COOPERHATE

COMBATE AO DISCURSO DE ÓDIO

Discurso de Ódio: Manual de Apoio e Glossário

Cofinanciado pela União Europeia

Autoras

Rita Guerra (CIS-Iscte), Raquel António (CIS-Iscte), Paula Carvalho (cllc, Universidade de Aveiro)

Autoras colaboradoras

Mónica Gonçalves & Vânia Sampaio (IPS_Innovative Prison Systems), Tânia Azevedo (ILGA Portugal), Anizabela Amaral & Beatriz Realinho (SOS Racismo)

COOPERHATE: Multidisciplinary cooperation approach to prevent and counter hate crime and hate speech

Declaração de financiamento: O projeto Cooperhate, ao abrigo do Acordo de Subvenção n.º 101213938, é financiado pela União Europeia. As visões e opiniões expressas neste documento são, no entanto, da exclusiva responsabilidade dos autores e não refletem necessariamente as da União Europeia. A União Europeia não pode ser responsabilizada por elas.

CC BY-ND 4.0

Como citar este relatório:

Guerra, R., António, R., & Carvalho, P. (2026). Discurso de Ódio: Manual de Apoio e Glossário. ISBN: 978-989-584-300-8.

Cofinanciado pela União Europeia

Índice

1. Introdução 3

2. Esclarecimento sobre Responsabilidade e Uso Ético do Manual 4

2.1. Importância do Contexto 4
2.2. A Evolução da Linguagem nos Meios Digitais 5
2.3. Utilização Responsável e Ética deste Manual 5

PARTE I - Compreender o Discurso de Ódio 8

3. O que é Discurso de Ódio? 9

3.1. Definição 9
3.2. Diferentes Tipos de Discurso de Ódio 9
3.3. Formas Prevalentes de Discurso de Ódio Online em Portugal 10

4. De que Modo o Discurso de Ódio se Manifesta e Funciona 13

4.1. Desumanização e Estereótipos Negativos 14
4.2. Narrativas de Ameaça 16
4.3. Negação do Ódio e Inversão de Papéis 17
4.4. Eufemismos e Linguagem Codificada 19
4.5. Referências Socio-Históricas 21
4.6. Falácias 24
4.7. Emoções Negativas 25
4.8. Contexto, Interação e Agravamento 27
4.9. Polarização, Desinformação e a Normalização do Ódio 35

5. Impactos do Discurso de Ódio 39

5.1. Impactos nas Pessoas 39
5.2. Impactos nas Comunidades-Alvo e nas Testemunhas 40
5.3. Impactos na Sociedade e na Democracia 40

Cofinanciado pela União Europeia

VOLTAR AO ÍNDICE

PARTE II - Grupos-Alvo e Padrões de Ódio 42

6. Grupos-Alvo e Formas Comuns de Ódio em Portugal 43

6.1. Processos de Racialização 43
6.2. Xenofobia 44
6.3. Preconceito Sexual 46
6.4. Intersecionalidade 48

PARTE III - Prevenção e Enquadramento Legal 49

7. Práticas de Prevenção e Sensibilização 50

7.1. Sensibilização: Campanhas Públicas 50
7.2. Iniciativas de Educação e Formação 51
7.3. Contra-Discurso e Narrativas Positivas 52
7.4. Medidas ao Nível das Plataformas 57

8. Enquadramento Legal e Institucional em Portugal 60

Conclusão 67

Referências 68

Anexos 82

A - Reconhecer o Discurso de Ódio: Indicadores Linguísticos, Simbólicos e Contextuais 82

Cofinanciado pela União Europeia

1. Introdução

O **Manual de Apoio e Glossário** constitui uma ferramenta de referência estruturada e acessível, destinada a promover uma compreensão partilhada e fundamentada em evidência científica do discurso de ódio no contexto português. Esclarece conceitos-chave, explica como o discurso de ódio se manifesta no seio de ecossistemas mais amplos de ódio e reúne termos, expressões e referências simbólicas relevantes frequentemente associados ao discurso de ódio.

Este manual baseia-se nos conhecimentos desenvolvidos no âmbito do projeto kNOwHATE (CERV-2021-EQUAL 101049306) e aprofunda-os, incorporando resultados teóricos e empíricos atualizados e aperfeiçoando uma interpretação contextual do discurso de ódio, à luz da investigação mais recente. Alinha a terminologia e as perspetivas analíticas com a abordagem mais ampla e baseada em evidências desenvolvida no âmbito do COOPERHATE, mantendo-se, ao mesmo tempo, um recurso educativo destinado ao público em geral.

O Manual de Apoio e Glossário combina um quadro conceptual sólido com uma caracterização geral dos grupos-alvo e dos padrões recorrentes, visando: i) promover um entendimento partilhado e conhecimento baseado em evidência científica entre as organizações da sociedade civil, as autoridades públicas e o público em geral; ii) reforçar a coerência na formação, sensibilização e prevenção; iii) incentivar a cooperação entre os diferentes intervenientes relevantes envolvidos no combate às ameaças crescentes colocadas pelos discursos de ódio; iv) e melhorar o reconhecimento de padrões prejudiciais, apoiando respostas mais eficazes ao discurso de ódio e aos seus múltiplos impactos.

É importante reforçar que o Manual de Apoio e Glossário tem como objetivo promover uma interpretação informada, contextualizada e responsável da linguagem, e não funciona como uma lista de palavras ou expressões proibidas ou uma ferramenta de classificação jurídica.

Cofinanciado pela União Europeia

2. Esclarecimento sobre Responsabilidade e Uso Ético do Manual

2.1. Importância do Contexto

A linguagem não deve ser interpretada de forma isolada. Tal como sublinhado nos Estudos Críticos do Discurso, o significado é socialmente construído e inerentemente dependente do contexto, sendo moldado por fatores históricos, políticos e ideológicos (van Dijk, 2023; Wodak, 2015). Para além do conteúdo da própria mensagem, o significado e o potencial dano são moldados pelo contexto socio-histórico e discursivo mais amplo em que esta é produzida e disseminada, bem como pela intenção comunicativa e pelo posicionamento da pessoa que emite a mensagem, incluindo, entre outros, fatores como o tom utilizado e a repetição da mesma (e.g., Calderón et al., 2021; Theofilopoulos, 2024). O potencial de dano é ainda influenciado pelo meio de difusão e pela suscetibilidade do público (Benesch, 2023).

O discurso de ódio funciona frequentemente através da interação de vários aspetos, incluindo a utilização específica de recursos linguísticos e simbólicos, a perpetuação das narrativas ao longo do tempo e o foco em grupos sociais específicos (e.g., Mannocci et al., 2024). Por conseguinte, a inclusão de uma palavra, frase ou símbolo neste manual não implica automaticamente que a sua utilização constitua discurso de ódio; pelo contrário, deve ser entendida como um potencial indicador que requer uma interpretação contextual integrada.

O manual foi concebido para promover uma sensibilização e reflexão informada, reconhecendo, simultaneamente, que cada situação exige abordagens cautelosas, sensíveis ao contexto e às diferenças culturais e, quando pertinente, uma análise legal adequada (ver secção 8).

Cofinanciado pela União Europeia

2.2. A Evolução da Linguagem nos Meios Digitais

A linguagem evolui rapidamente, sendo particularmente acelerada e volátil em ambientes digitais. As expressões consagradas na língua podem mudar de significado, novas referências codificadas podem emergir e certos símbolos podem ser apropriados ou reinterpretados em diferentes plataformas e comunidades (e.g., Magu & Luo, 2018). Nestes ambientes, as pessoas que utilizam as redes sociais também recorrem frequentemente à ironia, ao humor, a eufemismos e a linguagem cifrada (i.e., "*dog whistles*"[1]) para obscurecer, atenuar ou ocultar o discurso de ódio (Baider & Constantinou, 2020). Estas formas indiretas ou dissimuladas de discurso de ódio dependem do conhecimento partilhado dentro de comunidades específicas, tornando-as mais difíceis de identificar à medida que o discurso evolui (Baider, 2022; Baider & Constantinou, 2020). A natureza dinâmica da linguagem nas redes sociais e as múltiplas interpretações que a mesma pode adquirir em função do contexto tornam a deteção de discurso de ódio uma tarefa extremamente complexa (Geetanjali & Kumar, 2025). Por outro lado, o discurso de ódio não se limita ao texto escrito. À medida que as plataformas digitais evoluem, o mesmo pode ser materializado sob a forma de múltiplos formatos e modos, incluindo imagens, *memes*, áudio e vídeo (Prabhu & Seethalakshmi, 2025). Este caráter multimodal torna ainda mais difícil a sua identificação e combate.

Por estas razões, este manual deve ser entendido como um recurso dinâmico. Embora reflita os conhecimentos e os resultados da investigação atuais, não pretende ser encarado como uma lista fechada e exaustiva de expressões de discurso de ódio.

2.3. Utilização Responsável e Ética deste Manual

Este manual destina-se a fins educativos, preventivos, de sensibilização e de reforço de competências. Não deve, por isso, ser utilizado como uma ferramenta autónoma para a deteção ou classificação de discurso de ódio, nem para quaisquer ações relacionadas, incluindo a remoção de conteúdos ou

[1] Mensagens criadas para serem compreendidas por um público específico, mantendo-se, ao mesmo tempo, ambíguas ou passíveis de negação para o público em geral (Åkerlund, 2021)

resposta legal. Importa alertar que os exemplos incluídos neste manual não devem ser reproduzidos ou divulgados, com o propósito de justificar, difundir ou normalizar conteúdos nocivos.

As autoras ponderaram cuidadosamente sobre a decisão de reproduzir, ou não, conteúdos de ódio e a necessidade de aumentar a sensibilização e o conhecimento sobre esta forma perigosa de discurso. Os exemplos apresentados no manual incluem linguagem de ódio, depreciativa e ofensiva, tanto direta como indireta, obtida a partir de dados das redes sociais disponíveis publicamente. Esta decisão visa minimizar o risco de danos e seguir o Código de Conduta da APA para a investigação, que recomenda que "*o risco de danos não deve ser superior ao da vida quotidiana, ou seja, os indivíduos não devem ser expostos a riscos superiores ou adicionais aos que enfrentam nos seus estilos de vida normais*". Expor as pessoas a discursos de ódio pode desencadear consequências negativas injustificadas, as quais são discutidas na secção 5, juntamente com as estratégias existentes para o combater (secção 7).

Assim, a menção ou discussão dos termos e exemplos descritos neste manual deve evitar a revitimização ou a exposição desnecessária das comunidades visadas.

Por fim, uma utilização responsável exige sensibilidade, proporcionalidade e respeito pelos direitos fundamentais. Isto inclui a liberdade de expressão, um direito fundamental garantido pela Constituição da República Portuguesa no nº 1 do artigo 37.º. No entanto, a liberdade de expressão não é um direito absoluto. Este direito está consagrado em tratados internacionais e regionais de direitos humanos, incluindo a jurisprudência da Corte Interamericana e da Comissão Interamericana de Direitos Humanos, nomeadamente no nº 5 do artigo 13º. Existem também limites ao direito à liberdade de expressão estabelecidos pela lei portuguesa, nomeadamente quando a mesma infringe os direitos de outrem ao promover a discriminação, a violência e o incitamento ao ódio contra uma pessoa ou grupo de pessoas por motivos relacionados com a sua etnia, origem, religião, nacionalidade ou orientação sexual, entre outros, tal como previsto no artigo 240º do Código Penal português (ver Secção 8 para

uma discussão detalhada). Distinguir entre casos evidentes de discurso de ódio e expressões potencialmente legítimas, incluindo a sátira ou o sarcasmo, requer um julgamento cuidadoso e sensível ao contexto.

Ao promover a consciência contextual e o compromisso ético, este manual pretende contribuir para respostas mais informadas ao discurso de ódio, salvaguardando simultaneamente os valores democráticos e a dignidade humana.

Cofinanciado pela União Europeia

PARTE I

Compreender o Discurso de Ódio

3. O que é Discurso de Ódio?

O manual baseia-se na abordagem conceptual e nos resultados empíricos do **projeto kNOwHATE**, nomeadamente nas características linguístico-discursivas e sociopsicológicas do discurso de ódio *online*, bem como nos seus principais grupos-alvo.

