Comunicação e Sociedade, vol. 50, 2026, e026016
https://doi.org/10.17231/comsoc.50(2026).7084

Materialidade da Governação do Ódio no Instagram: Conflitos em Torno da Imigração em Portugal

Dalvacir Xavier de Oliveira Andrade
Centro de Estudos de Comunicação e Sociedade, Instituto de Ciências Sociais, Universidade do Minho, Braga, Portugal/ Programa de Pós-Graduação em Comunicação e Cultura Contemporâneas, Faculdade de Comunicação, Universidade Federal da Bahia, Salvador, Brasil

Resumo

Este artigo investiga a materialidade da governação do ódio no Instagram, refletindo sobre o papel desta arquitetura sociotécnica nos conflitos em torno da imigração em Portugal. O estudo insere-se num cenário de transformações na governação migratória, ascensão da extrema-direita e centralidade das plataformas digitais na disputa por visibilidade e participação. O objetivo central é interrogar como a definição e o controlo do ódio anti-imigrante são infraestruturalmente organizados pela plataforma. Para tal, adota-se a metodologia neomaterialista da comunicação associal (Lemos, 2020) e realiza-se uma análise material-discursiva (Barad, 2007; Kitchin & Dodge, 2011). O foco empírico é dado aos padrões da comunidade da Meta, especialmente nas secções "Conduta de ódio", "Violência e incitação" e "Bullying e assédio", assim como nas interfaces de denúncia disponíveis aos utilizadores do Instagram. A análise demonstra, por um lado, o reconhecimento formal de "refugiados, migrantes, imigrantes e pessoas que procurem asilo" como grupo protegido contra ataques graves. Por outro lado, o estudo evidencia hierarquias de dano que privilegiam a violência física imediata em detrimento de formas de exclusão política, económica e social, bem como zonas de ambiguidade produzidas por critérios de intenção, contexto, sátira e classificação interna das denúncias. Argumenta-se que a governação do ódio no Instagram funciona como mediador sociotécnico que simultaneamente promete proteção e redistribui responsabilidades de segurança pelos utilizadores, tornando o ódio anti-imigrante parcialmente visível nas políticas e pouco legível na interface. O artigo conclui discutindo as implicações desses achados para as condições de visibilidade, segurança e contestação de pessoas imigrantes no espaço digital.

Palavras-chave
governação de plataformas, discurso de ódio, imigração em Portugal, Instagram, conflito comunicacional

Materiality of Hate Governance on Instagram: Immigration-Related Conflicts in Portugal

Abstract

This article investigates the materiality of Instagram's hate governance, reflecting on the role of this sociotechnical architecture in conflicts surrounding immigration in Portugal. The study is situated within a context of transformations in migration governance, the rise of the far right, and the centrality of digital platforms in struggles over visibility and participation. Its primary objective is to examine how the definition and regulation of anti-immigrant hate are organised through the platform's infrastructure. To this end, the study adopts the neomaterialist methodology of asocial communication (Lemos, 2020) and conducts a material-discursive analysis (Barad, 2007; Kitchin & Dodge, 2011). The empirical focus is on Meta's community standards,

particularly the sections "Hateful conduct", "Violence and incitement", and "Bullying and harassment", as well as the reporting interfaces available to Instagram users. The analysis identifies the formal recognition of "refugees, migrants, immigrants, and asylum seekers" as a protected group against severe attacks. However, it also uncovers hierarchies of harm that prioritise immediate physical violence over political, economic, and social exclusion. Additionally, the study highlights areas of ambiguity arising from criteria related to intent, context, satire, and the platform's internal classification of reports. The article argues that Instagram's hate governance operates as a sociotechnical mediator, offering protection while shifting responsibility for safety onto users. This approach renders anti-immigrant hate only partially visible in policy and minimally discernible through the interface. The conclusion discusses the implications of these findings for the visibility, safety, and contestation experienced by immigrants within digital environments.

Keywords
platform governance, hate speech, immigration in Portugal, Instagram, communication conflict

1. Introdução

Nas primeiras décadas do século XXI, o aumento da mobilidade humana tem sido acompanhado por uma intensificação das controvérsias em torno das migrações. Crises económicas e políticas, desigualdades estruturais, conflitos armados e emergências climáticas contribuem para a multiplicação de deslocações forçadas e voluntárias, enquanto Estados e instituições disputam narrativas sobre fronteiras, segurança e direitos.

No caso português, estas dinâmicas inscrevem-se em sucessivas transformações na governação da imigração, marcadas pela extinção do Serviço de Estrangeiros e Fronteiras e pela criação da Agência para a Integração, Migrações e Asilo, em 2023, que passou a concentrar as competências administrativas em matéria de migração no país. Desde o início do seu funcionamento, têm-se multiplicado relatos de atrasos, falhas de comunicação e dificuldades de acesso a direitos, demandando intervenções de órgãos como a Provedoria de Justiça, a Ordem dos Advogados e organizações da sociedade civil. Em paralelo, mudanças de governo e revisões da legislação migratória ocorrem num contexto de ascensão eleitoral da extrema-direita, cuja agenda enfatiza o endurecimento das políticas de imigração, o discurso anti-imigração e a associação da imigração a questões como criminalidade e insegurança.

É neste cenário que as plataformas digitais se tornam centrais na forma como a migração é imaginada, planeada, experienciada e narrada. Ambientes como YouTube, Instagram e TikTok funcionam simultaneamente como fontes de informação, sociabilidade e visibilidade para pessoas imigrantes, onde circulam conselhos práticos, testemunhos do quotidiano e narrativas de recomeço. Todavia, essa visibilidade é permeada por conflitos comunicacionais marcados por ódio e moderação de conteúdo, que incidem sobre o direito de existir, trabalhar e falar enquanto imigrante no espaço público digital.

No contexto de "plataformização" da vida social (van Dijck et al., 2018) e no regime Plataformização, Datificação e Performatividade Algorítmica (PDPA; Lemos, 2021), estes conflitos são inseparáveis das infraestruturas sociotécnicas que organizam a sua

circulação. A moderação não é um aspeto secundário, mas um elemento constitutivo do que as plataformas são (Gillespie, 2018). Assim, além de observar o que se diz sobre imigração, importa interrogar a materialidade infraestrutural que condiciona essa circulação: as categorias através das quais as plataformas definem "ódio", os mecanismos de denúncia e filtragem, os fluxos de moderação humana e algorítmica, as sanções aplicadas e as *affordances* que convidam, limitam ou redirecionam determinados usos.

É a partir deste enquadramento que este artigo trata o conflito como controvérsia sociotécnica, na qual humanos e não humanos (utilizadores, algoritmos, interfaces, políticas de moderação, mecanismos de denúncia) coproduzem formas de visibilidade e vulnerabilidade (Latour, 1994, 2012). Propõe-se, assim, um mapeamento da materialidade da governação do ódio no Instagram, com ênfase nos dispositivos infraestruturais que enquadram conteúdos sobre imigração. Adota-se a metodologia neomaterialista da comunicação associal (Lemos, 2020) e realiza-se uma análise material-discursiva (Barad, 2007; Kitchin & Dodge, 2011) com o objetivo de analisar como o Instagram configura os modos de controlo do ódio anti-imigrante, a partir das suas políticas de comunidade e dos mecanismos de denúncia disponibilizados a utilizadores, refletindo sobre o papel dessa governação na disseminação de conflitos em torno da imigração em Portugal.

