceptualizar esta temática. De acor-

do com Zeng e Kaye (2022), a moderação de conteúdo é vista como um fenómeno de elevada complexidade sociotécnica e define-se como o processo no qual as plataformas esquematizam as trocas de informação de forma a decidirem e filtrarem o que é considerado apropriado de acordo com normas legais, culturais e políticas individuais.

Ainda neste registo, Grimmelmann (2015, como citado em Riedl et al., 2022) relaciona a moderação de conteúdo com mecanismos de governação, que têm como objetivo estruturar a participação no seio de uma comunidade com a finalidade de facilitar a cooperação e evitar abusos.

No entanto, autores como Steen et al. (2023) sublinham sobretudo o cariz opaco da moderação de conteúdo e da forma como plataformas como o TikTok esquematizam e desenvolvem os ambientes no seio da própria rede. Common (2020) vai mais longe ao referir que a abordagem adotada pela maioria das plataformas não está suficientemente desenvolvida e necessita de ser revista, sob pena de causar sérios problemas relacionados com recursos humanos.

2.3. Moderação Humana e Através de Inteligência Artificial: Vantagens e Desvantagens

Existem vários autores que apontam um conjunto de advertências para a moderação de conteúdo ao nível da moderação humana e da inteligência artificial. No entanto, importa tecer uma consideração relevante que deve ser tida em conta antes de analisar as vantagens e desvantagens de cada método de moderação. Zeng e Kaye (2022) referem que a moderação no TikTok tem um cariz especial, na medida em que esta rede social se gere sobre o princípio da “viralização”, o que, segundo os autores, contribui para a complexidade da gestão dos conteúdos. Isto explica-se na medida em que o próprio algoritmo privilegia vídeos que têm tendências para “viralizar”, dificultando em certa parte a moderação humana.

A criação de conteúdos virais é o principal meio de visibilidade e socialidade. Na tentativa de alcançar uma viralidade intencional ou “acidental”, os criadores do TikTok podem tentar ultrapassar os limites, produzindo conteúdos erráticos ou anómalos, bem como tendências virais prejudiciais. A imprevisibilidade e a incerteza da viralidade tornam difícil para os moderadores humanos de conteúdos manter as tendências, especialmente as tendências nocivas ou perigosas, sob controlo. (Zeng & Kaye, 2022, p. 80)

Tendo em conta a posição dos autores no que respeita à moderação feita por intervenção humana, Karabulut et al. (2022) apontam ainda como desvantagem a dificuldade que existe na análise exponencial de informação, visto que o ser humano, quando comparado com a tecnologia, tem uma capacidade limitada de análise

se intensiva de informação. Para além desse fator, uma análise a priori de todos os conteúdos levaria à perda da experiência de utilização, em particular no caso do TikTok:

a moderação de conteúdos de milhares de milhões de imagens por humanos, não é viável em muitos aspetos. Os seres humanos não conseguem atingir uma velocidade aceitável para analisar imagens em tempo real, e um atraso notável entre o carregamento da fotografia e a publicação causa uma experiência de utilizador menos atrativa. (Karabulut et al., 2022, p. 4440)

Ainda no ponto de vista da moderação humana, Karabulut et al. (2022) atribuem especial importância aos danos mentais que este tipo de moderação pode causar nos moderadores: “a exposição prolongada a conteúdos perturbadores pode causar problemas de saúde mental nos anotadores humanos” (p. 4440).

No que diz respeito à moderação tendo por base a inteligência artificial, nomeadamente os algoritmos, Zeng e Kaye (2022) sublinham a dificuldade acrescida que existe por parte da tecnologia em criar processos automatizados capazes de analisar vídeo e contextos, o que se evidencia como um ponto a favor da moderação humana: “os moderadores humanos desempenham um papel crucial na avaliação da adequação dos vídeos nas plataformas digitais, incluindo o TikTok” (p. 80).

Ainda neste registo, Steen et al. (2023) apontaram que a inteligência artificial muitas vezes não é capaz de interpretar contextos e, por isso, remove conteúdos de forma incoerente: a inteligência artificial “não compreende a utilização contextual de uma palavra num vídeo e, por conseguinte, interpreta-a incorretamente como uma violação das diretrizes da comunidade” (p. 9). Para além disso, os autores apontam também como problemas a aleatoriedade, a falta de precisão e o facto de muitas vezes a inteligência artificial atuar de forma enviesada (Steen et al., 2023).