3.1. Definição

Não existe uma definição única e universalmente aceite de **discurso de ódio**; no entanto, para efeitos do COOPERHATE, baseamo-nos na definição de discurso de ódio desenvolvida no âmbito do kNOwHATE, que se baseia nas recomendações do Conselho da Europa e na literatura em psicologia social (Guerra et al., 2025). O discurso de ódio é, portanto, abordado como um fenómeno intergrupal, que visa grupos ou indivíduos devido à sua pertença percebida a determinados grupos sociais. Especificamente, definimos o discurso de ódio *online* "*como linguagem depreciativa, motivada por preconceitos, que difunde, incita, promove ou justifica o ódio, a exclusão e/ou a violência/agressão, visando grupos ou indivíduos com base na sua pertença a um grupo (e.g., características percebidas como etnia, raça, orientação sexual, etc.)*" (Guerra et al., 2025, p. 2). É importante referir que o discurso de ódio se manifesta de múltiplas formas, envolvendo tanto expressões explícitas e diretas como formas mais dissimuladas e subtis.

3.2. Diferentes Tipos de Discurso de Ódio

Seguindo a abordagem desenvolvida por Guerra e colegas (2025), é importante distinguir entre dois tipos de discurso de ódio com base na sua forma de expressão: direto e indireto.

No **discurso de ódio direto**, verifica-se uma propagação ou justificação explícita do ódio, da exclusão, da discriminação e/ou da violência contra um grupo-alvo ou um indivíduo com base na perceção da sua pertença a um determinado grupo. Normalmente, contém linguagem preconceituosa e inflamatória, insultos e termos depreciativos. O exemplo seguinte ilustra este

padrão, no qual o grupo-alvo é apresentado como um exogrupo[2] e desumanizado através da metáfora desumanizante "parasitas", a par do reforço explícito de estereótipos negativos ("não querem fazer nada"):

> *"Racismo o c@ralho! se não fossem esses parasitas da sociedade que não querem fazer nada, Portugal era um paraíso."*

O discurso de ódio indireto evita o uso de linguagem explicitamente depreciativa ou insultuosa; em vez disso, a expressão de propagação, promoção ou justificação do ódio, da exclusão, da discriminação ou da violência é, normalmente, implícita e subtil. Do ponto de vista semântico, o seu significado não é muitas vezes literal e deve ser inferido pragmaticamente, com base no contexto social e histórico (Assimakopoulos et al., 2017; Baider, 2022). O exemplo seguinte ilustra este padrão, envolvendo estratégias como a vitimização do endogrupo[3] e a negação do racismo.

> *"Se um branco se zanga com um branco, tudo bem! Se um branco se zanga com um preto, é racismo!"*

3.3. Formas Prevalentes de Discurso de Ódio *Online* em Portugal

Não existe uma monitorização oficial do discurso de ódio *online* em Portugal, tal como referido no relatório de 2025 da Comissão Europeia contra o Racismo e a Intolerância (ECRI): "***Embora exista uma falta de dados oficiais e desagregados sobre incidentes de discurso de ódio em Portugal**, vários relatórios credíveis de organizações da sociedade civil e outras instituições independentes apontam para um aumento acentuado do discurso de ódio no país*" (ECRI, 2025). Não obstante, os resultados de projetos científicos realizados em Portugal (kNOwHATE; HateCovid; Racismo e Xenofobia em Portugal: A Normalização do Discurso de Ódio na Esfera Pública da Internet) convergem com as principais conclusões da ECRI e de outras organizações da

[2] Grupo social ao qual não se pertence ou com o qual não se identifica.
[3] Grupo social a que se pertence e com o qual se identifica.

sociedade civil (e.g., #MigraMyths Discurso de Ódio e imigração em Portugal, 2025; 8º Exercício de Monitorização do Discurso de Ódio Online, 2026; Ciberviolência de base sexual em Portugal: perspetivas de crianças, jovens, professores e pessoal técnico especializado, 2026): o discurso de ódio *online* tornou-se cada vez mais prevalente e normalizado. É importante referir que os resultados também demonstram que, no contexto *online* português, as **formas indiretas (subtis ou dissimuladas) de discurso de ódio** são mais comuns do que o discurso de ódio direto (ou explícito) (e.g., Carvalho et al., 2023; Guerra et al., 2025). Este padrão observa-se independentemente dos grupos-alvo e das plataformas de redes sociais envolvidas. Tendências semelhantes foram identificadas em investigações anteriores realizadas noutros contextos geopolíticos e culturais (e.g., Baider, 2023).

Em vez de se basear principalmente em insultos explícitos ou apelos diretos à exclusão ou à violência, o discurso de ódio *online* é, na sua maioria, transmitido de formas subtis e sofisticadas, envolvendo, por exemplo, a negação do racismo ou a inversão de papéis (Guerra et al., 2025). Estas formas recorrem frequentemente a recursos retóricos, tais como a ironia, o sarcasmo, as perguntas retóricas, a metáfora e a hipérbole (Carvalho et al., 2023; Guerra et al., 2025).

A prevalência de expressões indiretas e subtis torna particularmente importante compreender como é que o discurso de ódio se manifesta além dos insultos ou ameaças explícitas. A secção seguinte analisa estratégias discursivas recorrentemente utilizadas tanto na expressão do discurso de ódio direto como indireto.

Leituras recomendadas:

Carvalho, P., Caled, D., Silva, C., Batista, F., & Ribeiro, R. (2023). The expression of hate speech against Afro-descendant, Roma, and LGBTQ+ communities in YouTube comments. *Journal of Language Aggression and Conflict*. Advance online publication. https://doi.org/10.1075/jlac.00085.car

Cofinanciado pela União Europeia

Guerra, R., Carvalho, P., Marques, C., Carmona, M., Sarroeira, R., Batista, F., Ribeiro, R., Fonseca, A., Moro, S., & Silva, C. (2025). Unpacking online hate speech in Portuguese social media: a social-psychological and linguistic-discursive approach. *Humanities and Social Sciences Communications, 12*(1). https://doi.org/10.1057/s41599-025-05392-9

Silva, C., & Carvalho, P. (2023). When can compliments and humour be considered hate speech? A perspective from target groups in Portugal. *Comunicação e sociedade, 43*, e023006. https://doi.org/10.17231/comsoc.43(2023).4135

4. De que Modo o Discurso de Ódio se Manifesta e Funciona

O discurso de ódio direto é frequentemente mais fácil de identificar, uma vez que envolve linguagem depreciativa explícita, como insultos ou declarações abertamente discriminatórias. Pelo contrário, o discurso de ódio indireto tende a ser mais subtil. Em vez de recorrer a formas diretas de hostilidade, manifesta-se frequentemente através de insinuações, estratégias de enquadramento discursivo e ambiguidade estratégica (Baider, 2023).

Nestes casos, compreender a mensagem requer atenção **às estratégias linguístico-discursivas subjacentes**: ou seja, analisar o que é sugerido em vez do que é afirmado diretamente. É importante, no entanto, sinalizar que o discurso direto e indireto podem coexistir na mesma mensagem (Guerra et al., 2025).

Tanto o discurso de ódio direto como o indireto recorrem a uma variedade de estratégias discursivas, incluindo a desumanização, os estereótipos negativos e as ameaças, bem como a expressão de emoções negativas, como o ódio e a raiva. Estas são frequentemente reforçadas através de recursos retóricos (e.g., metáforas, comparações e ironia verbal) e de estratégias argumentativas falaciosas, nomeadamente apelos ao medo e apelos à ação. No entanto, o discurso de ódio indireto recorre ainda a mecanismos adicionais, tais como a negação do ódio e a inversão de papéis (Carvalho et al., 2023; Guerra et al., 2025).

As secções seguintes descrevem alguns dos padrões mais recorrentes observados no discurso de ódio *online*, com base nas dimensões sociopsicológicas e linguístico-discursivas desenvolvidas pelo consórcio kNOwHATE (Carvalho & Guerra, 2023; Guerra et al., 2025). No Anexo A são apresentados exemplos concretos de marcadores linguísticos, simbólicos e contextuais associados a estes padrões.

Cofinanciado pela União Europeia

4.1. Desumanização e Estereótipos Negativos

Um dos mecanismos mais poderosos associados ao discurso de ódio é a **desumanização**, um processo através do qual o grupo-alvo é desprovido de traços humanos positivos, sendo retratado como menos humano e comparado a animais. Deste modo, as considerações morais normalmente atribuídas aos outros seres humanos deixam de se aplicar aos grupos visados (Borinca et al., 2023; Haslam, 2006), retratando-os como seres inferiores ao endogrupo[4]. A desumanização materializa-se frequentemente através de comparações e metáforas que reduzem as pessoas a animais, objetos ou autómatos (Bahador, 2023; Kteily & Bruneau, 2017), como ilustrado no exemplo abaixo:

> *"Ciganos são como javalis, são animais selvagens."*

Neste caso, as pessoas ciganas são comparadas a "javalis" e "animais selvagens", excluindo-as simbolicamente da categoria de seres humanos. A investigação demonstra que a desumanização deliberada não só reforça os estereótipos negativos, como aumenta a perceção de ameaça, reduz a empatia e justifica ações prejudiciais contra os grupos-alvo (Borinca et al., 2023; Ghenai et al., 2025). No contexto do discurso de ódio, a desumanização surge frequentemente associada a emoções negativas, como o medo ou a raiva (Ghenai et al., 2025; Guerra et al., 2025).

Os **estereótipos** são crenças partilhadas sobre os atributos e características percebidos de um grupo, podendo ser positivos ou negativos. Os estereótipos moldam a forma como as pessoas percebem, interpretam e respondem aos outros (Dovidio et al., 2010). Os **estereótipos negativos** envolvem a atribuição de crenças e características negativas, imprecisas e injustas a grupos sociais específicos e são utilizados para menosprezá-los ou humilhá-los através de generalizações falaciosas (Paz et al., 2020; Silva & Carvalho, 2023). No contexto do discurso de ódio, estes estereótipos surgem frequentemente associados a emoções negativas, como o medo ou a raiva em relação ao grupo visado,

[4] Grupo social a que se pertence e com o qual se identifica.

reforçando as perceções de inferioridade e subordinação social (Guerra et al., 2025; Papcunová et al., 2023).

No exemplo que se segue, a pessoa que emite a mensagem retrata a comunidade cigana como praticante de comportamentos inaceitáveis do ponto de vista cultural, ético e legal, incluindo o incesto e o casamento infantil. Deste modo, os estereótipos reforçam os preconceitos sociais, contribuindo para a estigmatização e desumanização deste grupo-alvo.

> *"Coitad@. Vcs deitam se com primos, tios, irmaos etc praticam incest# entre familia que Casam crianças com adultos e forçam nas a engravidar."*

É essencial compreender a desumanização e os estereótipos, uma vez que estes precedem frequentemente formas mais explícitas de hostilidade (e.g., Haslam & Loughnan, 2014). O reconhecimento precoce destes padrões permite identificar dinâmicas prejudiciais antes que estas se transformem em incitamento explícito ou violência.

Para saber mais sobre desumanização e estereótipos:

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In *The SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 3–29). London, England: Sage.

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology, 65* (1), 399-423. https://doi.org/10.1146/annurev-psych-010213-115045

Sousa, Y., & Cabecinhas, R. (2025). Estereótipos Sociais. *Psicologia Social, Comunicação e Cultura, 69*. https://doi.org/10.21814/uminho.ed.157.6

4.2. Narrativas de Ameaça

O discurso de ódio funciona frequentemente através de narrativas que retratam determinados grupos como uma **ameaça** ao endogrupo (Guerra et al., 2025). Estas ameaças podem ser enquadradas como **realistas**, sugerindo que um grupo constitui uma ameaça ao poder, aos recursos e ao bem-estar geral, à saúde e à segurança do endogrupo, ou como **simbólicas**, alegando que um grupo constitui uma ameaça aos valores culturais, religião, tradições, sistema de crenças, ideologia, moralidade ou visão do mundo (Stephan & Stephan, 2000).

O primeiro exemplo a seguir apresentado reflete uma forma de ameaça realista, apresentando o grupo-alvo como associado ao crime e à insegurança, através do recurso a estereótipos negativos. O segundo ilustra uma ameaça simbólica, retratando o grupo-alvo (i.e., a comunidade LGBTI+) como algo que compromete os valores culturais e a identidade coletiva, através da perceção de uso indevido ou apropriação de símbolos nacionais.