Portugal funciona, neste artigo, como enquadramento empírico-político da controvérsia, e não como escala exclusiva da governação técnica do Instagram. As políticas e interfaces analisadas são corporativas e transnacionais; contudo, são acionadas em conflitos localmente situados, num momento em que a imigração se tornou um tema de forte disputa institucional, mediática e partidária no país. É nessa articulação entre arquitetura global de plataforma e conflito nacionalmente situado que o caso português torna particularmente críticas as ambiguidades da moderação do ódio anti-imigrante.

O artigo está organizado em cinco secções, além desta Introdução. A Secção 2 apresenta o enquadramento teórico sobre conflitos, migração e governação de plataformas. A Secção 3 explicita a abordagem metodológica e o recorte empírico. A Secção 4 analisa as políticas da Meta e o percurso de denúncia no Instagram. A Secção 5 reagrega os achados e discute implicações regulatórias. A Secção 6 apresenta as considerações finais.

2. Comunicação e Conflitos de Governação de Plataformas Digitais

2.1. Plataformização, Governação de Plataformas e Moderação Algorítmica

A noção de "plataformização" (van Dijck et al., 2018) e o regime PDPA (Lemos, 2021) ajudam a compreender como plataformas digitais assumem um papel estruturante na organização de setores sociais e económicos, atuando como infraestruturas que agregam dados, intermediações e interações em larga escala. Este processo impõe lógicas próprias de classificação e monetização, traduzidas em formas específicas de governação de conteúdos, nas quais decisões sobre o que é (ou não) visível são tomadas por regras corporativas e sistemas algorítmicos. Nesse sentido, plataformas como

Instagram, TikTok, YouTube ou X exercem poder infraestrutural que condiciona a circulação de informação e a cidadania digital (Plantin & Punathambekar, 2019).

A moderação de conteúdo deixa de ser um serviço auxiliar e torna-se parte do "produto" que as plataformas oferecem, juntamente com a promessa de um espaço seguro e ordenado (Gillespie, 2018, 2020). Essa governação sociotécnica envolve escolhas de desenho e políticas que definem, por exemplo, categorias de "ódio", modelos de orientação, sanções aplicadas e línguas e contextos culturais priorizados (Birhane, 2023; Birhane & Cummins, 2019).

Neste quadro, a moderação algorítmica assume-se como eixo central da governação de plataformas, articulando filtros automatizados, equipas humanas e denúncia num regime híbrido e opaco (Douek, 2022; Gillespie, 2018, 2020; Roberts, 2019). A pressão regulatória, como a exercida pelo Regulamento dos Serviços Digitais na União Europeia (2022), coexiste com imperativos comerciais de atenção e redução de custos, produzindo compromissos sobre o que é classificado como "ódio" e quando intervir. Para o caso em análise, isso significa que a forma como o Instagram define, deteta e sanciona o ódio anti-imigrante resulta de decisões infraestruturais situadas, inscritas nas materialidades normativas e operacionais da plataforma.

2.2. Conflitos Comunicacionais, Ódio e Migração na Cultura Digital

Os conflitos em torno da imigração extrapolam os limites da política institucional, dos meios de comunicação tradicionais e dos fóruns jurídicos, manifestando-se de forma crescente na cultura digital. Nesse contexto, plataformas sociotécnicas funcionam como infraestruturas cruciais na disputa por visibilidade, reconhecimento e poder, reorganizando mercados, trabalho, sociabilidade e participação cívica (Lemos, 2021; van Dijck et al., 2018).

A centralidade destas infraestruturas nas migrações transnacionais manifesta-se em múltiplos níveis (Leurs & Ponzanesi, 2018): da pesquisa de informação sobre rotas e burocracias à construção de comunidades afetivas e redes de apoio, passando pela produção de conteúdos que transformam a experiência migratória em mercadoria comunicacional (Andrade, 2025). Ao mesmo tempo, esses espaços são permeados por discursos que contestam a legitimidade da presença migrante, associando-a a questões como criminalidade, abuso de prestações sociais ou ameaça cultural.

Tais disputas discursivas assumem frequentemente a forma de enunciados hostis dirigidos a pessoas ou grupos, com base na sua origem nacional, raça, etnia, religião ou condição migrante, e são entendidos aqui, em sentido amplo, como formas de ódio anti-imigrante motivadas por preconceito, que procuram incitar, justificar ou normalizar exclusão, hostilidade ou violência (Smets et al., 2022; Waldron, 2012).

Nestes termos, o conflito comunicacional digital em torno da imigração deve ser compreendido como um fenómeno simultaneamente simbólico e material: envolve narrativas, enquadramentos e representações sobre quem "merece" estar num dado território, mas depende também de infraestruturas sociotécnicas, que condicionam o que circula, a quem chega e com que intensidade. É nesse cruzamento entre disputa simbólica e governação infraestrutural que se situa este artigo.

2.3. Xenofobia e Ódio Anti-Imigrante no Espaço Digital Português

No contexto europeu, o discurso de ódio é definido pelo Comité de Ministros do Conselho da Europa como todas as formas de expressão que "propaguem, incitem, promovam ou justifiquem ódio racial, xenofobia, antissemitismo e outras formas de ódio baseado na intolerância" (Keen & Georgescu, 2016, p. 11), incluindo hostilidade contra minorias, migrantes e pessoas de origem migrante. Trata-se de um dano que ultrapassa a ofensa individual, na medida em que visa minar o estatuto de igualdade e a dignidade pública dos grupos-alvo, procurando excluí-los da participação plena na sociedade (Waldron, 2012). No espaço europeu, o racismo e a xenofobia figuram entre as principais motivações reportadas para ameaças e ataques baseados no ódio (Organization for Security and Co-operation in Europe – Office for Democratic Institutions and Human Rights, 2021), e os dados disponíveis para Portugal indicam que pessoas imigrantes estão entre as vítimas mais frequentes dessas formas de violência (Comissão Europeia contra o Racismo e a Intolerância, 2025).

O relatório do projeto *Racismo e Xenofobia em Portugal: A Normalização dos Discursos de Ódio no Espaço Público da Internet* (2022) evidencia a banalização desses enunciados em comentários de notícias e redes sociais, mostrando como o discurso de ódio passa a integrar o quotidiano mediático. Narrativas que associam imigração a ameaça, criminalidade, incivilidade ou "abuso" de apoios sociais circulam em ambientes digitais, frequentemente articuladas a estereótipos raciais e étnicos.

Relatórios recentes aprofundam este diagnóstico. O projeto *#MigraMyths* mostra que o ódio dirigido a pessoas migrantes se tornou recorrente no debate público português, num contexto de polarização política e maior politização das migrações, em que movimentos populistas e extremistas exploram temas como "identidade nacional", "insegurança" e "crise económica" para produzir narrativas desinformadas contra migrantes e refugiados (Costa, 2025). O sexto relatório da Comissão Europeia contra o Racismo e a Intolerância (2025) assinala um aumento acentuado de discurso de ódio no país, dirigido sobretudo a migrantes, comunidades ciganas, pessoas negras e pessoas LGBTQIA+, alertando para a sua "trivialização" sob o pretexto da liberdade de expressão. Em paralelo, dados da Associação Portuguesa de Apoio à Vítima (2025) indicam uma subida das denúncias de situações de discriminação e incitamento ao ódio e à violência, incluindo através da internet.