2.4. De que Forma os Mecanismos de Moderação São Aplicados no TikTok?

Após terem sido articulados os diferentes tipos de moderação de conteúdo em conformidade com as vantagens e desvantagens que lhes são inerentes, importa agora perceber de que forma é que o processo de moderação é aplicado na prática. Ou seja, procura-se entender como é que estes mecanismos atuam e com que frequência intervêm no TikTok.

A própria rede social, de acordo com as informações publicadas no site, esclarece que os conteúdos publicados passam, numa primeira fase, pela inteligência artificial da plataforma e só depois são revistos por moderadores de conteúdo, no caso de o algoritmo detetar alguma inconformidade no que respeita às normas da aplicação (Steen et al., 2023). Não obstante, Steen et al. (2023)

apontam que o TikTok recorre com maior frequência a sistemas de inteligência artificial e procura desenvolver a capacidade de aprendizagem do algoritmo por forma a moderar os conteúdos publicados.

2.5. A Moderação Sob o Ponto de Vista dos Criadores de Conteúdo

Neste ponto importa compreender a forma como os mecanismos de moderação de conteúdo são entendidos e desconstruídos por parte dos criadores de conteúdo do TikTok. Esta é uma etapa importante na medida em que transparece uma visão prática da aplicabilidade destes mecanismos por forma a perceber a dimensão real da moderação de conteúdo.

Steen et al. (2023), através de um caso prático articulado com a recolha de dados baseada em várias entrevistas, procuraram perceber a perceção dos criadores de conteúdo relativamente à influência que a moderação de conteúdo tem no TikTok. Em particular, verificaram a utilização de *algospeak* com vista a manipular as estratégias utilizadas pelo algoritmo e, assim, conseguirem que os vários conteúdos fossem aceites pela inteligência artificial da plataforma. Desta forma, de acordo com os autores, a utilização de *algospeak* baseia-se sobretudo na utilização de termos ou emojis em substituição de palavras menos próprias, termos esses cujos referentes são entendidos pelo comum de utilizadores da plataforma.

De uma forma geral, os autores concluíram que a utilização desta técnica era recorrente e que muitas vezes os criadores de conteúdo recorriam a este método por forma a conseguirem que os conteúdos não fossem removidos. Isto explica-se na medida em que os criadores se apercebiam de que o que publicavam, apesar de não violar especificamente as normas da aplicação, era entendido pela inteligência artificial como uma ameaça e, consequentemente, bloqueado.

Em geral, descobrimos que os participantes adotaram o *algospeak* depois de verem no TikTok ou inventaram o seu próprio *algospeak* para evitar consequências por publicarem tópicos que consideravam secretamente indesejados, mas que não violavam explicitamente as diretrizes da comunidade do TikTok. Os participantes previram, em grande medida, a forma como o algoritmo do TikTok iria ler o seu conteúdo de vídeo e utilizaram o *algospeak* para evitar, principalmente, a moderação algorítmica injustificada do conteúdo. (Steen et al., 2023, p. 7)

Assim, na perspetiva dos criadores de conteúdos, o TikTok apresenta um sistema de moderação pouco eficaz e enviesado. Para além disso, segundo Steen et al. (2023), muitos participantes revelaram que o sistema de moderação é opaco e que só com processos de tentativa e erro é que os criadores de conteúdo conseguiram perceber a melhor forma de agir no que visa a publicação de

vídeos. “Todos os participantes eram utilizadores experientes do TikTok, mas a opacidade dos procedimentos de moderação de conteúdos da plataforma significava frequentemente que estavam apenas a ‘especular’ sobre o que a aplicação censurava” (Steen et al., 2023, p. 8).

3. Metodologia

Uma vez esclarecido, com sustentação teórica, o tema da moderação de conteúdos, falta então perceber de que forma os utilizadores estão (ou não) informados acerca do tema. Para além de explorar o conhecimento dos utilizadores sobre a moderação de conteúdo, o questionário procurou caracterizar os seus hábitos de uso das redes sociais online. Este bloco de perguntas permite contextualizar as respostas sobre moderação, tendo em conta a exposição, o tempo passado nas plataformas e a influência de criadores de conteúdo, fatores que podem impactar as perceções sobre a moderação.