> *"Quando os ciganos assaltam, matam ou agridem as forças de segurança não se diz ciganos, diz-se jovens, quando é para elogiar já se diz ciganos."*

> *"Sinto-me ofendido quando a bandeira do meu país é "transvirada" alterada e usada para uma causa "esquerdista" Vão para a cama com quem quiserem, mas o respeito e a ordem nos nossos símbolos não podem ser adulterados ou alienados, querem respeito...tenham respeito!!!!"*

A investigação sobre as relações intergrupais demonstra que as ameaças percebidas estão fortemente associadas a intenções e comportamentos negativos em relação a exogrupos[5], incluindo a discriminação e a agressão (Stephan & Stephan, 2016). As ameaças intergrupais também distorcem a perceção: moldam a forma como os membros do endogrupo interpretam o

[5] Grupo social ao qual não se pertence ou com o qual não se identifica.

exogrupo, reforçam preconceitos e estereótipos negativos e comprometem a capacidade de estabelecer interações intergrupais equilibradas e construtivas (Stephan & Stephan, 2016).

No contexto português, os preconceitos raciais, xenófobos e sexuais têm sido frequentemente mobilizados através de narrativas que retratam as comunidades visadas como ameaças, tanto realistas como simbólicas, para a sociedade, normalmente baseadas em argumentos falaciosos, tais como apelos ao medo e apelos à ação (Almeida & Pereira, 2026; Carvalho et al., 2023; Guerra et al., 2025; ver secção 4.6).

Para saber mais sobre narrativas de ameaça:

Pereira, C. R., & Souza, L. E. C. D. (2016). Fatores legitimadores da discriminação: Uma revisão teórica. *Psicologia: Teoria e Pesquisa, 32*(2), e322222. https://doi.org/10.1590/0102-3772e322222

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination*, (pp. 23-46). Lawrence Erlbaum Associates Publishers.

Stephan, W. G., & Stephan, C. W. (2016). Intergroup Threats. *The Cambridge Handbook of the Psychology of Prejudice*, 131–148. https://doi.org/10.1017/9781316161579.00

4.3. Negação do Ódio e Inversão de Papéis

Outra estratégia comum no discurso de ódio é a **negação do ódio**. As pessoas que emitem a mensagem apresentam frequentemente as suas declarações como neutras, factuais ou meras expressões de preocupação, antecipando a rejeição de quaisquer acusações de preconceito. Esta estratégia ajuda a proteger a credibilidade da pessoa que emite a mensagem, a reduzir a perceção de ódio e a manter a legitimidade das suas afirmações (van Dijk, 1992). No exemplo abaixo, a pessoa apresenta uma declaração de isenção de responsabilidade para minimizar ou negar o racismo, que enuncia

imediatamente a seguir, num comentário abertamente racista sobre o desejo de aniquilação de uma raça específica da Europa.

> *"Omd eu juro, nao sou racista mas se tivesse de tirar uma raça da Europa seria esta."*

Intimamente relacionado a este fenómeno, destaca-se a **inversão de papéis**, uma das formas mais fortes de negação de racismo (van Dijk, 1992). Neste caso, os membros dos grupos socialmente dominantes ou maioritários (o endogrupo) apresentam-se como as verdadeiras vítimas de discriminação ou preconceito, retratando os grupos minoritários como injustamente privilegiados ou ameaçadores (Guerra et al., 2025). No exemplo abaixo, os membros do exogrupo (i.e., pessoas negras) tendem a ser representados como os intolerantes e o endogrupo (i.e., pessoas brancas) como as vítimas.

> *"99,9% das pessoas negras são racistas fanáticas; estão sempre à espera de uma oportunidade para magoar"*

No contexto português, estas estratégias estão frequentemente associadas à influência persistente do **Lusotropicalismo**: uma ideologia que retrata a sociedade portuguesa como historicamente tolerante e com uma capacidade única para estabelecer relações interculturais harmoniosas (Bastos, 2019; Valentim, 2011). As referências a supostas "competências especiais" na gestão da diversidade são utilizadas para negar a existência de racismo ou de desigualdade sistémica, minimizando assim as experiências das comunidades marginalizadas e reformulando as discussões sobre discriminação como exageradas ou injustificadas.

Compreender estes mecanismos é essencial, pois muitas vezes parecem mais subtis e socialmente aceitáveis do que os insultos explícitos. As narrativas de ameaça, negação e inversão de papéis podem normalizar a exclusão, mantendo ao mesmo tempo uma aparência de razoabilidade, o que as torna particularmente poderosas no debate público contemporâneo.

Para saber mais sobre negação do ódio, inversão de papéis e narrativas coloniais:

Valentim, J. P., & Heleno, A. M. (2018). Luso-tropicalism as a social representation in Portuguese society: Variations and anchoring. *International Journal of Intercultural Relations, 62*, 34-42. https://doi.org/10.1016/j.ijintrel.2017.04.013

van Dijk, T. A. (1992). Discourse and the Denial of Racism. *Discourse & Society, 3*(1), 87–118. https://doi.org/10.1177/0957926592003001005

Van Nieuwenhuyse, K., Bentrovato, D., & Valentim, J. P. (2026). The colonial past and/in history textbooks: a literature review. *Current Opinion in Psychology, 68*,102264. https://doi.org/10.1016/j.copsyc.2025.102264 .

4.4. Eufemismos e Linguagem Codificada

O discurso de ódio é frequentemente transmitido através de **eufemismos e linguagem codificada,** estratégias linguísticas que permitem às pessoas que emitem a mensagem expressar hostilidade, enquanto suavizam a sua aparência ou evitam a moderação e a sanção social (e.g., Magu & Luo, 2018). Embora estas estratégias possam ser utilizadas com objetivos idênticos, diferem na forma como transmitem o significado oculto.

Os **eufemismos** são utilizados para substituir expressões abertamente discriminatórias, de modo a conferir maior neutralidade, ou mesmo humor, ao discurso (e.g., Magu & Luo, 2018). Isto pode tornar as ideias excludentes menos explícitas e, por conseguinte, mais socialmente aceitáveis, permitindo que o significado pretendido seja compreendido por quem o reconhece. Por exemplo, a expressão “mandá-lo para casa” no caso abaixo pode ser entendida como uma referência eufemística à expulsão do indivíduo mencionado do país:

> *“Afinal o homem [Ativista Negro antirracista] tem que promover e denunciar o racismo, senão vai para as obras, ou com sorte, servir à mesa. Fica mais barato mandá-lo para casa, apoio.”*

No entanto, os eufemismos raramente ocorrem isoladamente. Neste exemplo em concreto, a aparente suavização é estratégica, reforçando, de forma indireta e humorística, ideias discriminatórias contra o indivíduo visado.

As **palavras-código** (i.e., linguagem codificada) são expressões que veiculam um significado codificado, partilhado apenas pelos membros de determinadas comunidades, incluindo redes extremistas ou de extrema-direita, sendo utilizadas para transmitir mensagens ou posições ideológicas de forma implícita, permitindo contornar normas sociais, mecanismos de moderação ou críticas explícitas (Åkerlund, 2021; Calderón et al., 2021; Magu & Luo, 2018). À primeira vista, estes termos podem parecer inofensivos. No entanto, em contextos específicos, remetem para narrativas de conspiração, hierarquias raciais ou hostilidade contra grupos específicos. A ambiguidade inerente a estes termos permite às pessoas que emitem a mensagem contornar os sistemas de deteção e negar intenções de ódio, caso sejam questionados. No exemplo abaixo, a pessoa recorre ao termo "*Google*" para se referir a comunidades afrodescendentes, sendo o significado explicitado através da *hashtag* #voltemparaaafrica:

> *"@utilizador Estou farto destes googles inúteis »#voltemparaafrica"*
>
> *Palavra-código: Google*
>
> *Palavra real: pessoas negras*

Nos ambientes digitais, a linguagem codificada evolui rapidamente. À medida que certos termos se tornam reconhecidos e sujeitos a moderação, surgem novas variações lexicais ou substitutos dos mesmos. Isto realça a importância da interpretação contextual: identificar discurso de ódio requer atenção não só à interpretação literal do texto, mas sobretudo aos padrões de utilização, e à sua interpretação pragmática num contexto discursivo mais amplo (e.g., Calderón et al., 2021).

Cofinanciado pela União Europeia

Por conseguinte, reconhecer referências codificadas é essencial para compreender como o discurso de ódio contemporâneo, frequentemente expresso de forma subtil, indireta, irónica ou codificada, se adapta, persiste e circula.

Para saber mais sobre eufemismos e palavras-código:

Magu, R., & Luo, J. (2018). Determining Code Words in Euphemistic Hate Speech Using Word Embedding Networks. *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*. https://doi.org/10.18653/v1/w18-5112

4.5. Referências Socio-Históricas

O discurso de ódio pode recorrer a interpretações seletivas da história para legitimar a exclusão no presente. As **referências socio-históricas** podem invocar a nostalgia colonial, a retórica autoritária ou da era da ditadura, a negação ou distorção de acontecimentos históricos bem documentados, podendo ser mobilizadas em teorias da conspiração (e.g., Almeida et al., 2023; Baider, 2022; NCTV, 2024; Papcunová et al., 2023).

Estas referências apresentam, frequentemente, o passado como uma época de ordem, unidade ou grandeza nacional, contrastando-o, implícita ou explicitamente, com um presente descrito como enfraquecido pela diversidade, pela migração ou pela mudança social. Ao idealizar determinados períodos históricos, tais narrativas podem normalizar visões hierárquicas do mundo, a que alguns grupos pertencem "naturalmente", enquanto outros não (e.g., Couperus et al., 2023; Homolar & Löfflmann, 2021; Kentmen-Cin, 2025). A investigação mostra que estas narrativas evocam tipicamente um passado glorioso e vitorioso, enquadrando depois a sua perda como uma humilhação partilhada dirigida contra aqueles considerados como não pertencentes ao "verdadeiro" povo (i.e., o endogrupo; Homolar & Löfflmann, 2021).

Noutros casos, estas referências baseiam-se na negação ou minimização histórica, como a minimização da discriminação, do racismo, da violência colonial ou da opressão sistémica do passado, para deslegitimar as reivindicações contemporâneas de igualdade e reconhecimento (Castelo, 2021; van Dijk, 1992). No caso de Portugal, estas referências estão frequentemente ligadas ao apoio ao Lusotropicalismo (ver secção 4.3).

O uso de referências históricas no discurso de ódio geralmente mobiliza estratégias discursivas como a negação do racismo (ver secção 4.3), o que neutraliza eficazmente a resistência (van Dijk, 1992): quando a injustiça passada e presente é negada, os apelos a políticas antirracistas ou à reparação legal tornam-se desnecessários. Por exemplo, o discurso populista de direita tem mobilizado a reabilitação da glória nacional, minimizando crimes coletivos (e.g., o colonialismo) através da reavaliação seletiva ou da fabricação de relatos históricos (Couperus et al., 2023).

A reinterpretação de acontecimentos históricos passados, que justifica a desconfiança, a exclusão ou a hostilidade em relação a grupos específicos (e.g., o Holocausto; Williams, 2022), está frequentemente presente em teorias da conspiração. Estas narrativas recorrem à memória coletiva e a referências históricas, reutilizando histórias de longa data de desconfiança, conflito, marginalização e traição institucional entre grupos para conferir coerência a novas alegações sobre ameaças atuais (Wagoner et al., 2026). Assim, as teorias da conspiração criam uma sensação de continuidade entre ameaças passadas e presentes, reforçando o medo e a desconfiança.

O exemplo abaixo ilustra a interseção entre o pensamento conspirativo e o discurso de ódio através da reinterpretação da história colonial. Em vez de reconhecer as assimetrias históricas, a declaração reinterpreta o colonialismo como um processo benevolente, atribuindo “infraestruturas”, “tecnologia” e “cultura” à presença colonial portuguesa. Isto constitui uma forma de revisionismo histórico, típica das narrativas conspirativas que reaproveitam a memória coletiva para construir relatos simplificados das relações passadas e presentes (Inwood & Zappavigna, 2023; Wagoner et al., 2026).

Cofinanciado pela União Europeia

> *"Se existisse algum tipo de reparação, deveria ser a indemnização a Portugal pelas antigas colónias, por todas as infraestruturas e tecnologias deixadas pelos portugueses e pela alma e cultura dadas ao mundo.*
>
> *Não vou permitir mais que envergonhem o nosso país!"*

Este exemplo combina uma reinterpretação histórica baseada em teorias da conspiração com estratégias de discurso de ódio, tais como a inversão da culpa coletiva, a glorificação do endogrupo e a exclusão implícita do exogrupo, normalizando, assim, uma narrativa excludente sem recorrer à linguagem depreciativa explícita.