A comunidade brasileira tem sido alvo particular de hostilidade digital: reportagens e análises apontam para uma "onda de ódio" contra brasileiros em Portugal, em que o discurso anti-imigração e a estigmatização circulam em comentários, *memes* e publicações nas redes sociais, alimentando perceções de insegurança e não-pertença (Santos, 2023). Esses discursos articulam-se com a ascensão eleitoral da extrema-direita e com a atuação de grupos organizados de ódio, que utilizam plataformas digitais para difundir propaganda anti-imigração e mobilizar apoiantes.

Os dados do relatório da Casa do Brasil de Lisboa (Costa, 2025) ajudam a precisar o lugar das plataformas neste cenário: 79,8% das pessoas imigrantes inquiridas referiram já ter sofrido discurso de ódio online em Portugal, em particular no Instagram (36,9%). Para efeitos deste artigo, interessa menos quantificar esses episódios e mais

compreender como as infraestruturas de plataforma moldam as condições da sua emergência, visibilidade e controlo. O Instagram, enquanto ambiente híbrido de socialização, autopromoção e trabalho digital, é um dos espaços onde tais dinâmicas se manifestam, seja através de conteúdos e comentários hostis, seja por meio de denúncias, remoções e (des)recomendações mediadas por algoritmos.

É neste ponto que se torna necessário um quadro teórico-metodológico capaz de tratar políticas de comunidade, categorias de "ódio", interfaces de denúncia e ferramentas de moderação como actantes na controvérsia (Latour, 1994, 2012; Lemos, 2020) em torno da imigração, enquadramento que apresentamos na secção seguinte.

3. Princípios Teórico-Metodológicos e Procedimentos

3.1 Recorte Exploratório: O Instagram

O recorte empírico concentra-se no Instagram, por ser uma plataforma central no quotidiano digital de pessoas imigrantes em Portugal, como local de informação, sociabilidade, visibilidade pública e, em muitos casos, fonte de rendimento. Como já referido, inquéritos recentes sobre discurso de ódio contra pessoas imigrantes em Portugal indicam o Instagram como o principal espaço online em ocorrências deste tipo (Costa, 2025), o que reforça a pertinência deste foco.

O Instagram constitui também um caso paradigmático da plataformização da vida social (van Dijck et al., 2018) e do regime PDPA (Lemos, 2021), articulando formatos audiovisuais curtos, sistemas de recomendação algorítmica e métricas de engajamento que modulam as ações. Além disso, é um serviço sujeito a crescente escrutínio regulatório e público, o que torna particularmente pertinente interrogar a materialidade dos seus mecanismos de governação do ódio.

Este recorte traduz-se numa análise material-discursiva (Barad, 2007; Kitchin & Dodge, 2011) de um conjunto delimitado de elementos infraestruturais do Instagram: os padrões da comunidade da Meta (com especial atenção às secções "Conduta de ódio", "Violência e incitação" e "Bullying e assédio"), as interfaces de denúncia acessíveis na aplicação e páginas relevantes do Centro de Ajuda, que descrevem os fluxos de denúncia, revisão e informação ao utilizador.

Embora o foco analítico recaia no Instagram, o caso é representativo de disputas mais amplas em torno da governação do ódio em plataformas, nas quais diferentes empresas articulam, de modos distintos, compromissos com regulação, interesses económicos e visões sobre liberdade de expressão e segurança.

3.2. Metodologia

Adota-se a metodologia neomaterialista da comunicação associal (Lemos, 2020), partindo da premissa de que a comunicação digital é um processo de mediação distribuída entre humanos e não humanos (Latour, 1994, 2012); assim, políticas de comunidade, interfaces, sistemas e mecanismos são tratados como actantes que participam ativamente na controvérsia em torno do ódio anti-imigração.

Operacionalmente, o estudo mobiliza as quatro etapas da cartografia neomaterialista (Lemos, 2020): (a) *modo* — delimitação da controvérsia na forma como o Instagram define, classifica e modera conteúdos de ódio anti-imigrante e como essa governação incide sobre os conflitos em torno da imigração em Portugal; (b) *inventário* — identificação e descrição sistemática dos actantes infraestruturais relevantes: documentos normativos e interfaces de denúncia acessíveis a utilizadores; (c) *transdução* — análise das mediações produzidas por estes actantes, isto é, de como categorias, normas e fluxos de denúncia transformam possibilidades de circulação, visibilidade e controlo de conteúdos de ódio; (d) *reagregação* — síntese dos resultados em eixos analíticos que permitem discutir a governação do ódio como processo de mediação sociotécnica com efeitos específicos sobre as condições de segurança, visibilidade e contestação de pessoas imigrantes.

3.3. Procedimentos: Análise Material-Discursiva de Documentos e Interfaces

Os procedimentos empíricos adotados inscrevem-se na análise material-discursiva (Barad, 2007; Kitchin & Dodge, 2011), entendida como uma abordagem que trata textos, imagens, códigos e interfaces como materiais com agência, capazes de configurar modos de ação e significação.

O trabalho de *inventário* e *transdução* organiza-se em dois conjuntos principais de materiais:

- Documentos normativos e de orientação — padrões da comunidade da Meta, com destaque para as secções "Conduta de ódio", "Violência e incitação" e "Bullying e assédio", bem como as páginas do Centro de Ajuda do Instagram sobre privacidade, denúncia e segurança. As versões foram selecionadas por corresponderem às categorias diretamente mobilizadas pelo fluxo de denúncia de conteúdos de ódio, violência e assédio. A recolha ocorreu a 4 de dezembro de 2025, em páginas de língua portuguesa do Centro de Transparência da Meta e do Centro de Ajuda. A análise concentrou-se nas categorias de ódio e grupos protegidos, nos exemplos, nas sanções e nas garantias de confidencialidade, revisão e recurso.
- Interfaces de denúncia e pedidos de suporte no Instagram — a partir de uma conta de utilizador, foram registados os percursos de denúncia de publicações, comentários e contas, com capturas de ecrã e notas de campo sobre *affordances*, categorias e caminhos disponíveis para reportar ódio ou comportamentos abusivos, bem como as mensagens informativas. As capturas apresentadas referem-se ao percurso de denúncia de uma publicação observada a 9 de dezembro de 2025, num dispositivo móvel Samsung, com sistema operativo Android, aplicação móvel configurada em português e utilizada em Portugal. Como se pretendia verificar o fluxo de denúncia, e não avaliar um conteúdo específico, foi selecionada uma publicação aleatória de um perfil público. Por razões éticas, a denúncia não foi enviada, tendo o procedimento sido interrompido antes da submissão final. Como as interfaces podem variar segundo dispositivo, sistema operativo, versão da aplicação, região, idioma e tipo de conteúdo, estes elementos são tratados como condições da materialidade observada.

4. Políticas de Comunidade, Interfaces e Fluxos de Denúncia no Instagram

As políticas que regem o Instagram integram-se nos padrões da comunidade da Meta, documento transversal às tecnologias da empresa (Facebook, Instagram, Messenger e Threads). Segundo a Meta, esses padrões procuram conciliar "espaço para

a expressão" e segurança, definindo o que é ou não permitido nas plataformas e aplicando-se a todos os tipos de conteúdo, incluindo conteúdos gerados por inteligência artificial (Meta, s.d.-a, s.d.-b, s.d.-c).

Cada secção organiza-se em torno de uma fundamentação lógica da política, seguida de linhas específicas, que distinguem conteúdos proibidos dos que exigem contexto adicional para avaliação. Para fins desta análise, foram selecionadas as secções "Conduta de ódio", "Violência e incitação" e "Bullying e assédio" por constituírem o núcleo da governação do ódio, violência e assédio no Instagram, em articulação com as orientações práticas de denúncia e acompanhamento de pedidos de suporte.