Relativamente ao questionário, este instrumento foi construído em dois blocos, contando com 85 respostas, considerando-se uma amostra por conveniência. O primeiro bloco procurou caracterizar os hábitos dos utilizadores nas redes sociais online — como frequência de uso, finalidades e relação com influenciadores digitais —, por forma a oferecer um enquadramento comportamental. O segundo bloco abordou especificamente a perceção dos mecanismos de moderação, a familiaridade com o papel dos moderadores e as experiências diretas com esse tipo de intervenção. Das questões apresentadas, apenas uma era de resposta aberta, de forma a não obter resultados demasiado inconclusivos.

O questionário esteve disponível online durante uma semana, em dezembro de 2023, e foi partilhado, essencialmente, com alunos da Universidade de Trás-os-Montes e Alto Douro (UTAD), através da Associação Académica dessa universidade. Assim, a amostra utilizada neste estudo foi composta por estudantes do ensino superior da UTAD, sendo uma amostra não probabilística por conveniência. A escolha deste grupo justifica-se não apenas pela acessibilidade direta — uma vez que os autores também são estudantes da UTAD —, mas também por razões estratégicas. O público jovem é o que apresenta maior presença e envolvimento na plataforma TikTok, tal como demonstram estudos recentes (Anjos, 2024). Assim, optou-se por inquirir principalmente estudantes do ensino superior, por se tratar de um grupo representativo dos utilizadores mais expostos às dinâmicas de moderação de conteúdo nesta plataforma. Esta opção permitiu obter respostas contextualizadas a partir de experiências reais com a rede social em estudo.

4. Análise e Discussão de Resultados

Tendo em consideração que a utilização destas plataformas tem como públicos centrais os mais jovens, procurámos direccionar o questionário para estas faixas etárias. Neste sentido, a maioria dos inquiridos enquadram-se na faixa dos 18 aos 22 anos (69) e 13 têm entre 23 e 28 anos, sendo que os restantes têm idades superior às mencionadas. Este dado confirma a adequação da amostra ao foco do estudo, uma vez que os jovens constituem o grupo mais exposto e ativo nas dinâmicas da plataforma TikTok — sendo, por isso, também os mais afetados (ou potencialmente afetados) pelos mecanismos de moderação.

No que diz respeito ao nível de formação^[2], o espectro é bastante abrangente. No entanto, destacam-se maioritariamente alunos do 1.º ano de licenciatura (24), do 1.º ano de mestrado (19) e do 3.º ano de licenciatura (17).

Relativamente ao tempo despendido nas redes sociais online por dia, 39 inquiridos afirmam passar entre três e seis horas nas plataformas, o mesmo número de pessoas que afirma passar entre uma a três horas do tempo nas redes sociais, por dia. No entanto, importa destacar que seis inquiridos afirmam passar entre seis e 12 horas por dia nas redes, o que apesar de parecer irrisório no contexto da amostra pode ser, porém, alarmante. Este comportamento intensivo de uso reforça a importância de se estudar a moderação de conteúdo, uma vez que o grau de exposição pode aumentar o impacto (positivo ou negativo) das intervenções algorítmicas e humanas sobre os utilizadores.

Ao nível da utilização das diferentes redes sociais online, com a questão “que redes sociais utiliza com maior frequência?”, procurámos perceber de que forma é que o TikTok tem, ou não, grande preponderância no quotidiano dos inquiridos. Como era expectável, esta rede social enquadra-se num pódio de plataformas mais utilizadas, sendo superada apenas pelo Instagram. Neste sentido, 66 inquiridos afirmam utilizar o TikTok com bastante frequência. De salientar que plataformas como o Facebook e o Messenger, que há uns anos vigoravam neste espectro, aparecem na realidade atual como pouco utilizadas pela amostra, com apenas nove e dez inquiridos, respetivamente, a afirmar utilizarem estas plataformas com elevada frequência.