Compreender o uso de referências socio-históricas é importante, uma vez que estas criam uma camada de significado mais profunda ao discurso de ódio. Ao ancorar o preconceito em memórias aparentemente partilhadas ou na identidade nacional, tais narrativas podem fazer com que a exclusão pareça justificada, inevitável ou mesmo patriótica. Reconhecer como a história é mobilizada ajuda a identificar os quadros discursivos mais amplos dentro dos quais o discurso de ódio opera. Exemplos de tais marcadores simbólicos e historicamente enraizados podem ser encontrados no Anexo A.

Para saber mais sobre referências socio-históricas:

Almeida, P., Pereira, J., & Candido, D. (2023). Online hate speech on social media in Portugal: extremism or structural racism? *Social Identities, 29*(5), 419–435. https://doi.org/10.1080/13504630.2024.2324277

Castelo, C. (2021). *"Portuguese non racism": On the historicity of an invented tradition*. https://cesa.rc.iseg.ulisboa.pt/afroport/portuguese-non-racism-on-the-historicity-of-an-invented-tradition/

Wagoner, B., Jørgensen, M. S., & Pahuus, K. (2026). Conspiracy theories through the lens of collective memory. *Current Opinion in Psychology, 68*, 102227. https://doi.org/10.1016/j.copsyc.2025.102227

Cofinanciado pela União Europeia

4.6. Falácias

O discurso de ódio recorre frequentemente a estratégias argumentativas falaciosas que moldam a forma como o público interpreta e responde a determinadas situações. Frequentemente, as mensagens recorrem a **falácias** e a uma linguagem emocionalmente carregada, afastando-se dos padrões de uma discussão equilibrada e crítica. Entre outras estratégias de argumentação falaciosa, são particularmente relevantes, no contexto do discurso de ódio, o *apelo ao medo* e o *apelo à ação* (Carvalho et al., 2023; Guerra et al., 2025).

O **apelo ao medo** é uma estratégia persuasiva que não envolve uma ameaça explícita, mas sim um aviso de que ocorrerá um resultado negativo se o destinatário não adotar a ação recomendada (explícita ou implicitamente) (Tindale, 2007), conforme ilustrado no exemplo seguinte:

> *"Minorias? Somos nós a minoria, se olharmos para as coisas numa perspetiva global... Passámos de 30% da população mundial em 1930 para menos de 11%, e se a tendência se mantiver, daqui a 30 anos seremos menos de 7%...."*

Neste comentário, as estatísticas demográficas são utilizadas de forma seletiva para construir uma narrativa de declínio demográfico e perda do estatuto de maioria. Os dados populacionais são reinterpretados como um aviso implícito de futura marginalização do endogrupo, suscitando assim insegurança e ansiedade por meio da projeção de uma inversão numérica.

O **apelo à ação** é uma estratégia intimamente relacionada com a anterior, envolvendo um pedido (explícito ou implícito) de uma resposta comportamental com o objetivo de reverter uma situação percebida como negativa. Normalmente, é transmitido através de um tom carregado de emoção, como ilustrado no exemplo seguinte:

> *"Agradeçam à União Europeia por destruir a Europa e Portugal! Deixem de votar em políticos de esquerda!"*

Aqui, imperativos diretos ("Deixem de votar") e linguagem avaliativa constroem uma narrativa de declínio social atribuída a atores políticos e instituições supranacionais associados a exogrupos. O enquadramento emotivo e a atribuição de culpa servem para mobilizar o público no sentido de uma mudança comportamental concreta, nomeadamente no contexto eleitoral, assente num discurso de deterioração política e cultural.

Para saber mais sobre falácias:

Carvalho, P., Caled, D., Silva, C., Batista, F., & Ribeiro, R. (2023). The expression of hate speech against Afro-descendant, Roma, and LGBTQ+ communities in YouTube comments. *Journal of Language Aggression and Conflict*. Advance online publication. https://doi.org/10.1075/jlac.00085.car

Tindale, C. W. (2007). *Fallacies and Argument Appraisal*. Cambridge University Press. https://doi.org/10.1017/cbo9780511806544

4.7. Emoções Negativas

As emoções desempenham um papel fundamental na moldagem das atitudes e comportamentos intergrupais, estando diferentes emoções associadas a tendências motivacionais distintas, tais como o confronto ou o evitamento (Cottrell & Neuberg, 2005; Mackie & Smith, 2015). Certas **emoções negativas**, como o ódio e a raiva, desempenham um papel central na formação, expressão e justificação do discurso de ódio *online* (Ghenai et al., 2025; Guerra et al., 2025).

O **ódio** é uma emoção negativa poderosa ou, com o tempo, um sentimento mais estável que surge quando indivíduos ou grupos são percebidos como

tendo intenções maliciosas de prejudicar o endogrupo (Fischer et al., 2018). No discurso de ódio, esta emoção reflete-se na representação dos grupos-alvo como inerentemente imorais, perigosos e imutáveis (Guerra et al., 2025). O ódio está frequentemente associado a processos de desumanização, silenciamento e ameaças realistas e simbólicas, sendo também um forte preditor de intenções e comportamentos prejudiciais. De facto, o ódio motiva ações destinadas não só a ferir, mas também a excluir, humilhar ou mesmo eliminar o alvo a nível psicológico, social ou físico (Fischer et al., 2018). Neste sentido, o ódio vai além da hostilidade momentânea, envolvendo, frequentemente, um desejo de punição, vingança ou a remoção da fonte percebida de ameaça.

Conforme ilustrado nos exemplos abaixo, esta emoção é reforçada através de várias estratégias discursivas, incluindo a humilhação, a desumanização, apelos explícitos à ação (nomeadamente, a exclusão ou mesmo a aniquilação do grupo-alvo) e manifestações de intensa hostilidade:

> *"EXPULSA-LOS PRA FORA DO PAÍS COMO UNS CÃES VADIOS QUE SÃO!!! NOJO DESTA GENTE!!! CHEGAAAAAA"*
>
> *"MATAR O CARALHO..VAI PRÁ TUA TERRA.MAS AFINAL DE QUE TRIBO ÉS..[Ativista Negro antirracista] .E NÂO HÁ QUÊM LHE. TRATE DA SAÚDINHA."*

A **raiva** é uma emoção que surge quando as ações de um exogrupo são percebidas como injustas, injustificadas ou contrárias às normas sociais estabelecidas (Fischer et al., 2018). Ao contrário do ódio, que muitas vezes tem origem na perceção de uma intenção maliciosa inerente, a raiva é normalmente dirigida a comportamentos e situações específicos.

No discurso de ódio, a raiva reflete-se em mensagens que atribuem a responsabilidade ao exogrupo, apresentando-o como um obstáculo aos objetivos, valores ou bem-estar do endogrupo (Guerra et al., 2025). Estas podem incluir ameaças percebidas aos recursos económicos, à ordem social ou aos direitos. A raiva motiva frequentemente o confronto, a crítica ou tentativas de corrigir ou desafiar os outros.

Esta emoção pode envolver a intenção de causar dano, mas, ao contrário do ódio, está normalmente associada a um desejo de responder a uma injustiça percebida ou de eliminar obstáculos, em vez de eliminar o alvo. Tal como ilustrado no exemplo abaixo, a raiva está assim associada a narrativas de ameaça realista e simbólica, bem como a estratégias discursivas como a inversão de papéis (Carvalho & Guerra, 2023; Guerra et al., 2025).

> *"Força [Radical de extrema-direita], fizeste muito bem. Essa escumalha vai ter que aprender a respeitar os Portugueses e parar de gozar com os cidadãos de bem. Viva Portugal".*

Para saber mais sobre ódio e raiva:

Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why we hate. *Emotion Review, 10*(4), 309-320. https://doi.org/10.1177/1754073917751229

Halperin, E. (2011). The emotional roots of intergroup aggression: The distinct roles of anger and hatred. In P. R. Shaver & M. Mikulincer (Eds.), *Human aggression and violence: Causes, manifestations, and consequences* (pp. 315–331). American Psychological Association. https://doi.org/10.1037/12346-017

4.8. Contexto, Interação e Agravamento

A comunicação *online* organiza-se através de trocas conversacionais, incluindo cadeias de mensagens, respostas, partilhas, "gostos", *hashtags* e imagens partilhadas. Assim, o significado e o potencial impacto negativo do discurso de ódio *online* tornam-se frequentemente mais claros quando analisados no contexto conversacional mais amplo em que se insere (Calderón et al., 2021).

Por exemplo, um comentário ambíguo ou potencialmente inofensivo pode ser desambiguado quando lido em conjunto com mensagens anteriores, referências repetidas ou respostas coordenadas. Estes padrões de interação podem revelar hostilidade codificada, significados em constante mudança ou

formas indiretas de agressão que permanecem opacas quando consideradas isoladamente (e.g., Calderón et al., 2021; Fonseca et al., 2024).

Os indicadores seguintes podem ajudar a identificar como o discurso de ódio se desenvolve, se intensifica e se propaga:

- **Padrões de repetição:** o uso repetido de insultos, *memes* (e.g., *memes* de extrema-direita; NCTV, 2024) ou linguagem codificada/palavras-código (e.g., Calderón et al., 2021; Magu & Luo, 2018), incluindo repetições de letras e elementos simbólicos (e.g., Mansur et al., 2024), pode reforçar narrativas de ódio e indicar um ataque coordenado ou persistente a grupos (e.g., Arce-García et al., 2025; Ghenai et al., 2025; Magu & Luo, 2018; Mansur et al., 2024).

> *Por exemplo, verificou-se que os emojis de raiva, riso ou vómito, que aparecem frequentemente em sequências de três ou mais elementos, ou em combinações padronizadas, são comuns no discurso de ódio dirigido às minorias de género (Theofilopoulos, 2024).*

- **Publicações coordenadas**: A existência de várias contas ou de membros de um grupo que partilham conteúdos de ódio semelhantes ou sincronizados (e.g., através de *hashtags*; Mannocci et al., 2024) pode indicar uma amplificação organizada ou atividade de *bots*[6], aumentando a visibilidade e o alcance das mensagens de ódio (e.g., Arce-García et al., 2025; Pontes et al., 2024).

> *Um exemplo recente no contexto português é a disseminação de narrativas específicas de desinformação (e.g., "Portugal está a ser invadido", "islamização de Portugal") durante as eleições de 2024 e 2025, que envolveram uma coordenação entre contas anónimas para amplificar conteúdos alinhados com o discurso do partido Chega. As contas apresentavam uma forte sobreposição de seguidores e partilhas*

[6] Os *bots* são algoritmos desenvolvidos para gerar conteúdo automaticamente e interagir com utilizadores humanos (Uyheng & Carley, 2020).

sincronizadas, tendo alcançado centenas de milhares de visualizações em momentos-chave (Cardoso et al., 2025).

- **Sinais de rede**: Os padrões de interação entre as pessoas que utilizam as redes sociais podem ajudar a identificar como o discurso de ódio se propaga e ganha visibilidade *online*. Em vez de nos concentrarmos apenas nas mensagens individuais, é importante considerar a forma como o conteúdo circula pelas redes, por exemplo, através da partilha repetida, da publicação coordenada ou da atividade de contas de grande visibilidade (isto pode ser identificado através da Análise de Redes Sociais[7]). A investigação mostra que os conteúdos de ódio tendem a organizar-se em comunidades fortemente interligadas (i.e., *clusters*), o que facilita a sua persistência e disseminação nos espaços *online* (Ghenai et al., 2025). Estas estruturas interligadas permitem que narrativas prejudiciais circulem mais facilmente, reforçando crenças partilhadas e aumentando o seu alcance. Analisar padrões de interação mais amplos, tais como quem partilha o conteúdo, com que frequência é repetido e de que forma os grupos de utilizadores se relacionam entre si, pode ajudar a identificar as dinâmicas complexas da propagação do discurso de ódio *online* (Fonseca et al., 2024; Pontes et al., 2024).

- **Escalada no discurso**: O discurso de ódio pode intensificar-se ao longo do tempo através de interações repetidas, conversas e dinâmicas de grupo, evoluindo progressivamente de expressões subtis ou indiretas (por exemplo, estereótipos negativos) para hostilidade explícita, apelos à ação ou incitamento à violência. Esta escalada é ilustrada na Pirâmide do Ódio (Williams, 2022, adaptado de Gordon Allport, The Nature of Prejudice, 1954), onde, em condições específicas, a rejeição verbal evolui para a exclusão social, a violência e o genocídio.

[7] Análise de Redes Sociais (ARS) é uma abordagem metodológica e uma metodologia de investigação que analisa em profundidade os padrões de relações entre atores interligados (i.e., indivíduos, grupos, organizações ou outras unidades). A ARS representa estas relações sob a forma de redes, nas quais os "nós" representam os atores e as "arestas" simbolizam as ligações entre eles.