Antes de avançar para a análise, convém estabilizar a relação entre os termos mobilizados neste artigo e as categorias da Meta. "Ódio" é usado como categoria analítica ampla para práticas de hostilidade, desumanização, intimidação ou exclusão dirigidas a grupos vulnerabilizados. "Discurso de ódio" refere-se às suas formas expressivas, como enunciados, imagens, comentários ou símbolos que propagam, incitam ou justificam hostilidade e discriminação. "Xenofobia" designa a rejeição ou inferiorização de pessoas percebidas como estrangeiras. "Ódio anti-imigrante" nomeia formas de hostilidade dirigidas a pessoas em razão da condição migrante, origem nacional ou estatuto de estrangeiridade. Estes termos não equivalem diretamente às categorias corporativas da Meta: "Conduta de ódio", "Violência e incitação" e "Bullying e assédio" são aqui tratadas como formas infraestruturais de classificação, pelas quais a plataforma traduz, hierarquiza e torna moderáveis determinados danos.

À luz do quadro neomaterialista, estas secções e interfaces são tratadas como mediadores que desenham o perímetro do que pode ser dito e feito em relação a pessoas e grupos na plataforma. Nas subsecções seguintes, examina-se como estas políticas e dispositivos configuram categorias protegidas, níveis de gravidade e áreas de ambiguidade que definem as condições de possibilidade para a disseminação e o controlo do ódio anti-imigrante no Instagram.

4.1. "Conduta de Ódio": Definição, Categorias Protegidas e Lugar do Estatuto Migrante

A secção "Conduta de ódio" (ou "Comportamento que incita ao ódio") dos padrões da comunidade da Meta (s.d.-b) começa por afirmar que as pessoas "utilizam a sua voz e estão em contacto mais livremente quando não se sentem atacadas com base em quem são" (para. 1) e, por isso, declara não permitir comportamentos que incitam ao ódio nas suas plataformas.

No registo de alterações, uma revisão datada de 7 de janeiro de 2025 apresenta o texto anteriormente associado a "Discurso de ódio" sob a rubrica atual "Conduta de ódio". Embora o histórico não explicite literalmente a mudança de título, a comparação entre versões permite inferir uma alteração terminológica que desloca o foco de uma ênfase no "discurso" para uma ênfase em "conduta", sugerindo que a política se aplica não apenas a enunciados linguísticos, mas a formas mais amplas de atuação hostil.

Considerando a prática material-discursiva (Barad, 2007), tal mudança indica o caráter processual e adaptativo da governação do ódio: a própria definição do fenómeno é objeto de ajustes infraestruturais contínuos.

Na versão consultada, a Meta (s.d.-b) define comportamentos que incitam ao ódio como "ataques diretos contra pessoas, em vez de conceitos ou instituições, com base no que denominamos caraterísticas protegidas" (para. 2). Estas incluem: raça, etnia, nacionalidade, deficiência, afiliação religiosa, casta, orientação sexual, sexo, identidade de género e doença grave, considerando ainda a idade quando associada a outra característica protegida. O texto acrescenta que a Meta "também [protege] refugiados, migrantes, imigrantes e pessoas que procurem asilo dos ataques mais graves (Nível 1 abaixo), apesar de permitirmos comentários e críticas a políticas de imigração" (Meta, s.d.-b, para. 2). Ao inscrever explicitamente "refugiados, migrantes, imigrantes e pessoas que procurem asilo" no âmbito da proteção contra ataques graves, a política integra o estatuto migrante na sua gramática interna, aproximando-o de outros marcadores considerados sensíveis.

A remoção de conteúdos é organizada em níveis de gravidade. O Nível 1 abrange ataques considerados mais severos, como discurso desumanizador (comparações com animais, organismos patogénicos ou "formas de vida sub-humanas"), alegações de imoralidade ou criminalidade grave e desejos de que determinadas pessoas sofram danos (como contrair doenças, morrer ou sofrer desastres naturais), quando os alvos são definidos por caraterísticas protegidas ou estatuto de imigrante. O Nível 2 inclui incentivos ou apoios à exclusão ou segregação (por exemplo, negar direitos políticos, económicos ou sociais a um grupo), bem como insultos e expressões de desprezo, repugnância ou nojo dirigidos com base nas mesmas caraterísticas.

Do ponto de vista da proteção de populações migrantes, a hierarquização entre Nível 1 e Nível 2 é reveladora. Enquanto o primeiro concentra ameaças diretas e desumanização explícita, o segundo agrupa práticas que, no plano jurídico e sociológico, se aproximam de violência estrutural e discriminação sistemática: exclusão política, económica e social, incentivos à segregação e insultos reiterados. Considerá-las "menos graves" do que ameaças individuais de dano físico privilegia a violência física direta em detrimento de formas de exclusão, que, de modo difuso e prolongado, afetam a capacidade de participação de grupos vulnerabilizados, entre os quais se incluem pessoas imigrantes.

A política reconhece ainda que nem todo conteúdo com linguagem potencialmente ofensiva deve ser automaticamente removido. Admite que utilizadores partilhem insultos ou discurso de ódio de terceiros (para condenar ou denunciar) ou usem termos potencialmente ofensivos (de forma autorreferencial ou "fortalecedora") e declara permitir esse tipo de discurso "quando a intenção do orador é clara" (Meta, s.d.-b, para. 4), admitindo que, quando tal não se verifica, os conteúdos podem ser removidos. Acrescenta que, em certos contextos políticos e religiosos (como debates sobre direitos de pessoas transgénero, imigração ou homossexualidade), as políticas foram "concebidas para dar espaço a estes tipos de discurso" (para. 5), bem como a algum uso de sátira, desde que os elementos em infração estejam a ser criticados ou ridicularizados.

Estas disposições introduzem uma segunda camada de ambivalência. Ao admitir que conteúdos potencialmente ofensivos podem ser mantidos quando são interpretados como denúncia, uso autorreferencial ou humorístico, e ao prever margem adicional para debates políticos e religiosos (incluindo sobre imigração), a política desloca parte da decisão de moderação para um exercício interpretativo situado. Na prática, enunciados hostis dirigidos a pessoas imigrantes podem ser enquadrados como opinião legítima sobre "controlo de fronteiras" ou "segurança", escapando à remoção. A margem concedida à sátira e a discursos ambíguos amplia essas áreas de indeterminação, sobretudo em contextos de forte polarização em torno da imigração, como o português.

A secção "Conduta de ódio" funciona, assim, como mediador infraestrutural que define quais atributos identitários merecem proteção reforçada, hierarquiza formas de violência e desenha limites e exceções que condicionam, em simultâneo, visibilidade, vulnerabilidade e possibilidades de contestação por parte de pessoas imigrantes.

4.2. "Violência e Incitação": Dano Offline e Graduações de Proteção

A política de "Violência e incitação" tem como objetivo "impedir potencial violência offline que possa estar relacionada com conteúdos" (Meta, s.d.-c, para. 1) nas plataformas da Meta. Declara remover linguagem que incite ou facilite violência, bem como ameaças consideradas credíveis à segurança pública ou pessoal, incluindo discurso violento dirigido a pessoas ou grupos "com base nas respetivas caraterísticas protegidas ou estatuto de imigrante" (para. 1).