Ao serem questionados relativamente ao propósito de utilização das redes sociais, as respostas seguem um padrão bastante decalcado, uma vez que a questão foi colocada de forma aos inquiridos conseguirem selecionar múltiplas opções. A ideia de fazer amigos já não tem um peso relevante para os inquiridos (cinco) e a amostra transparece sobretudo a ideia de utilização destas plataformas com vista a passar o tempo, que corresponde à resposta dada por 82 inquiridos. Este aspeto é importante na medida em que redes sociais online como o TikTok focam-se, sobretudo, no

entretenimento e procuram cativar e “prender” os utilizadores com base no algoritmo. Não obstante, a segunda resposta mais selecionada, com 56 escolhas, é a utilização das redes com vista a uma interação social esporádica e casual.

Relativamente à questão “normalmente, utiliza as redes sociais quando...”, seguiu a mesma linha que a questão anterior, em forma de caixa de verificação, as respostas mais registadas foram: “tenho tempo livre para isso” (81); “à noite, antes de ir dormir” (72) e, com menos expressão, “enquanto estou na escola/universidade/trabalho” (35). Neste sentido, embora a esmagadora maioria tenha respondido que normalmente utiliza as redes sociais quando tem tempo livre, percebe-se que também são usadas noutras situações do quotidiano. A utilização destas plataformas em situações sociais e durante as refeições surge de forma notória, na medida em que dez inquiridos afirmam fazê-lo nestas circunstâncias. Mais uma vez, apesar de ser um valor em minoria, é um dado relevante e sobre o qual importa tecer reflexões. Em conformidade, a utilização das redes sociais em situações de ocupação laboral ou universitária (35) é um dado que realça ainda mais esta perspetiva. Além disso, a conexão que existe entre os utilizadores e as redes transparece a necessidade que os inquiridos têm em consultar estas plataformas à noite, imediatamente antes de se deitarem.

No entanto, quando os inquiridos são confrontados com esta realidade de forma direta, nomeadamente através da questão colocada imediatamente a seguir — “verifica todas as suas redes sociais antes de ir dormir?” —, os resultados entram numa certa contradição com as respostas apresentadas na questão anterior. Assim, apenas 29 dos respondentes afirmam consultar todas as redes sociais, 35 responderam que não o faz em todas as situações e 20 afirmam que não o fazem de todo. Uma pessoa não respondeu. Através de uma breve análise aos valores facilmente se pode concluir que inquiridos que afirmavam que normalmente recorriam às redes antes de ir dormir agora apontam para um menor uso. Este fator pode ser explicado pela exposição direta à questão tendo em conta o que é tido como correto pelo senso comum.

No que diz respeito à questão “sente que as redes sociais e o conteúdo lá apresentado altera o seu humor?”, as respostas dadas seguem uma tendência neutra, na medida em que cerca de metade dos inquiridos selecionou os níveis dois e três na escala de Likert que abrangia valores de zero a cinco. Apesar de tudo, 15 pessoas selecionaram o nível quatro, revelando assim que as redes sociais online motivavam alterações significativas de humor nestes inquiridos. Não obstante, importa também destacar que 11 respondentes afirmaram não ser influenciados por estas plataformas ao nível do humor. Em contraste, apenas uma pessoa afirmou que as redes sociais online alteravam completamente o hu-

mor — nível cinco. De uma forma geral, se dividirmos a escala em duas partes uniformes (zero a dois e três a cinco) verificamos uma tendência negativa que se traduz em alterações mais baixas de humor nos respondentes por influência das redes sociais, no entanto, a diferença entre ambas as partes não é significativa.

Quando os inquiridos são confrontados com a questão sobre o facto de seguirem ou não influenciadores digitais, 69 responderam de forma positiva, e apenas 16 afirmaram não seguir qualquer influenciador. Assim, no que diz respeito a esta questão, as expectativas entraram em conformidade com os resultados obtidos, confirmando o papel que os influenciadores digitais ocupam na sociedade, principalmente nas camadas mais jovens.