- **Dinâmicas de escalada e correlações *online-offline***: Embora o ódio *online* não cause diretamente violência no mundo real (*offline*), existem atualmente evidências sólidas de como a violência *online* e no mundo real se correlacionam, e de como a disseminação do discurso de ódio nas redes sociais antecede incidentes de crimes de ódio no mundo físico (e.g., Calderón et al., 2024; Madriaza et al., 2025; Williams et al., 2020). Por exemplo, estudos mostram que publicações antirrefugiados na página do *Facebook* do partido de extrema-direita Alternative für Deutschland (AfD) estavam associadas a crimes violentos *offline* contra imigrantes na Alemanha (Müller & Schwart, 2021). No Reino Unido, o discurso de ódio *online* dirigido às comunidades negras e muçulmanas correlacionou-se com crimes de ódio racial e religioso em Londres e precedeu os ataques (Williams et al., 2020). Conclusões semelhantes foram encontradas nos EUA e em Espanha. Nos Estados Unidos, verificou-se que o discurso de ódio *online* dirigido a comunidades racializadas e LGBTI+ correlacionava-se com crimes de ódio *offline* contra estes grupos (Bozhidarova et al., 2023). Da mesma forma, uma análise de publicações no *Twitter* e no *Facebook* realizadas em Espanha, entre 2016 e 2018, revelou correlações entre o discurso de ódio *online* e crimes de ódio dirigidos a comunidades de migrantes e pessoas LGBTI+ (Calderón et al., 2024).

- **Fatores Contextuais que Impulsionam o Discurso de Ódio**

  - **Fatores Desencadeadores Sociais, Políticos e Relacionados com Crises:** Os grandes acontecimentos sociais podem funcionar como catalisadores do discurso de ódio, observando-se aumentos de discurso de ódio na sequência de ataques terroristas, bem como durante momentos políticos altamente polarizadores, como o *Brexit* no Reino Unido ou os Acordos de Paz na Colômbia, frequentemente amplificados pela desinformação (Madriaza et al., 2025). Em momentos de crise, as sociedades procuram explicações ou bodes expiatórios, e os grupos socialmente minoritários podem tornar-se alvos convenientes para a culpa, particularmente quando a

desinformação ou narrativas de conspiração circulam amplamente (Šrol et al., 2022).

Por exemplo, isto foi especialmente visível durante a pandemia da COVID-19, descrita pelas Nações Unidas como tendo desencadeado um "tsunami de ódio" online, marcado pela proliferação de conteúdos xenófobos, misóginos, transfóbicos e conspirativos (Vergani et al., 2024).

- **Ambientes Informativos e Amplificação de Crises:** Estas dinâmicas podem ser mais pronunciadas em contextos afetados por conflitos ou de alto risco, onde instituições frágeis, acesso limitado a informação fiável e restrições à liberdade de imprensa criam ambientes informativos vulneráveis. Nesses contextos, a incerteza acentuada e a baixa confiança facilitam a rápida propagação de rumores, informação incorreta, desinformação (ver secção 4.9) e discurso de ódio, muitas vezes, intensificando as divisões existentes e amplificando as suas consequências no mundo real (Wardle, 2024).
- **Instrumentalização Política do Discurso de Ódio:** O ódio é também frequentemente usado ou estrategicamente deturpado para fins políticos (Bilewicz & Soral, 2020; Williams, 2022). O debate político pode tornar-se um poderoso catalisador do discurso de ódio, particularmente durante períodos eleitorais, reformas legislativas ou controvérsias públicas envolvendo temas como migração, igualdade de género ou direitos de grupos estigmatizados (Kentmen-Cin, 2025; Said-Hung et al., 2023). A investigação mostra que o ódio *online* é utilizado estrategicamente para promover objetivos políticos, tais como mobilizar apoio, reforçar as fronteiras entre endogrupos e exogrupos e influenciar o discurso público e as políticas (Kentmen-Cin, 2025).

Por exemplo, nos Estados Unidos, os desenvolvimentos políticos contemporâneos têm sido associados a uma crescente desumanização das

pessoas imigrantes, da classe trabalhadora e das mulheres, tanto nas redes sociais como nos meios de comunicação social (Karantzas & Simpson, 2026).

Em Portugal, uma análise de mais de 10 000 publicações de figuras-chave do partido político de extrema-direita CHEGA na plataforma X, entre 2019 e 2024, revelou que a prevalência de temas relacionados com a identidade e a imigração coincidia com os ciclos eleitorais. Isto sugere a utilização estratégica de conteúdos xenófobos com forte carga emocional para fins de mobilização política (Cardoso et al., 2025).

Em geral, as declarações de ódio proferidas por políticos podem dessensibilizar gradualmente as pessoas em relação à linguagem depreciativa, tornando-as mais propensas a aceitar os estereótipos e a discriminação, contribuindo assim para um clima mais generalizado de hostilidade intergrupal. Esta retórica pode ainda alterar as normas do que configura um comportamento aceitável, enfraquecendo as normas sociais existentes que protegem grupos socialmente minoritários (Bilewicz & Soral, 2020). As plataformas digitais amplificam ainda mais estas narrativas, permitindo que os atores de extrema-direita expandam o seu alcance para além das suas redes imediatas e contribuam para a criação de "caixas de ressonância" (i.e., *echo chambers;* ver secção 4.9) políticas que ajudam a difundir e a normalizar pontos de vista mais radicais (Kentmen-Cin, 2025; Williams et al., 2020).

- **Fatores Desencadeadores do Discurso de Ódio *Online* Relacionados com o Desporto:** Os eventos desportivos podem também tornar-se fatores que desencadeiam o discurso de ódio *online*. As rivalidades entre equipas, a forte identificação com o grupo e o envolvimento emocional podem contribuir para a utilização de linguagem agressiva ou discriminatória (Kearns et al., 2022; Miranda et al., 2023). Embora grande parte da hostilidade relacionada com o desporto possa ser enquadrada como rivalidade, esta pode interligar-se com o racismo, a xenofobia, a homofobia e a misoginia (Kearns et

al., 2022; Miranda et al., 2023; Montesinos-Cánovas et al., 2023). De facto, o discurso de ódio pode manifestar-se durante eventos desportivos, como no caso do jogador do FC Porto, Moussa Marega, que abandonou o campo em 2020 após ter sido alvo de cânticos racistas por parte dos adeptos adversários, ou, mais recentemente, os insultos racistas e homofóbicos dirigidos a Vinicius Júnior durante um jogo entre o Benfica e o Real Madrid, que levou a uma suspensão por conduta discriminatória.

As redes sociais ampliaram estas dinâmicas para além dos estádios, e os eventos desportivos têm sido repetidamente identificados como fatores desencadeadores de picos de ódio *online* (e.g., Carvalho et al., 2022).

Por exemplo, a final da UEFA EURO 2020 entre a Itália e a Inglaterra foi seguida por uma onda de insultos racistas dirigidos a três jogadores negros da seleção inglesa, um padrão também observado durante torneios subsequentes, como o EURO Feminino de 2022 e o Campeonato do Mundo de 2022.

Estes incidentes ilustram como momentos de grande visibilidade e intensidade emocional podem rapidamente traduzir-se em hostilidade direcionada *online*. Dinâmicas semelhantes também surgiram nos debates em torno da participação de atletas transgénero no desporto, onde as discussões *online* são frequentemente acompanhadas por narrativas transfóbicas, essencialismo biológico e retórica excludente (Murib, 2022).

Por exemplo, durante os Jogos Olímpicos de Paris 2024, onde o caso de Imane Khelif gerou uma onda de discurso transfóbico e racializado em todas as plataformas (Taha & Sailofsky, 2025).

Em conjunto, estes indicadores revelam que o discurso de ódio não se resume apenas ao que é dito, mas também à forma como circula, evolui e ganha

Cofinanciado pela União Europeia

significado ao longo do tempo. A interação entre a expressão individual e os contextos sociais mais amplos molda o discurso de ódio. O debate político, as crises e até mesmo os eventos desportivos, bem como as estruturas das plataformas digitais, influenciam de forma conjunta a forma como a hostilidade surge, circula e se intensifica. Reconhecer estes ecossistemas permite respostas mais informadas e sensíveis ao contexto. Assim, abordar o discurso de ódio de forma eficaz requer não só identificar expressões prejudiciais, mas também compreender os contextos em que estas surgem, circulam e se normalizam.

De acordo com recomendações recentes, que salientam que a análise do discurso de ódio *online* deve ir além da simples deteção, recomendamos que o discurso de ódio seja avaliado tendo em conta não só a presença de características linguísticas específicas, mas também a sua **combinação com fatores extralinguísticos,** incluindo**, repetição, agravamento ao longo do tempo e contextos mais amplos** (e.g., dinâmicas da plataforma, eventos desencadeadores).

Para saber mais sobre contexto, interação e escalada:

Bilewicz, M., & Soral, W. (2020). Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology, 41*(S1), 3–33. https://doi.org/10.1111/pops.12670

Magu, R., & Luo, J. (2018). Determining Code Words in Euphemistic Hate Speech Using Word Embedding Networks. *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*. https://doi.org/10.18653/v1/w18-5112

Williams, M. (2022). *The Science of Hate: How Prejudice Becomes Hate and What We Can Do to Stop It*. Faber & Faber, Ltd

4.9. Polarização, Desinformação e a Normalização do Ódio

À medida que o discurso de ódio circula, se repete e intensifica nas interações, pode acabar por se integrar em narrativas mais amplas que influenciam a forma como os grupos e as questões sociais são compreendidos. Nos ambientes *online*, o discurso de ódio surge frequentemente associado a processos de **polarização** e **desinformação**, incluindo **teorias da conspiração**, contribuindo para um discurso público cada vez mais dividido e hostil.

A **polarização** pode assumir diferentes formas nos contextos *online* e sociais. Pode referir-se a posições extremadas por parte de atores individuais (polarização individual), a grupos que reforçam pontos de vista mais radicais no âmbito de discussões internas do grupo (polarização de grupo), a grupos opostos que se tornam mais hostis e distantes uns dos outros (polarização intergrupal ou política) ou a divisões ideológicas e temáticas crescentes, em torno de questões sociais ou políticas (polarização temática). Assim, a polarização não deve ser entendida como uma condição fixa, mas sim como um processo dinâmico multinível moldado pelo desacordo e pela interação social (Bliuc et al., 2024).

A polarização intergrupal, em particular, tende a reforçar a dinâmica "nós contra eles", isto é, a construção do "outro" (*othering*), um processo em que indivíduos ou grupos são retratados como fundamentalmente diferentes, inferiores ou mesmo ameaçadores para o endogrupo (Smith et al., 2024). Este processo é uma estratégia discursiva frequentemente utilizada no discurso de ódio, tal como descrito no modelo ideológico proposto por van Dijk (1992), em que os indivíduos que promovem ou difundem discurso de ódio recorrem a uma autoapresentação positiva (*Nós*) e a uma apresentação negativa do outro (*Outros*). Nesses contextos, o recurso a estereótipos negativos, narrativas de ameaça e representações desumanizantes é frequente e mais difícil de contestar (e.g., Morales et al., 2025).

A polarização está também relacionada com a formação de **"caixas de ressonância"** (i.e., "*eco-chambers*", ou "bolhas de filtro", i.e., "*filter bubbles*"), ou seja, comunidades de utilizadores com opiniões semelhantes, nas quais a informação circula praticamente sem contestação e a exposição a pontos de

vista divergentes é limitada, em parte devido a algoritmos de classificação que reduzem a exposição a perspetivas contrárias (e.g., Gallacher et al., 2021; Harel et al., 2020; Williams, 2022). Os algoritmos reforçam frequentemente a exposição a informação polarizada influenciando, deste modo, a dinâmica dos debates e o modo como as pessoas interagem em ambientes digitais (Williams, 2022).

Dentro destas comunidades, constituídas por pessoas com perspetivas semelhantes, a polarização, a **desinformação**[8], a ***informação incorreta***[9] e a ***informação maliciosa***[10] propagam-se mais facilmente, uma vez que é menos provável que sejam questionadas ou contestadas (Harel et al., 2020; Williams et al., 2020; Zhang, 2018).

Estes mecanismos podem fomentar ainda mais o discurso de ódio, ao fornecerem uma justificação aparentemente factual para a hostilidade, conferindo às narrativas preconceituosas, que retratam determinados grupos sociais como ameaças, uma aparência de legitimidade (Bradshaw, 2024; Carvalho et al., 2025; Madriaza et al., 2025). De facto, embora as plataformas digitais desempenhem um papel significativo na configuração e aceleração destes processos, muitas vezes estes assentam em padrões de preconceito e divisão social estabelecidos (Udupa, 2025).