A secção organiza-se em torno de um conjunto de "proteções universais" (que abrangem todas as pessoas face a ameaças de violência de elevada e média gravidade) e de "proteções adicionais" atribuídas a sujeitos definidos como mais vulneráveis, entre os quais crianças, pessoas em situação de alto risco e indivíduos ou grupos visados com base em caraterísticas protegidas. Na medida em que ataques a pessoas imigrantes podem ser formulados a partir de origem nacional, etnia, religião ou do próprio estatuto migrante, a política abre espaço para que, em teoria, beneficiem deste patamar reforçado de intervenção.

Para além das ameaças diretas, a secção inclui "outros tipos de violência", como pedidos ou ofertas de serviços violentos (por exemplo, contratação de assassinos), instruções para uso de armas ou explosivos com intenção de causar dano grave e ameaças de levar armas para locais sensíveis (instituições de ensino, locais de culto, espaços eleitorais). Em todos estes casos, a política volta a admitir exceções contextuais (por exemplo, em situações de autodefesa, treino desportivo ou militar), o que reforça a centralidade da interpretação situada.

Uma parte significativa do texto é dedicada às situações em que a aplicação dos padrões exige "informações e/ou contexto adicionais", como ameaças codificadas ou implícitas, referências a episódios históricos de violência, convites a terceiros para praticar agressões ou insinuações sobre o conhecimento de dados sensíveis (morada, local de trabalho, rotinas diárias). Também aqui a intervenção depende da avaliação da plataforma sobre a credibilidade e a iminência do dano.

Esta política reforça a primazia do critério de dano offline: aquilo que justifica a remoção é, sobretudo, o potencial de produzir agressões físicas fora do ecrã. Para a controvérsia em torno da imigração em Portugal, isso sugere um duplo efeito. Por um lado, a referência ao estatuto de imigrante confirma o reconhecimento de pessoas imigrantes como grupo vulnerável. Por outro, ao privilegiar ameaças credíveis e risco físico imediato, a política tende a subvalorizar intimidação e violência simbólica, que alimentam um ambiente hostil e de deslegitimação de populações migrantes. As exceções e avaliações contextuais mantêm margens nas quais enunciados hostis podem circular como opinião política, indignação ou retórica figurada.

4.3. "Bullying e Assédio": Ataques Personalizados e Assédio em Massa

A política de "Bullying e assédio" desloca o foco da proteção de grupos definidos por caraterísticas protegidas para a proteção de pessoas singulares e figuras públicas (Meta, s.d.-a). A Meta (s.d.-a) declara não tolerar comportamentos que façam com que as pessoas "não se sintam seguras e respeitadas" (para. 1) e estabelece uma distinção entre figuras públicas e utilizadores comuns, reforçando as salvaguardas para estes últimos, em especial menores.

A secção organiza diferentes níveis de proteção, que abrangem desde contacto malicioso repetido, ameaças de divulgação de dados pessoais (*doxing*), incitações à automutilação, humilhação e sexualização indesejada, até ataques à aparência física e assédio em massa coordenado. Menores, pessoas singulares adultas e figuras públicas involuntárias recebem camadas adicionais de proteção, sobretudo quando são elas próprias a denunciar o conteúdo.

No contexto da imigração, esta política é relevante porque enquadra as formas de ataque mais frequentes dirigidas a pessoas imigrantes com visibilidade pública: contacto hostil insistente, comentários depreciativos recorrentes, tentativas de humilhação e intimidação através da exposição de informação pessoal. Embora a secção não nomeie explicitamente "imigrantes", menciona "pessoas em risco elevado de danos offline", incluindo defensores de direitos humanos, minorias étnicas e religiosas em zonas de conflito e vítimas de tragédias violentas. Assim, a formulação pode abranger criadores e ativistas migrantes como alvo de campanhas coordenadas de assédio devido ao conteúdo que produzem.

Ao mesmo tempo, a política reforça a dependência da autodenúncia: em diversos casos, a intervenção da plataforma exige que seja a própria vítima, ou um representante autorizado, a sinalizar que o conteúdo é indesejado. Para pessoas imigrantes que trabalham com criação de conteúdo, isso significa que a proteção está condicionada à sua capacidade de reconhecer o ataque, denunciá-lo e gerir as ferramentas de moderação disponíveis, acumulando camadas adicionais de trabalho e desgaste emocional.

A política de "Bullying e assédio" funciona como mediador que regula a fronteira entre crítica legítima, exposição pública e violência simbólica dirigida a sujeitos concretos. Em articulação com "Conduta de ódio" e "Violência e incitação", compõem um núcleo normativo-infraestrutural que, enquanto promete proteção reforçada, produz

limites e ambivalências quanto ao que conta como ataque inaceitável, com impactos diretos na experiência de segurança e na sustentabilidade do trabalho digital de pessoas imigrantes no Instagram e em outras plataformas da Meta.

4.4. Fluxo de Denúncia e Pedidos de Suporte no Instagram: Entre Condução e Opacidade

A partir do ponto de vista da utilizadora, o percurso documentado de denúncia de uma publicação começa ao clicar no ícone de três pontos, associado a conteúdos no *feed*, *reels* ou *stories*. O primeiro ecrã (Figura 1) apresenta opções como "legendagem opcional", "porque é que está a ver esta publicação?", "sem interesse", "com interesse", "sobre esta conta", seguidas de um link para "gerir preferências de conteúdos". A opção "denunciar", destacada a vermelho, marca a passagem de um regime de personalização de *feed* para um regime de governação de conteúdos: ao ser acionada, traduz-se uma experiência subjetiva de ódio ou violência em categorias infraestruturais reconhecidas pela plataforma.

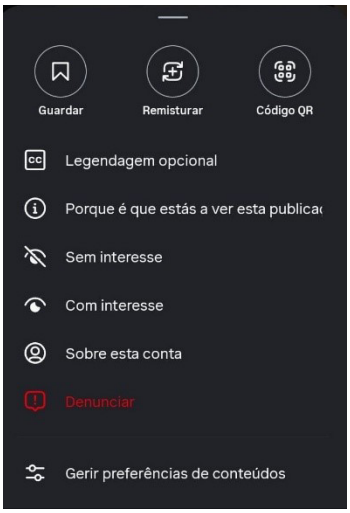

Figura 1. *Primeiro ecrã do percurso de denúncia*

Nota. Captura de ecrã, 9 de dezembro de 2025.

No ecrã seguinte (Figura 2), a utilizadora é confrontada com a pergunta "porque está a denunciar esta publicação?", acompanhada da indicação de que "a tua denúncia é anónima" e da instrução para contactar serviços de emergência em caso de perigo iminente. Em seguida, são apresentadas opções de categorização, entre as quais: "simplesmente não gosto", "bullying ou contacto indesejado", "suicídio, automutilação ou distúrbios alimentares", "violência, ódio ou exploração", "venda ou promoção de artigos restritos", "nudez ou atividade sexual", "burla, fraude ou *spam*", "informações falsas", "propriedade intelectual" e "denunciar como ilegal".