De seguida, foi colocada uma outra pergunta: “caso tenha respondido ‘sim’ na questão anterior, sente que estes apresentam uma influência sobre si (forma de vestir, alimentação, etc.)?”, na qual 28 dos inquiridos afirmam sentir-se influenciados e 25 responderam não ter a certeza, optando por seleccionar a opção “talvez”. De todos os inquiridos, apenas uma franja de 16 das pessoas disseram não sofrer qualquer tipo de influência. Assim, ao articularmos as respostas afirmativas em conjunto com os que respondem “talvez”, chegamos à conclusão de que grande parte dos inquiridos sente algum tipo de influência sobre si por parte dos influenciadores digitais. Este aspeto é relevante para a presente investigação, pois mostra que os utilizadores nem sempre estão plenamente conscientes da forma como o conteúdo (moderado ou não) os afeta, o que pode interferir na forma como percebem a função da moderação e a sua eficácia.

A questão que se segue está relacionada em concreto com a rede social que é o tema principal desta investigação e que diz respeito à utilização do TikTok por parte dos inquiridos.

Numa primeira análise, procuramos perceber se a amostra segue os padrões da sociedade em que estamos inseridos e, de certa forma, podemos concluir que, tal como esperávamos, o TikTok é utilizado pela maioria dos inquiridos (69).

No que diz respeito à percepção sobre a moderação por parte dos inquiridos foram desenhadas duas questões de resposta dicotómica. À primeira questão “sabe o que significa moderar um comentário ou um post nas redes sociais?”, 49 dos inquiridos responderam de forma positiva. A segunda questão procura perceber as percepções dos indivíduos, questionando se “alguma vez sentiu o efeito da moderação em alguma rede social?”. Através das respostas percebe-se de forma ainda mais clara que a grande maioria dos respondentes nunca sofreu qualquer ação por parte de sistemas de moderação de conteúdo, ou, se sofreu, não o sabe. Assim, 35 dos inquiridos afirmaram nunca ter sentido qualquer efeito da moderação nas redes sociais online, 32 não sabem e ape-

nas 18 já tiveram algum tipo de contacto com uma experiência de moderação nas redes sociais online.

No seguimento destas questões, procurou-se medir os conhecimentos dos inquiridos sobre o trabalho de um moderador de redes sociais online em particular.

Neste sentido, à questão “conhece o trabalho de um moderador?” apenas nove declararam conhecer o trabalho destes profissionais e 35 disseram não ter qualquer tipo de conhecimento sobre o tema. Não obstante, importa realçar que 41 dos inquiridos afirmaram que já tinham ouvido falar sobre o tema, mas que não tinham grande conhecimento. Com base na análise desta questão, e em conformidade com as perguntas anteriores, podemos concluir que a moderação das redes sociais online ainda é um território desconhecido por uma parte dos utilizadores. Assim, alguns inquiridos não têm qualquer conhecimento sobre o tema e muitos dizem já ter ouvido falar, mas que, no entanto, não sabem muito bem em que consiste e como esta questão se articula e desenvolve nas redes sociais online.

A última questão do inquérito procurou recolher possíveis contributos dos inquiridos relativamente aos conteúdos que não deviam ser publicados nas redes sociais online. Não foi possível identificar um conjunto organizado de sugestões, uma vez que as respostas foram relativamente dispersas, sendo as mais frequentes: conteúdos de incitação à violência; pornografia infantil; conteúdos racistas e xenófobos; conteúdos que incitem o ódio e a discriminação; conteúdos sensíveis que podem ser considerados impróprios ou desconfortáveis para terceiros; conteúdos que levem à desinformação.

No entanto, há também quem defenda que todos os conteúdos devem estar presentes nas redes sociais online, mas de forma devidamente moderada. Ainda neste registo, um respondente afirma também que todos os conteúdos são importantes, mas que reconhece que as políticas de moderação não são eficazes no que diz respeito à filtração de conteúdos. Neste sentido, existe uma relação entre as conclusões retiradas por Steen et al. (2023) e esta percepção.

Destaca-se ainda a influência das redes sociais online nas pessoas e na forma de estar e agir socialmente, sendo que um respondente afirma que conteúdos que definam como as pessoas se devem vestir, falar e estar levam à definição de rótulos que por sua vez originam problemas do foro psicológico, como é o caso de estereótipos sociais em relação ao corpo feminino e à forma de vestir.