Em particular, as **teorias da conspiração** desempenham um papel importante nestas dinâmicas, reforçando o preconceito e a hostilidade entre grupos (Bilewicz, 2025). Ao apresentarem alegações infundadas numa estrutura narrativa que se assemelha a uma explicação factual, as teorias da conspiração atenuam a fronteira entre opinião e "evidência", aumentando, assim, a perceção de credibilidade de narrativas hostis (Burnham et al., 2022; Carvalho et al., 2025; Ghenai et al., 2025). Em alguns casos, o discurso de ódio recorre diretamente a narrativas conspirativas, contribuindo para a disseminação de visões de mundo excludentes e prejudiciais (Bilewicz, 2025). Um exemplo

[8] Divulgação intencional de informações incorretas, com o objetivo de enganar e partilhadas para causar danos graves (Wardle, 2024).
[9] "Misinformation": Divulgação não intencional de informações incorretas partilhadas de boa-fé por pessoas que não têm consciência de que estão a divulgar falsidades (Wardle, 2024).
[10] "Malinformation": Informação baseada na realidade, mas utilizada fora do contexto com o intuito de causar danos (Greene, 2025).

proeminente é a teoria da conspiração da "Grande Substituição" (i.e., *Great Replacement*), que alega falsamente que as populações brancas estão a ser deliberadamente substituídas por imigrantes não brancos, no contexto mais amplo de um esquema orquestrado pelas elites. Este tipo de discurso é frequentemente usado por grupos extremistas para aumentar a polarização social, explorando a retórica de invasão e a apresentação do "outro" como uma ameaça simbólica ou realista. Esta polarização *online* pode ser transferida para o contexto *offline*, podendo traduzir-se em resultados concretos, tais como apoio financeiro, participação em manifestações ou mesmo crimes de ódio (Williams, 2022).

A investigação mostra que estas teorias são particularmente suscetíveis de surgir em contextos sociais polarizados, caracterizados pela presença de exogrupos hostis ou ideologicamente diferentes, e que a mera existência de um exogrupo que inspira desconfiança é suficiente para gerar acusações conspiratórias infundadas contra o mesmo (van Prooijen & Douglas, 2017; van Prooijen, 2021).

A exposição repetida a discursos de ódio, teorias da conspiração e outras formas de desinformação pode contribuir para a **normalização do ódio** e para **processos de dessensibilização** (e.g., Madriaza et al., 2025; Soral et al., 2018; ver secções 5.2 e 5.3).

Em geral, o discurso de ódio pode simultaneamente espelhar a polarização ou reforçá-la, fomentando a hostilidade entre grupos e contribuindo para climas de medo e divisão social (Ghenai et al., 2025; Madriaza et al., 2025). Este fenómeno opera não apenas ao nível da expressão individual, mas também no contexto de outros processos sociais que influenciam perceções, atitudes e comportamentos ao longo do tempo. Estes processos são frequentemente amplificados em contextos sociais e institucionais mais amplos, contribuindo para a normalização do ódio (ver Secção 4.8).

Leituras recomendadas:

Bilewicz, M., & Soral, W. (2020). Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology, 41*(S1), 3–33. https://doi.org/10.1111/pops.12670

Carvalho, P., Caled, D., Silva, M. J. (2025). The Thin Line Between Conspiracy Theories and Opinion: Why Humans and AI Struggle to Differentiate Them. *International Journal of Communication, 19*, 565–591. https://ijoc.org/index.php/ijoc/article/view/22182

Smith, L. G. E., Thomas, E. F., Bliuc, A.-M., & McGarty, C. (2024). Polarization is the psychological foundation of collective engagement. *Communications Psychology, 2*(1). https://doi.org/10.1038/s44271-024-00089-2

Wardle, C. (2024). *A conceptual analysis of the overlaps and differences between hate speech, misinformation and disinformation*. United Nations, Department of Peace Operations and Office of the Special Adviser on the Prevention of Genocide. https://peacekeeping.un.org/sites/default/files/report_-_a_conceptual_analysis_of_the_overlaps_and_differences_between_hate_speech_misinformation_and_disinformation_june_2024.pdf

5. Impactos do Discurso de Ódio

O discurso de ódio *online* não é necessariamente menos prejudicial do que os atos físicos. A literatura demonstra que tem impacto direto nas **pessoas** visadas, nas **comunidades** e na **sociedade em geral** (Hassan et al., 2022; Madriaza et al., 2025; Silva & Carvalho, 2023). Devido ao anonimato que a internet proporciona, é provável que os perpetradores se sintam menos inibidos a partilhar, sem qualquer tipo de filtro, discurso de ódio (Williams, 2022).

5.1. Impactos nas Pessoas

Para as pessoas diretamente visadas, o discurso de ódio pode provocar sentimentos de medo, desconfiança, ansiedade, sintomas depressivos, *stress*, humilhação e isolamento (e.g., Dreißigacker et al., 2024; Hassan et al., 2022; Kentmen-Cin, 2025; Madriaza et al., 2025; Wachs et al., 2022). As experiências de vitimização também têm sido associadas a comportamentos autolesivos, ideação e comportamentos suicidas (Madriaza et al., 2025). Adicionalmente, a exposição repetida ao ódio reduz a qualidade de vida e está associada a um aumento da fobia social (Madriaza et al., 2025).

Os ambientes *online* podem intensificar estes efeitos (e.g., de Roos & Caon, 2026; Madriaza et al., 2025). A rapidez, visibilidade e permanência da comunicação digital pressupõem que as mensagens prejudiciais chegam rapidamente a um vasto público e permanecem acessíveis ao longo do tempo (e.g., Ghenai et al., 2025). As vítimas podem sofrer assédio em várias plataformas, criando uma sensação de exposição constante e insegurança que se transfere dos espaços *online* para os espaços *offline* (Dreißigacker et al., 2024). As pessoas que emitem mensagens de ódio também são afetadas, apresentando níveis significativamente mais elevados de raiva, ansiedade e emoções negativas, o que sugere que os ambientes de ódio *online* afetam os estados emocionais não apenas das pessoas que são alvo de ódio, mas de todos os que estão envolvidos, direta ou indiretamente, nestes ambientes (Ghenai et al., 2025).

Cofinanciado pela União Europeia

5.2. Impactos nas Comunidades-Alvo e nas Testemunhas

O discurso de ódio não afeta apenas aqueles que são diretamente visados. Tem também impacto nas comunidades como um todo, uma vez que a sua exposição a este fenómeno nas redes sociais reforça preconceitos, estereótipos e visões polarizadas, aprofundando divisões sociais (e.g., Kentmen-Cin, 2025) e contribuindo para comportamentos violentos na vida real (de Roos & Caon, 2026). Mesmo que não sejam diretamente visados nas mensagens de ódio, os membros dos grupos-alvo são igualmente afetados por esta exposição ao ódio (i.e., vitimização vicariante/indireta; Madriaza et al., 2025; Pickles, 2021).

Ao longo do tempo, a exposição frequente a conteúdos de ódio pode dessensibilizar as testemunhas (i.e., *bystanders*: pessoas que testemunham o discurso de ódio sem serem diretamente visadas), reduzindo a empatia, normalizando a discriminação e diminuindo a intervenção nessas situações (Kentmen-Cin, 2025; Madriaza et al., 2025; Soral et al., 2018). Assim, à medida que a linguagem prejudicial se torna mais generalizada, pode ser percebida como mais aceitável ou menos extrema, afetando as normas sociais, levando os indivíduos a acreditar que o discurso de ódio é comum e socialmente tolerado simplesmente porque o encontram regularmente (Bilewicz & Soral, 2020). Por exemplo, a investigação mostra que a exposição frequente à linguagem depreciativa atenua a sensibilidade à dor dos outros, sugerindo que o discurso de ódio compromete a empatia (Pluta et al., 2023). Com o tempo, a dessensibilização e a normalização podem reduzir a capacidade das pessoas de reconhecer discurso de ódio e as suas consequências, enfraquecendo desta forma as normas sociais que o desencorajariam (Bilewicz & Soral, 2020).

5.3. Impactos na Sociedade e na Democracia

A nível societal, o discurso de ódio pode colocar em causa os valores democráticos e comprometer a coesão social (Hassan et al., 2022). O debate público depende do respeito, do pluralismo e da participação equitativa. Quando certos grupos são sistematicamente retratados como inferiores, perigosos ou ilegítimos, as suas vozes podem ser marginalizadas ou silenciadas

(Kentmen-Cin, 2025). Com base na teoria da espiral do silêncio, a investigação sugere que, à medida que o discurso de ódio se intensifica, assumindo formas mais extremas de abuso, pode silenciar e isolar as pessoas e as comunidades visadas, desencorajando-os de expressar as suas opiniões e de participar no discurso público, aprofundando assim a marginalização política e a polarização (Kentmen-Cin, 2025).

A exposição ao discurso de ódio também está relacionada com mudanças nas convicções políticas e pode influenciar a disposição dos indivíduos para apoiar partidos políticos radicais ou extremistas e incentivar a adesão a ideologias extremistas (Madriaza et al., 2025). A existência prolongada de hostilidade *online* pode contribuir para padrões mais amplos de tensão social e, em alguns contextos, pode preceder aumentos em comportamentos de agressão no mundo real ou em crimes de ódio (e.g., Calderón et al., 2024; Hassan et al., 2022; Madriaza et al., 2025). Embora o discurso *online* possa não causar automaticamente violência, pode influenciar atitudes e normas sociais que tornam a violência mais aceitável.

À medida que o panorama digital continua a evoluir, compreender e dar resposta às consequências psicológicas e sociais do discurso de ódio é uma prioridade urgente (Dreißigacker et al., 2024).

Leituras recomendadas:

Hassan, G., Rabah, J., Madriaza, P., Brouillette-Alarie, S., Borokhovski, E., Pickup, D., Varela, W., Girard, M., Durocher-Corfa, L., & Danis, E. (2022). PROTOCOL: Hate online and in traditional media: A systematic review of the evidence for associations or impacts on individuals, audiences, and communities. *Campbell Systematic Reviews, 18*(2). Portico. https://doi.org/10.1002/cl2.1245

Madriaza, P., Hassan, G., Brouillette-Alarie, S., Mounchingam, A. N., Durocher-Corfa, L., Borokhovski, E., Pickup, D., & Paillé, S. (2025). Exposure to hate in online and traditional media: A systematic review and meta-analysis of the impact of this exposure on individuals and communities. *Campbell Systematic Reviews, 21*, e70018. https://doi.org/10.1002/cl2.70018

PARTE II

Grupos-Alvo e Padrões de Ódio

6. Grupos-Alvo e Formas Comuns de Ódio em Portugal

De acordo com relatórios recentes da UE e estudos realizados em Portugal, os grupos sociais mais frequentemente visados pelo discurso de ódio *online* no país são as **comunidades racializadas** (incluindo comunidades afrodescendentes e ciganas), **migrantes**, as **comunidades LGBTI+** e as **mulheres** (Carvalho & Guerra, 2023; Carvalho et al., 2024; EUAFR, 2025; Comissão Europeia contra o Racismo e a Intolerância, 2025; Instituto Europeu para a Igualdade de Género, 2025; Guerra et al., 2025). Compreender as nuances no conteúdo e as motivações subjacentes ao discurso de ódio dirigido a diferentes grupos sociais é crucial para desenvolver respostas eficazes de prevenção e combate a este fenómeno. Embora possam existir algumas semelhanças nos padrões do discurso de ódio, que são transversais a todos os grupos sociais, é importante ter em conta a existência de processos específicos de discriminação em função do grupo social envolvido.

6.1. Processos de Racialização

O **processo de racialização** consiste na construção social de identidades racializadas, que orientam a perceção, categorização e hierarquização de determinados grupos sociais. O racismo, enquanto ideologia, baseia-se na crença de superioridade do endogrupo, o que justifica comportamentos discriminatórios e depreciativos em relação ao exogrupo racializado, contribuindo para a normalização dos privilégios a ele associados (Jones, 1997).