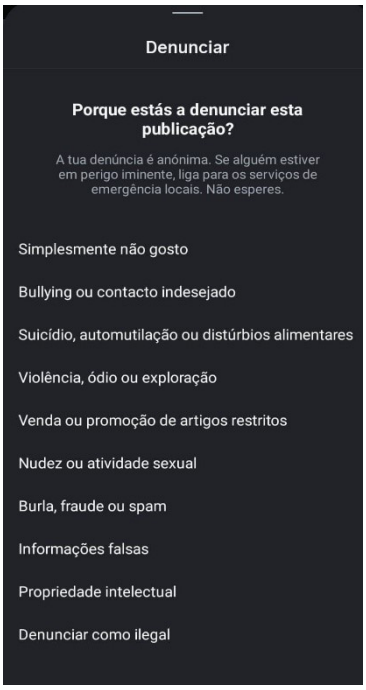

Figura 2. *Categorias inicias para denúncia*

Nota. Captura de ecrã, 9 de dezembro de 2025.

Quando a utilizadora escolhe "violência, ódio ou exploração", um novo ecrã (Figura 3) solicita maior especificidade: "de que forma se enquadra na opção violência, ódio ou exploração?", oferecendo subcategorias como "ameaça credível à segurança", "parece terrorismo ou crime organizado", "parece exploração", "discurso ou símbolos de incentivo ao ódio", "incentiva a violência", "mostra violência, morte ou ferimentos graves" ou "agressão a animais".

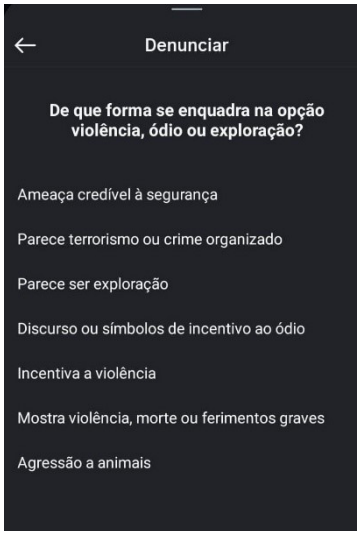

Figura 3. *Subcategorias de "violência, ódio ou exploração" no percurso de denúncia*

Nota. Captura de ecrã, 9 de dezembro de 2025.

No ecrã seguinte (Figura 4), o Instagram apresenta um resumo da denúncia: "vais enviar uma denúncia. Apenas removemos conteúdos que desrespeitam os nossos Padrões da Comunidade", com um link direto para esse documento, seguido dos campos "porque estás a denunciar esta publicação?" e "de que forma se enquadra na opção violência, ódio ou exploração?", agora preenchidos com as escolhas feitas ("violência, ódio ou exploração"/"discurso ou símbolo de incentivo ao ódio") e o botão "enviar". Esta etapa final inscreve explicitamente o fluxo de denúncia na arquitetura normativa analisada nas secções anteriores: a plataforma sublinha que nem tudo o que é ofensivo será removido, apenas o que se enquadrar nas suas próprias definições de violação.

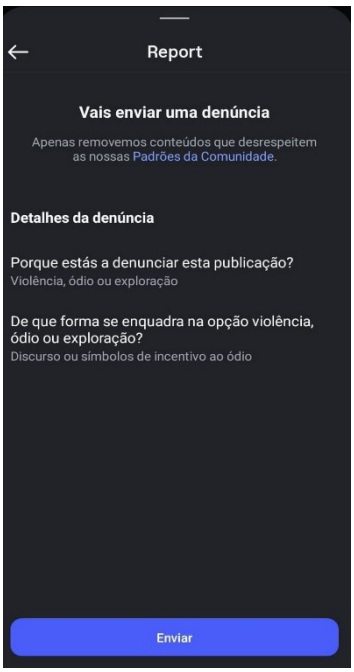

Figura 4. *Ecrã de confirmação antes do envio da denúncia*

Nota. Captura de ecrã, 9 de dezembro de 2025.

Esta organização reforça a leitura das políticas: embora a Meta reconheça "refugiados, migrantes, imigrantes e pessoas que procurem asilo" como alvo protegido nos padrões da comunidade, o estatuto "migrante" não aparece explicitamente na interface de denúncia. À utilizadora cabe localizar a experiência de ódio anti-imigrante dentro da categoria genérica de "discurso ou símbolos de incentivo ao ódio". Em termos de *transdução* (Lemos, 2020), experiências situadas de xenofobia são traduzidas numa gramática normativa abstrata, sem emergir na superfície da aplicação.

Desse modo, este percurso evidencia três aspetos centrais da governação do ódio no Instagram:

- Redistribuição da responsabilidade de classificação — a interface delega na utilizadora a tarefa de traduzir a sua experiência vivida num código interno de categorias, realizando um trabalho de classificação que alimenta o processo de moderação e a datafication do conflito.
- Deslocamento entre experiência migrante e gramática normativa — a condição migrante é reconhecida no plano textual das políticas, mas não se torna diretamente visível no fluxo de denúncia, que trata o ódio anti-imigrante como caso genérico de "discurso de ódio".

- Produção de zonas de ambiguidade — a copresença de opções como "simplesmente não gosto" e "violência, ódio ou exploração", bem como a necessidade de distinguir ameaça "credível" de linguagem figurada, pode deslocar experiências de ódio anti-imigrante para a esfera do desacordo, da opinião política ou da sátira.

As instruções do Centro de Ajuda (s.d.) reforçam e complementam esta representação. Na secção "Privacidade e denúncia", textos sobre denúncia de assédio, bullying ou conteúdo abusivo indicam que contas, comentários e publicações podem ser denunciados diretamente na aplicação através da ferramenta integrada de denúncia. O utilizador é também encorajado a bloquear ou restringir a conta denunciada, e o Centro de Ajuda reitera que a denúncia é anónima e que nenhuma informação sobre quem denuncia é partilhada com a pessoa denunciada.

O Centro de Ajuda (s.d.) descreve ainda a possibilidade de verificar o estado de algo denunciado por meio da área "Pedidos de suporte", acessível no menu de configurações. A utilizadora pode consultar algumas denúncias anteriores, ver se a plataforma tomou alguma medida, ler explicações adicionais sobre as políticas aplicadas e, em certos casos, solicitar uma nova análise da decisão. Ao mesmo tempo, é explicitado que nem todas as denúncias ficam visíveis em "Pedidos de suporte" e que nem sempre existe a opção de recurso, o que introduz um grau de opacidade quanto à forma como a plataforma prioriza e comunica determinados casos.

Lido em conjunto com as políticas de "Conduta de ódio", "Violência e incitação" e "Bullying e assédio", este fluxo ilustra uma dupla dinâmica. Por um lado, há um investimento em condução: a interface simplifica a classificação, oferece garantias de anonimato, remete para recursos de segurança e sugere ações complementares, como bloquear ou restringir contas. A responsabilidade pela ativação da proteção é, assim, atribuída à utilizadora, que deve reconhecer a situação, navegar pelas categorias corretas, acompanhar o estado da sua denúncia e, se for o caso, solicitar revisão. Por outro lado, persistem zonas de opacidade: a utilizadora não acede aos critérios concretos que determinam se um conteúdo denunciado por ódio ou violência será removido, mantido ou apenas limitado; nem sabe por que certas denúncias aparecem em "Pedidos de suporte" e outras não, ou em que condições é possível recorrer da decisão.

Em contextos de polarização em torno da imigração, como o português, essa combinação de condução e opacidade configura, de modo situado, as condições de segurança, visibilidade e contestação de pessoas imigrantes no Instagram, mediando a forma como experiências de ódio anti-imigrante são (ou não) reconhecidas, traduzidas ou desclassificadas pela plataforma.