Estes resultados, aliados ao tempo elevado de uso das redes e à influência percebida por parte de criadores de conteúdo, reforçam a necessidade de compreender como os mecanismos de moderação afetam diferentes perfis de utilizadores, em especial os

mais jovens, mais ativos e mais suscetíveis à lógica de viralização da plataforma.

5. Conclusões

Numa época em que existem biliões de utilizadores de plataformas digitais, que influenciam e se deixam influenciar por grande parte do conteúdo que consomem no online todos os dias, é importante refletir sobre a moderação de conteúdos e de como é feita esta gestão.

Ao longo deste estudo, demos ênfase à moderação da rede social TikTok onde identificamos a moderação humana, a moderação por parte da inteligência artificial, mas também as estratégias dos próprios utilizadores face a estes mecanismos. Esta rede social é bastante conhecida pela “viralização” de determinados vídeos que aí são publicados e, daí, advém uma preocupação acrescida com a moderação que aqui se desenvolve, uma vez que este processo desempenha um papel crucial na manutenção de um ambiente seguro e respeitoso.

Neste mesmo ponto é importante sublinhar que uma das maiores desvantagens desta moderação de conteúdo recai sobre o ser humano, ou seja, a moderação humana. Quem faz a moderação de conteúdos acaba por ter de lidar com vídeos bastante sensíveis e perturbadores para a sua saúde mental, pelo próprio facto de estar exposto à análise destes mesmos conteúdos.

Conseguimos também perceber que a moderação da rede social do TikTok é, primeiramente, analisada pela inteligência artificial e, seguidamente, passa pelos próprios moderadores de conteúdo para uma revisão mais detalhada.

Em resposta à questão de investigação se existe conhecimento por parte dos utilizadores, parece que a grande maioria não tem muito conhecimento sobre estes mecanismos.

Assim, de uma forma resumida, os principais resultados deste estudo revelam que a maioria dos jovens utilizadores do TikTok não possui qualquer conhecimento sobre os mecanismos de moderação ou do papel dos moderadores, não tendo assim consciência que existe essa moderação, sugerindo uma opacidade nos critérios aplicados. Existe ainda um elevado grau de utilização das plataformas digitais, o que acaba por tornar mais alarmante o facto de desconhecerem os sistemas que regulam os conteúdos. Porém, os inquiridos são capazes de reconhecer a existência de conteúdos impróprios e de manifestar consenso na importância de políticas de moderação mais eficazes e claras, reconhecendo também que tanto a moderação automática como a humana acabam por ter fragilidades, saliento os riscos iminentes para a saúde mental no caso da moderação humana.

Para terminar, importa salientar que, à data a que se conclui este estudo, ainda se sentia uma considerável dificuldade em encontrar referências bibliográficas acerca da temática. Apesar de a rede social estar na ordem do dia e nas *trends* diárias, a comunidade científica ainda não se havia debruçado de forma significativa sobre o tema, principalmente no que diz respeito aos investigadores portugueses. Praticamente a totalidade de artigos científicos e estudos utilizados neste artigo estão em inglês, perante a presença de um menor número de referências portuguesas. Assim, como sugestões futuras, sugere-se um maior aprofundamento acerca desta rede social, não só no caso da moderação humana, mas da inteligência artificial, uma vez que se torna cada vez mais uma ferramenta nestes processos. Como metodologia de investigação, talvez a opção por grupos de foco possa esclarecer e clarificar mais alguns conceitos. Tendo em conta as condições que reuníamos, o questionário foi o método mais eficaz para retirar estas pequenas conclusões que esperamos que sejam um contributo para a área das Ciências Humanas e Sociais.