Neste enquadramento, destacam-se as comunidades afrodescendentes e ciganas em Portugal (Almeida & Pereira, 2026; Almeida et al., 2023; Carvalho et al., 2023; Guerra et al., 2025). O legado colonial e a ideologia partilhada do Lusotropicalismo (ver secção 4.3) afetam as atitudes e comportamentos intergrupais, incluindo formas subtis de derrogação racial. Por exemplo, a nostalgia e o excecionalismo coloniais são narrativas prevalentes no discurso de ódio dirigido a grupos racializados, com a negação do racismo a constituir uma das estratégias relevantes mais mobilizadas (Almeida et al., 2023; Guerra

Cofinanciado pela União Europeia

et al., 2025). No mesmo sentido, a comunidade cigana enfrenta séculos de discriminação e exclusão social, reproduzidos nos discursos sociais e meios de comunicação tradicionais, que perpetuam estereótipos de criminalidade e marginalização social. De facto, a investigação mostra que o discurso de ódio *online* contra as pessoas ciganas em Portugal é normalizado e desumanizante, construindo, sistematicamente, estas comunidades como uma ameaça para a sociedade (Almeida & Pereira, 2026; Guerra et al., 2025).

A investigação revela ainda que as pessoas tendem a agrupar diferentes formas de preconceito no mesmo discurso, como ilustrado no exemplo abaixo:

> *"Os portugueses são famosos por serem um povo sereno, e agora os africanos, a comunidade LGBT e os ciganos são os melhores; em breve, os europeus estarão de joelhos, a pedir desculpa por terem sido derrotados por pessoas do Terceiro Mundo e pelas minorias em geral"*

Neste caso, os termos associados ao grupo dominante (no exemplo, indivíduos portugueses brancos, cisgénero e heterossexuais) coocorrem com termos que se referem a vários grupos marginalizados, incluindo as comunidades afrodescendentes, ciganas e LGBTI+. Neste comentário, a pessoa que emite a mensagem inverte estrategicamente as posições de domínio e poder entre grupos dominantes (portugueses/europeus) e grupos marginalizados (pessoas africanas, LGBTI+, ciganas, populações do "Terceiro Mundo" e minorias em geral). Ao retratar os grupos dominantes como potenciais vítimas, o discurso mascara desigualdades estruturais e reinterpreta as relações de poder existentes como injustamente invertidas.

6.2. Xenofobia

A **xenofobia** envolve preconceito, medo ou hostilidade em relação a pessoas consideradas estrangeiras ou pessoas percebidas como não pertencentes ao grupo dominante (e.g., portugueses), **visando frequentemente pessoas migrantes e de origem imigrante** (Sanchez-Mazas & Licata, 2015). Estas dinâmicas têm frequentemente origem em crenças etnocêntricas que

posicionam a própria cultura, valores ou modo de vida como superiores aos dos outros, reforçando o favoritismo endogrupal[11] e a derrogação do exogrupo[12]. Por exemplo, as pessoas migrantes são vistas como ameaças à identidade nacional ou à estabilidade económica. Tanto as ameaças percebidas como realistas (e.g., recursos económicos) como as simbólicas (e.g., culturais ou baseadas em valores) alimentam o preconceito e a discriminação, conforme ilustrado no exemplo abaixo (ver secção 4.2). Estas dinâmicas desempenham um papel crucial nos padrões contemporâneos de discurso de ódio dirigido às comunidades migrantes (Guerra et al., 2025).

> *"Há alguns anos, antes da onda de imigração, a televisão raramente noticiava casos de violência contra civis ou agentes da autoridade – talvez uma notícia por mês. Agora, se ligarmos a CMTV, há dezenas de notícias desse tipo todos os dias. Não pode ser mera coincidência"*

Neste exemplo, o discurso xenófobo é construído através de uma relação causal implícita entre a imigração e o aumento da violência. O comentário ilustrado apresenta as pessoas migrantes como sendo as responsáveis pela insegurança e desordem pública, recorrendo a evidências seletivas para sustentar uma generalização mais ampla. Isto reflete a mobilização de ameaças realistas (i.e., violência), mantendo simultaneamente uma negação plausível.

É de salientar que o discurso de ódio também pode visar pessoas com base na sua **identidade religiosa** percebida, particularmente quando a religião se cruza com aspetos relacionados com a **migração ou a etnia**. Por exemplo, em Portugal a retórica antimuçulmana e islamofóbica interliga-se frequentemente com narrativas xenófobas e racializadas (Comissão Europeia contra o Racismo e a Intolerância, 2025). De facto, a investigação recente mostra que a islamofobia é altamente prevalente *online*, sendo as pessoas refugiadas e

[11] Tratamento desigual através da demonstração de favoritismo para com o grupo próprio.
[12] Discriminação contra o exogrupo.

imigrantes percebidas como muçulmanas, independentemente das suas diversas etnias e religiões (Almeida et al., 2023).

6.3. Preconceito Sexual

O **preconceito sexual** envolve atitudes negativas em relação às pessoas com base na sua orientação sexual, identidade de género, expressão de género ou características sexuais percebidas. Este preconceito está intimamente relacionado com o estigma sexual - a crença de que as identidades e comportamentos não heterossexuais são intrinsecamente errados ou inferiores - e com ameaças percebidas (Herek, 2004). As suas manifestações incluem a homofobia, a bifobia, a transfobia e a intersexofobia, que predizem discriminação e agressão, tais como assédio verbal, violência física e sexual e exclusão social (Katz-Wise & Hyde, 2012). O exemplo abaixo ilustra como as perceções de ameaça podem ser expressas de forma exagerada e generalizada, refletindo medos subjacentes sobre a mudança social.

> *"Estas pessoas LGBTI+ estão a dominar tudo"*

Em particular, as **pessoas trans** em Portugal enfrentam desafios específicos relacionados tanto com o preconceito social como com barreiras institucionais (e.g., Neves et al., 2023). Embora o reconhecimento legal do género seja possível por autodeterminação a partir dos 16 anos[13], as pessoas trans continuam a ser particularmente vulneráveis ao discurso de ódio, ao assédio e à exclusão social (e.g., Neves et al., 2023). Quando as pessoas acreditam que certos grupos desafiam os seus valores ou modo de vida, isso pode levar a sentimentos de repulsa ou raiva e resultar em exclusão ou agressão para defender essas crenças (Guerra et al., 2025). Abordar as realidades específicas das comunidades trans é, pois, essencial para compreender de forma mais abrangente e informada as manifestações do discurso de ódio anti-LGBTI+.

[13] Entre os 16 e os 18 anos é possível, mas apenas com a autorização dos seus representantes legais. No entanto, está atualmente em curso no Parlamento português um processo legislativo para revogar a Lei nº 38/2018. Não obstante, a lei ainda não foi oficialmente revogada e os seus procedimentos legais continuam em vigor.

> *"Maricas e travestis... Vá lá, vamos preparar os fornos [referência ao crematório nos campos de concentração nazis]"*

A violência e a discriminação baseadas no **género** continuam a ser desafios persistentes em Portugal, afetando particularmente as mulheres e as raparigas através de formas de discurso de ódio *online*, assédio e abuso sexual. Apesar dos progressos em matéria de direitos legais e representação política, as mulheres continuam a sofrer elevadas taxas de violência baseada no género, incluindo violência doméstica e feminicídio (e.g., Instituto Europeu para a Igualdade de Género, 2025). Um relatório de 2026 (Silva & Brázia, Ciberviolência de Género em Portugal: Perspetivas de Crianças e Jovens, Escolas e Autoridades de Aplicação da Lei) mostra que a violência *online* contra as mulheres está profundamente ligada a estruturas mais amplas de sexismo e desigualdade de género. Além disso, mostra que o discurso de ódio *online* dirigido às mulheres é também sustentado por estereótipos sexistas persistentes, hipersexualização e atitudes de culpabilização da vítima que normalizam a agressão contra as mulheres nos espaços digitais. Relatórios internacionais (e.g., FRA, 2023) apontam para os mesmos resultados, concluindo que os espaços *online* refletem e amplificam cada vez mais narrativas **misóginas** (i.e., hostilidade ou preconceito especificamente dirigidos às mulheres), perpetuando estereótipos e atitudes hostis para com as mulheres. O exemplo abaixo ilustra o incitamento à violência e ao ódio dirigido a uma mulher.

> *"Vitimizaçao. Algo de muito comum numa mulher. Para alem de Gorda é fraca"*
>
> *"Eu nem era totalmente anti-aborto, mas depois desta merda toda, acabei por perceber que vocês são só umas putas que querem levar na cona sem consequência. Aborto não é contraceptivo."*

6.4. Intersecionalidade

Por fim, as pessoas podem ser alvo de discriminação com base na perceção da sua pertença a vários grupos sociais marginalizados. A **intersecionalidade** destaca a forma como as pessoas se posicionam no seio de sistemas de opressão e supremacia com raízes históricas, associados ao racismo, à cultura, ao género, à classe social e a outras categorias sociais. Estas múltiplas identidades intersetam-se, produzindo experiências de pertença e exclusão que não podem ser compreendidas através da análise isolada de cada identidade. Os marcadores de identidade social raramente funcionam de forma independente, dando frequentemente origem a alvos de discurso de ódio complexos e sobrepostos, em vez de grupos isolados e separados (Carvalho & Guerra, 2023). O exemplo abaixo ilustra uma publicação com um alvo intersecional (i.e., mulheres e a religião).

> *"És uma retardada psicótica, nem acredito que te atreves a aparecer, lol, não tens um futuro promissor pela frente; haverá consequências graves para os traidores do país e para as p**** que abraçam o Islão. É só esperar."*

Leituras recomendadas:

Almeida, P. (2022). *Relatório do projeto de Investigação "Racismo e Xenofobia em Portugal: a normalização dos discursos de ódio no espaço público da internet"*. https://racismoexenofobia.cria.org.pt/wp-content/uploads/2022/07/Relato%CC%81rio_Racismo_Xenofobia.pdf

European Commission Against Racism and Intolerance. (2025). *Sixth report on Portugal* [Report]. Council of Europe.

Guerra, R., Carvalho, P., Marques, C., Carmona, M., Sarroeira, R., Batista, F., Ribeiro, R., Fonseca, A., Moro, S., & Silva, C. (2025). Unpacking online hate speech in Portuguese social media: a social-psychological and linguistic-discursive approach. *Humanities and Social Sciences Communications, 12*(1). https://doi.org/10.1057/s41599-025-05392-9

PARTE III

Prevenção e Enquadramento Legal

7. Práticas de Prevenção e Sensibilização

O combate ao discurso de ódio *online* deve ser um processo contínuo, em vez de intervenções isoladas, que deve ir além da mera identificação de conteúdos nocivos depois de estes terem sido amplamente circulados. Dada a complexidade das manifestações do discurso de ódio, e dos ecossistemas em que este opera, os esforços de prevenção devem obrigatoriamente abranger diferentes níveis e esferas de ação, e estar fortemente articulados com as políticas públicas. Estes esforços devem envolver medidas abrangentes, incluindo respostas legais e regulamentares, políticas educativas, tanto no domínio da literacia digital como da educação para os direitos humanos, bem como a sensibilização da sociedade em geral e o empoderamento das comunidades afetadas.

7.1. Sensibilização: Campanhas Públicas

As **campanhas públicas** desempenham um papel importante no combate ao discurso de ódio *online*, ao alertar para as consequências prejudiciais deste fenómeno, ao incentivar a denúncia e ao reforçar a mensagem de que o comportamento digital acarreta implicações sociais e legais. Em Portugal, por exemplo, a recente campanha "Ódio online, mata offline", promovida pela Polícia Judiciária, destacou o perigo e as consequências do extremismo e dos conteúdos de ódio *online*. Outros exemplos de campanhas de sensibilização e prevenção em Portugal incluem a campanha #CortarOMalPelaRaiz do projeto kNOwHATE, o Movimento #respectbattles: Combater o Ódio com Respeito da APAV (Associação Portuguesa de Apoio à Vítima) e "O discurso de ódio não é um argumento" da Rede Europeia de Combate à Pobreza em Portugal.

Embora as campanhas de sensibilização, isoladamente, não possam eliminar o discurso de ódio, contribuem para influenciar de forma positiva as normas sociais e aumentar a sensibilidade coletiva em relação a este fenómeno e às suas consequências negativas.