5. Governação do Ódio, Imigração e Plataformas

5.1. Eixos de Reagregação da Controvérsia

A partir do percurso metodológico da comunicação associal (Lemos, 2020), a *reagregação* implica retomar o caminho traçado pelo *modo*, *inventário* e *transdução* para

propor uma leitura propositiva da controvérsia. No caso em análise, três eixos sintetizam a relação entre governação do ódio, imigração e plataformas digitais.

O primeiro eixo diz respeito ao reconhecimento parcial do estatuto migrante. No plano normativo, a secção "Conduta de ódio" inscreve explicitamente "refugiados, migrantes, imigrantes e pessoas que procurem asilo" entre os alvos protegidos contra ataques graves, aproximando a condição migrante de outras características sensíveis, como raça, etnia, nacionalidade ou religião.

No entanto, tal reconhecimento permanece maioritariamente circunscrito à camada textual das políticas: na interface de denúncia, a experiência de ódio anti-imigrante é absorvida em categorias genéricas como "violência, ódio ou exploração" ou "discurso ou símbolos de incentivo ao ódio", sem qualquer referência explícita à origem nacional ou ao estatuto de imigrante. A condição migrante, central na controvérsia sociopolítica portuguesa, aparece assim visível na gramática normativa e invisível na superfície operacional da aplicação.

O segundo eixo diz respeito às hierarquias de dano. A estrutura em níveis, da secção "Conduta de ódio", associa o Nível 1 a ataques desumanizadores, alegações de criminalidade grave e desejos de dano físico, enquanto o Nível 2 agrupa incentivos à exclusão política, económica e social, à segregação e a insultos reiterados.

Do ponto de vista da experiência migrante, esta hierarquização é problemática porque práticas de exclusão, estigmatização e deslegitimação sistemática são formalmente classificadas como "menos graves" do que ameaças diretas de dano físico, ainda que possam contribuir para a produção de medo, retraimento e precarização da participação pública. De modo semelhante, a política de "Violência e incitação" privilegia o critério de "risco credível de dano offline", tendendo a subvalorizar formas de intimidação simbólica, que não se traduzem de imediato em agressões físicas, mas integram o ambiente hostil vivido por pessoas imigrantes.

O terceiro eixo diz respeito à redistribuição assimétrica da responsabilidade. Tanto as políticas como o fluxo de denúncia deslocam para o utilizador a tarefa de reconhecer o ódio, traduzi-lo em categorias internas e acionar ferramentas de proteção. Em paralelo, a Meta reserva para si a decisão final sobre o que viola, ou não, os padrões da comunidade, introduzindo opacidade quanto aos critérios concretos de decisão, à priorização de casos e à possibilidade de recurso.

Quando a imigração se torna eixo de forte polarização política e afetiva, essa combinação de condução e ambiguidade abre brechas pelas quais discursos hostis contra pessoas imigrantes podem ser reclassificados como opinião política, sátira ou retórica figurada, escapando à remoção mesmo quando contribuem para a banalização do ódio anti-imigrante.

Se percebidos a partir do regime PDPA (Lemos, 2021), estes achados ganham uma camada adicional. As políticas e interfaces analisadas definem o que pode ser dito e estruturam a transformação de denúncias em dados, alimentando sistemas de recomendação, deteção automatizada e métricas de "integridade" que modulam a visibilidade de conteúdos e perfis. A governação do ódio em torno da imigração é, assim, um processo de

mediação sociotécnica em que categorias normativas, fluxos de denúncia e infraestruturas algorítmicas se entrelaçam na produção de diferentes graus de exposição, vulnerabilidade e possibilidade de contestação para pessoas imigrantes.

5.2. Implicações Regulatórias no Contexto Europeu

Estes resultados também têm implicações para a leitura regulatória da moderação no contexto europeu. O Regulamento dos Serviços Digitais (União Europeia, 2022) não define substantivamente o que deve contar como ódio anti-imigrante em cada situação nacional, mas desloca as plataformas para um regime de maior diligência, transparência e contestabilidade. Entre outros pontos, exige mecanismos de notificação e ação acessíveis, informação clara sobre decisões de moderação, sistemas internos de reclamação e, no caso de plataformas de muito grande dimensão, avaliação e mitigação de riscos sistémicos ligados à disseminação de conteúdos ilegais e aos efeitos sobre direitos fundamentais.

À luz desses parâmetros, a proteção de pessoas imigrantes não depende apenas da existência formal de uma categoria protegida. Depende também da legibilidade dessa categoria na interface, da possibilidade de a denúncia captar dimensões contextuais como origem nacional e estatuto migrante, da explicação das decisões tomadas e da existência de recursos efetivos quando a plataforma mantém conteúdos hostis. A questão regulatória deixa, assim, de se limitar a saber se um conteúdo isolado infringe uma norma interna; passa também a abranger a questão de saber se a arquitetura de denúncia e moderação reconhece danos sociotécnicos recorrentes, avalia riscos contextuais e responde de modo proporcional às vulnerabilidades dos grupos afetados.

6. Considerações Finais

Partindo de um contexto português marcado por mudanças na governação migratória, atrasos institucionais e forças políticas anti-imigração, este artigo investigou a materialidade da governação do ódio no Instagram, com foco nos padrões da comunidade da Meta e no fluxo de denúncia acessível a utilizadores. Com a metodologia neomaterialista da comunicação associal (Lemos, 2020), realizou-se uma análise material-discursiva (Barad, 2007; Kitchin & Dodge, 2011), que tratou políticas, interfaces e categorias de moderação como actantes que participam ativamente na controvérsia em torno do ódio anti-imigrante. As etapas de *modo*, *inventário*, *transdução* e *reagregação* permitiram uma análise das infraestruturas normativas e operacionais que enquadram o conflito.

Empiricamente, os resultados indicam que, embora a Meta reconheça formalmente refugiados, migrantes, imigrantes e pessoas que procuram asilo como alvos protegidos, esse reconhecimento é reduzido por três movimentos principais: (a) a invisibilidade do estatuto migrante na interface de denúncia, que obriga a inscrever o ódio anti-imigrante em categorias genéricas; (b) a hierarquização, que considera menos graves formas de exclusão política, económica e social centrais no ódio anti-imigrante; e (c) a combinação

entre delegação de responsabilidade no utilizador e opacidade decisória, que favorece a permanência de conteúdos hostis sob a forma de opinião política ou sátira. Em conjunto, estes elementos configuram um regime de proteção simultaneamente necessário e limitado, no qual pessoas imigrantes são convidadas a confiar na plataforma, mas permanecem expostas a ambientes marcados por incerteza e assimetria de poder.

Para o contexto português, estes achados extravasam o plano infraestrutural. Num cenário em que dados empíricos apontam a centralidade das plataformas — e do Instagram em particular — na experiência quotidiana de ódio dirigido a pessoas imigrantes, a forma como a plataforma define, hierarquiza e operacionaliza o "ódio" integra, ainda que de modo opaco, a ecologia mais ampla da governação migratória. Esta constatação também interpela o enquadramento regulatório europeu, pois a proteção de grupos vulnerabilizados depende não apenas da remoção de conteúdos ilícitos, mas da transparência, contestabilidade e legibilidade dos sistemas de moderação.

Este artigo corresponde a uma investida inicial no estudo da controvérsia do ódio anti-imigrante em Portugal, no contexto das plataformas digitais, centrada no Instagram e num conjunto específico de materiais (políticas e fluxo de denúncia). Investigações futuras poderão ampliar o foco para outras plataformas, analisando estas tensões infraestruturais em articulação com abordagens como a análise sistemática de conteúdos, a observação online/etnografia digital e a escuta direta de pessoas imigrantes, aprofundando relações, agenciamentos e afetos entre actantes humanos e não humanos na constituição do conflito (Fox & Alldred, 2022; Latour, 1994, 2012; Lemos, 2020).