Referências

- Agência O Globo. (2022, 5 de abril). *Como o TikTok atua no cérebro e vicia jovens em seus vídeos curtos*. Exame. <https://exame.com/ciencia/como-o-tiktok-atua-no-cerebro-de-jovens-com-vidEOS-curtos-e-personalizados/>
- Anjos, M. (2024, 10 de janeiro). TikTok e os jovens: A máquina de fazer extremistas. *Visão*. <https://visao.pt/atualidade/sociedade/2024-01-10-tiktok-e-os-jovens-a-maquina-de-fazer-extremistas/>
- Barroso, P. (2020, 6 de agosto). O que é, afinal, o TikTok? Respondemos a todas as suas dúvidas. *Visão Júnior*. <https://visao.pt/visaojunior/familia/2020-08-06-o-que-e-afinal-o-tiktok-respondemos-a-todas-as-suas-duvidas/>
- Common, M. F. (2020). Fear the Reaper: How content moderation rules are enforced on social media. *International Review of Law, Computers & Technology*, 34(2), 126–152. <https://doi.org/10.1080/13600869.2020.1733762>
- Fialho, J., Saragoça, J., Baltazar, M., & Santos, M. O. (Eds.) (2018). *Redes sociais. Para uma compreensão multidisciplinar da sociedade*. Edições Sílabo.
- Karabulut, D., Ozcinar, C., & Anbarjafari, G. (2022). Automatic content moderation on social media. *Multimedia Tools and Applications*, 82, 4439–4463. <https://doi.org/10.1007/s11042-022-11968-3>
- Riedl, M., Whipple, K., & Wallace, R. (2022). Antecedents of support for social media content moderation and platform regulation: The role of presumed effects on self and others. *Information, Communication & Society*, 25(11), 1632–1649. <https://doi.org/10.1080/1369118X.2021.1874040>
- Santo, D. (2022, 12 de setembro). *Ser moderador do TikTok é ver o “pior da humanidade” por 900€*. Rádio Renascença. <https://rr.sapo.pt/fotoreportagem/vida/2022/09/13/ser-moderador-do-tiktok-e-ver-o-pior-da-humanidade-por-900/295550/>
- Steen, E., Yurechko, K., & Klug, D. (2023). You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media + Society*, 1–17. <https://doi.org/10.1177/20563051231194586>
- TikTok. (s.d.). *Sobre o TikTok*. <https://www.tiktok.com/about?lang=pt-BR>
- TikTok. (2024, 13 de fevereiro). *Em Portugal, mais de três milhões de pessoas utilizam o TikTok para partilhar as suas histórias, paixões e interesses*. <https://newsroom.tiktok.com/pt-pt/em-portugal-mais-de-tres-milhoes-de-pessoas-utilizam-o-tiktok>
- Zeng, J., & Kaye, D. B. V. (2022). From content moderation to visibility moderation: A case study of platform governance on TikTok. *Policy & Internet*, 14, 79–95. <https://doi.org/10.1002/poi3.287>

Notas

1 Este trabalho integra o número especial “Comunicação Pop: Abordagens Teóricas e Perspetivas Metodológicas”. Esta secção é um espaço de publicação de trabalhos de alunos do Mestrado em Ciências da Comunicação da Universidade de Trás-os-Montes e Alto Douro (2023/2024). As secções especiais ampliam a resposta ao desígnio da *Revista Comunicando* enquanto espaço de aprendizagem.

2 Importa referir que 25 participantes não indicaram o seu nível de formação.

* Contribuições

Daniela Tavares: concetualização, análise formal, investigação, metodologia, redação do rascunho original, redação — revisão e edição

Inês Saraiva: concetualização, análise formal, investigação, metodologia, redação do rascunho original, redação — revisão e edição

José Leão: concetualização, análise formal, investigação, metodologia, redação do rascunho original, redação — revisão e edição

Información adicional

redalyc-journal-id: 7747



Disponible en:

<https://www.redalyc.org/articulo.oa?id=774783198003>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en redalyc.org

Sistema de Información Científica Redalyc
Red de revistas científicas de Acceso Abierto diamante
Infraestructura abierta no comercial propiedad de la
academia

Daniela Tavares, Inês Saraiva, José Leão

**O Lado Negro do TikTok: A Perspetiva da Moderação
The Dark Side of TikTok: The Moderation Perspective
El Lado Oscuro de TikTok: La Perspectiva de la
Moderación**

Revista Comunicando

vol. 14, núm. Special, Esp. e025008, 2025

Associação Portuguesa de Ciências da Comunicação,
Portugal

diretorcomunicando@sopcom.pt

ISSN-E: 2182-4037

ISSN-L: 2182-4037

DOI: <https://doi.org/10.58050/comunicando.v14iSpecial.403>



CC BY 4.0 LEGAL CODE

Licencia Creative Commons Atribución 4.0 Internacional.