Ao indicar que a governação do ódio não é apenas um problema de "moderação de conteúdos", mas um processo de mediação sociotécnica (Latour, 1994, 2012; Lemos, 2020) que envolve normas, interfaces e algoritmos, entre outros atores, o artigo contribui para o debate sobre comunicação e conflitos em torno da imigração, propondo uma agenda de investigação sobre a materialidade das plataformas e suas implicações na disseminação de práticas de ódio anti-imigrante, bem como na proteção e participação migrante no espaço digital.

Agradecimentos

Este trabalho é financiado no âmbito do financiamento plurianual do Centro de Estudos de Comunicação e Sociedade 2025-2029, referência UID/00736/2025, pela Fundação para a Ciência e a Tecnologia.

Referências

Andrade, D. (2025). *YouTubers e migração: A dinâmica comunicacional sociotécnica do trabalho digital de imigrantes brasileiros em Portugal* [Tese de doutoramento, Universidade do Minho]. RepositóriUM. https://hdl.handle.net/1822/98534

Associação Portuguesa de Apoio à Vítima. (2025). *Estatísticas APAV: Relatório anual 2024*. APAV.

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Birhane, A. (2023). Algorithmic colonization of Africa. In S. Cave & K. Dihal (Eds.), *Imagining AI: How the world sees intelligent machines* (pp. 247–260). Oxford University Press. https://doi.org/10.1093/oso/9780192865366.003.0016

Birhane, A., & Cummins, F. (2019). *Algorithmic injustices: Towards a relational ethics*. arXiv. https://doi.org/10.48550/arXiv.1912.07376

Centro de Ajuda. (s.d.). *Como denunciar algo no Instagram*. Retirado a 4 de dezembro de 2025, de https://help.instagram.com/2922067214679225/

Comissão Europeia contra o Racismo e a Intolerância. (2025). *Relatório da ECRI sobre Portugal: Sexto ciclo de monitorização*. Conselho da Europa. https://rm.coe.int/sixth-report-on-portugal-translation-in-portuguese-/1680b6668f

Costa, A. P. (2025). *Discurso de ódio e imigração em Portugal*: *Relatório* #MigraMyths – *Desmistificando a imigração*. Casa do Brasil de Lisboa; #migramyths.

Douek, E. (2022). Content moderation as systems thinking. *Harvard Law Review*, *136*(2), 58–144.

Fox, N., & Alldred, P. (2022). New materialism. In P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *SAGE research methods foundations* (pp. 1–16). SAGE. https://doi.org/10.4135/9781526421036768465

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, *7*(2), 1–5. https://doi.org/10.1177/2053951720943234

Keen, E., & Georgescu, M. (2016). *Manual para o combate do discurso de ódio online através da educação para os direitos humanos: Edição revista com a inclusão do guia dos direitos humanos para os utilizadores da internet*. Conselho da Europa; Fundação Calouste Gulbenkian.

Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. The MIT Press.

Latour, B. (1994). On technical mediation - Philosophy, sociology, genealogy. *Common Knowledge*, *3*(2), 29–64.

Latour, B. (2012). *Reagregando o social: Uma introdução à teoria do ator-rede*. EDUFBA; EDUSC.

Lemos, A. (2020). Epistemologia da comunicação, neomaterialismo e cultura digital. *Galáxia*, (43), 54-66. https://doi.org/10.1590/1982-25532020143970

Lemos, A. (2021). *A tecnologia é um vírus: Pandemia e cultura digital*. Editora Sulina.

Leurs, K., & Ponzanesi, S. (2018). Connected migrants: Encapsulation and cosmopolitanization. *Popular Communication*, *16*(1), 4–20.

Meta. (s.d.-a). *Padrões da comunidade: Bullying e assédio*. Meta Transparency Center. Retirado a 4 de dezembro de 2025, de https://transparency.meta.com/policies/community-standards/bullying-harassment/

Meta. (s.d.-b). *Padrões da comunidade: Conduta de ódio*. Meta Transparency Center. Retirado a 4 de dezembro de 2025, de https://transparency.meta.com/policies/community-standards/hate-speech/

Meta. (s.d.-c). *Padrões da comunidade: Violência e incitação*. Meta Transparency Center. Retirado a 4 de dezembro de 2025, de https://transparency.meta.com/policies/community-standards/violence-incitement/

Organization for Security and Co-operation in Europe – Office for Democratic Institutions and Human Rights. (2021). *2020 hate crime data: Key findings*. https://hatecrime.osce.org/sites/default/files/2021-11/2020%20Infographic_EN.pdf

Plantin, J.-C., & Punathambekar, A. (2019). Digital media infrastructures: Pipes, platforms, and politics. *Media, Culture & Society*, *41*(2), 163–175. https://doi.org/10.1177/0163443718818376

Racismo e xenofobia em Portugal: A normalização dos discursos de ódio no espaço público da internet. (2022). Centro de Rede de Investigação em Antropologia. https://racismoexenofobia.cria.org.pt/wp-content/uploads/2022/07/Relato%CC%81rio_Racismo_Xenofobia.pdf

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Santos, J. (2023, 17 de dezembro). Onda de ódio contra brasileiros em Portugal. *Visão*. https://visao.pt/atualidade/sociedade/2023-12-17-a-onda-de-odio-contra-os-brasileiros-em-portugal/

Smets, K., Toffano, G., & Almenara Niebla, S. (2022). Refugee and mediated lives. In M. Powers (Ed.), *Oxford research encyclopedia of communication* (pp. 1–22). Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.1263

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

União Europeia. (2022). *Regulamento (UE) 2022/2065 de 19 de outubro de 2022 relativo a um mercado único de serviços digitais (Regulamento dos Serviços Digitais)*. https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32022R2065

Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.

Nota Biográfica

Dalvacir Xavier de Oliveira Andrade é doutora em Ciências da Comunicação pela Universidade do Minho, com tese sobre a dinâmica comunicacional sociotécnica do trabalho digital de YouTubers brasileiros em Portugal. A sua investigação tem foco nas mediações infraestruturais das plataformas digitais, controvérsias sociotécnicas, migração e trabalho digital em ecossistemas de plataforma. Tem colaboração com o Centro de Estudos de Comunicação e Sociedade, na Universidade do Minho, e com o Lab404 – Laboratório de Pesquisa em Mídia Digital, Redes e Espaço, na Universidade Federal da Bahia, integrando projetos sobre cultura digital, plataformização, datificação e performatividade algorítmica. Os seus interesses de pesquisa incluem ainda metodologias neomaterialistas, etnografias digitais e estudos críticos sobre precariedade laboral e desigualdades no ambiente digital.

ORCID: https://orcid.org/0000-0001-8886-4038

Email: dalvacirx@gmail.com

Morada: Lab404 – Laboratório de Pesquisa em Mídia Digital, Redes e Espaço,

Programa de Pós-Graduação em Comunicação e Cultura Contemporâneas, Faculdade de Comunicação, Universidade Federal da Bahia, Rua Barão de Jeremoabo, s/n, Campus Universitário de Ondina, 40170-115 Salvador – BA, Brasil

Submetido: 30/12/2025 | Aceite: 19/06/2026