7.2. Iniciativas de Educação e Formação

Outro conjunto importante de estratégias para combater e limitar a propagação do discurso de ódio *online* centra-se na **educação e na formação** (Isasi & Juanatey, 2016; Madriaza et al., 2025; Mossou & Lane, 2018). Estas estratégias são fundamentais para capacitar as pessoas, especialmente os jovens, com as competências necessárias para reconhecer e combater o discurso de ódio *online*, permitindo, simultaneamente, o equilíbrio entre responsabilidade, liberdade de expressão e a participação ativa nos espaços digitais (Isasi & Juanatey, 2016). Embora todos os setores da sociedade estejam expostos a este tipo de conteúdos prejudiciais, os jovens estão particularmente mais vulneráveis devido à utilização intensiva que fazem das redes sociais, tornando-os um foco central de muitas iniciativas de educação e formação. Estas iniciativas podem ser agrupadas, em termos gerais, em duas categorias: as que se destinam ao **público em geral**, que procuram promover a "cidadania digital"[14] através do desenvolvimento de conhecimentos e competências relevantes; e as direcionadas para pessoas já envolvidas na prevenção e combate ao discurso de ódio, incluindo **profissionais e ativistas**, que necessitam de ferramentas mais específicas para responder de forma eficaz a este fenómeno (Isasi & Juanatey, 2016).

No **contexto português**, um exemplo destas iniciativas de sensibilização, foi o desenvolvimento do Manual Referências. Este manual, concebido para apoiar o Movimento Contra o Discurso de Ódio e a Campanha da Juventude do Conselho da Europa contra o Discurso de Ódio Online, é útil para os **educadores** que trabalham nesta área, dentro e fora do sistema de ensino formal. O manual destina-se a jovens com idades compreendidas entre os 13 e os 18 anos, embora as atividades possam ser adaptadas a outros grupos etários (Conselho da Europa, 2016).

De um modo geral, as iniciativas de sensibilização que promovem a literacia digital, a "cidadania digital", o pensamento crítico e a literacia mediática podem ajudar as pessoas a:

[14] Adaptar o conceito de educação para a cidadania, de modo a abranger os conhecimentos e as competências necessárias para uma interação responsável em contextos digitais (Isasi & Juanatey, 2016).

- Reconhecer a desinformação e narrativas manipuladoras;
- Identificar formas codificadas ou indiretas de discurso de ódio;
- Reconhecer racismo, sexismo, xenofobia, entre outros, presentes no discurso de ódio *online*;
- Refletir sobre os impactos psicológicos e sociais do discurso de ódio;
- Refletir antes de partilhar conteúdos próprios ou de terceiros.

As iniciativas de formação destinadas a educadores, jornalistas, funcionários públicos e líderes comunitários são importantes, uma vez que estes atores desempenham frequentemente um papel mediador no discurso público, reforçando ou contestando narrativas prejudiciais (e.g., UNESCO e Gabinete das Nações Unidas para a Prevenção do Genocídio e a Responsabilidade de Proteger, 2023).

Em última instância, as iniciativas de educação e formação preventivas não pretendem suprimir o debate ou a liberdade de expressão, mas sim promover uma comunicação responsável, o envolvimento democrático e a sensibilização para os direitos fundamentais (Kentmen-Cin, 2025; Zendeli, 2017).

7.3. Contra-Discurso e Narrativas Positivas

A investigação recente tem-se centrado, cada vez mais, na forma como surgem novas normas sociais na comunicação *online*, incluindo a utilização crescente do contra-discurso para combater o discurso de ódio (Bilewicz & Soral, 2020). O **contra-discurso** consiste em respostas que desafiam e enfraquecem o discurso de ódio, introduzindo mensagens alternativas que promovem a inclusão, o esclarecimento factual ou a solidariedade (e.g., Bilewicz & Soral, 2020; Gagliardone et al., 2015; Williams, 2022). Deste modo, o contra-discurso caracteriza-se, essencialmente, pela introdução de perspetivas diferentes na conversa, e não pelo silenciamento de vozes prejudiciais (e.g., remoção de conteúdos, Baider, 2023).

A investigação mostra que o contra-discurso pode ajudar a reduzir a propagação do ódio *online*, por exemplo, através da utilização de *hashtags*

inclusivos (Williams, 2022). Esta estratégia pode surgir espontaneamente ou ser mais organizada, através de campanhas coordenadas de *hashtags* (e.g., #ichbinhier, #brasileirasnaosecalam), podendo assumir diferentes formas, incluindo texto e imagens (Benesch et al., 2016). Estudos mostram que certas formas de contra-discurso são mais eficazes, como por exemplo, alertar para as consequências negativas do discurso prejudicial ou recorrer a emoções positivas, como a empatia ou o humor (Benesch et al., 2016; Chung et al., 2024). Estas estratégias podem levar a uma mudança de tom por parte da pessoa que emite a mensagem de ódio ou influenciar o público mais alargado (i.e., testemunhas/*bystanders*), tornando menos provável o seu apoio ao discurso de ódio.

Embora a investigação existente saliente o potencial do contra-discurso (e.g., Baider, 2022; Gennaro et al., 2025), a sua eficácia no combate ao discurso de ódio não é consensual, e vários estudos revelam que nem todas as formas de contra-discurso são igualmente eficazes (e.g., Baider, 2023; Wang et al., 2026; Williams, 2022). Por exemplo, o contra-discurso pode aumentar, involuntariamente, a visibilidade, o envolvimento e a intensidade emocional de conteúdos de ódio, especialmente tendo em conta que os algoritmos das redes sociais são desenhados para maximizar o envolvimento dos utilizadores. A utilização de insultos ou de um tom hostil nas mensagens de contra-discurso também não é apontada como uma estratégia eficaz, agravando a interação e originando mais conteúdos prejudiciais (Baider, 2023; Williams, 2022). Assim, a investigação sugere que o contra-discurso é mais eficaz quando segue alguns princípios-chave, incluindo: evitar a utilização de linguagem insultuosa ou de ódio, apresentar argumentos claros e consistentes e solicitar provas quando as alegações são falsas ou questionáveis. Incentivar outras pessoas a participarem no contra-discurso e, nos casos em que as respostas pareçam ser falsas ou automatizadas (*bot*), denunciá-las à respetiva plataforma também têm sido apontadas como estratégias úteis (Chung et al., 2024; Williams, 2022).

Resultados anteriores do projeto kNOwHATE, baseados em dois *corpora* anotados compostos por comentários e respostas associadas recolhidos no *YouTube* e no *Twitter/X* em Portugal entre 2021 e 2022, revelaram que as estratégias de contra-discurso mais prevalentes assentam na utilização de

contra-estereótipos, empatia, identidades inclusivas e apelo a normas legais e sociais.

↘ Contra-estereótipos

Os contra-estereótipos referem-se a crenças ou representações que desafiam crenças culturais amplamente difundidas sobre um grupo (Čehajić-Clancy & Bilewicz, 2021). Os contra-estereótipos põem em causa preconceitos e promovem perceções mais precisas e complexas dos grupos sociais, ao tornarem saliente informação que contraria os estereótipos comuns (Čehajić-Clancy & Bilewicz, 2021). No exemplo abaixo, a pessoa que emite a mensagem contesta explicitamente o estereótipo de que a comunidade cigana beneficia, de forma desproporcional, de subsídios estatais:

> *"Já tentaste descobrir a verdade sobre os subsídios atribuídos à comunidade cigana? Acho que não, mas há vários sites que podem comprovar a tua enorme estupidez disfarçada de preconceito étnico. Se precisares, posso enviar-te alguns"*

↘ Empatia

A empatia desempenha também um papel importante no combate ao discurso de ódio (Bilewicz & Soral, 2020), sendo uma das estratégias de contra-discurso mais testadas, incluindo através da utilização de *bots* de contra-discurso (e.g., Gennaro et al., 2025; Wang et al., 2026). A empatia envolve tanto respostas emocionais, como sentir preocupação pelas outras pessoas, como elementos mais cognitivos, como ver as situações da perspetiva do outro, e está associada a uma maior tolerância e comportamentos de ajuda (Batson, 2009). O exemplo abaixo ilustra uma estratégia de contra-discurso empática:

> *"Tenta vestir-te como uma mulher e vai procurar emprego e arranja habitação para veres se é fácil! Desculpa, mas não devias ser tão retrógrado quando se trata de apoiar as pessoas [...]"*

Cofinanciado pela União Europeia

Neste comentário, a pessoa que emite a mensagem convida à adoção de uma nova perspetiva, incentivando o interlocutor a imaginar as experiências vividas e as dificuldades estruturais enfrentadas pelas mulheres. Este apelo à empatia é utilizado para contestar pontos de vista preconceituosos ou menosprezantes, colocando em evidência as desigualdades e barreiras sociais existentes.

↘ Identidades inclusivas

Esta estratégia é altamente eficaz para reduzir o preconceito e os conflitos intergrupais (e.g., Gaertner & Dovidio, 2000; Gaertner et al., 2016). Envolve a saliência de identidades partilhadas, incentivando as pessoas a verem-se como parte de um "nós" mais amplo e inclusivo, em detrimento de distinções de "nós vs. eles". Estas identidades inclusivas podem ser alcançadas através da promoção da cooperação, da interdependência ou da saliência de pontos em comum, redirecionando processos psicológicos de favorecimento endogrupal e promovendo atitudes e comportamentos mais positivos (e.g., maior empatia) (Gaertner & Dovidio, 2000; Gaertner et al., 2016). Por exemplo, a afirmação apresentada abaixo ilustra um apelo a uma identidade partilhada. Este tipo de discurso reenquadra as diferenças (e.g., cor da pele, nacionalidade) como secundárias em relação a uma identidade comum mais inclusiva, promovendo assim a unidade entre diversos grupos sociais.

> *"Muito bem. Os portugueses têm de se unir. Sejam brancos, negros, ciganos ou o que for, temos uma coisa em comum: somos portugueses".*

↘ Normas sociais

As normas sociais são expectativas partilhadas sobre atitudes e comportamentos considerados adequados dentro de um grupo ou sociedade (McDonald & Crandall, 2015). Podem ser definidas como *"uma espécie de gramática das interações sociais. Tal como uma gramática, um sistema de normas especifica o que é aceitável e o que não é aceitável numa sociedade ou num grupo"* (Bicchieri et al., 2023). Por isso, as normas sociais são padrões

poderosos que orientam as perceções, os julgamentos e os comportamentos dos indivíduos.

Normas sociais que enquadram o discurso de ódio como socialmente inaceitável conseguem dissuadir as pessoas de promoverem e aceitarem este tipo de discurso, limitando a sua propagação. Deste modo, as normas sociais podem incentivar respostas coletivas, especialmente entre aqueles que reconhecem o discurso de ódio como uma forma de injustiça e estão motivados para o combater (Bilewicz & Soral, 2020). O recurso ao apelo a normas legais no contexto de contra-discurso, pode também dissuadir a propagação e aceitação do discurso de ódio *online*, ao sinalizar que este tipo de discurso tem, em alguns casos, consequências legais. Estas regulamentações legais reforçam, simultaneamente, as normas sociais que enquadram o discurso de ódio como inaceitável, desencorajando a sua utilização e apoiando esforços coletivos para o contestar (Bilewicz & Soral, 2020).

Os exemplos apresentados ilustram como as normas sociais podem ser mobilizadas para promover a inclusão e regular o comportamento.

> *"Quem nasce em Portugal é português. Não importa se és etnicamente europeu, africano, americano, asiático, etc..."*

> *"Ninguém deve ser perseguido na rua desta maneira nojenta e ameaçadora, deveriam ambos ser multados ou presos pelo que fizeram. Isto não é a maneira de se "fazerem ouvir", isto é extremismo, querem falar? Falem por meios democráticos."*

O primeiro comentário reforça uma norma inclusiva de pertença nacional. Ao sublinhar que a identidade portuguesa não depende da origem étnica, este discurso contribui para estabelecer uma expectativa comum de que a exclusão com base na origem é inadequada ou mesmo ilegítima. Da mesma forma, o segundo comentário ilustra a aplicação de normas sociais e legais contra o assédio e a agressão. Ao enquadrar esse comportamento como inaceitável e

apelar a consequências, a pessoa que emite a mensagem reforça a ideia de que as ações relacionadas com o ódio violam tanto os padrões sociais como os limites legais.

Estes exemplos ilustram a forma como as normas sociais podem promover pertenças sociais mais inclusivas, desencorajando comportamentos prejudiciais e legitimando respostas coletivas ao discurso de ódio.

Concluindo, o contra-discurso é também uma estratégia que destaca o papel importante das testemunhas (i.e., *bystanders)* no combate ao discurso de ódio (Wang et al., 2026; Gennaro et al., 2025). A investigação mostra que menos de metade das pessoas que testemunham discurso de ódio intervém, seja para apoiar a vítima ou para confrontar o agressor (Williams, 2022). Existem várias formas de resposta ao discurso de ódio em contextos *online* que podem evitar o seu agravamento, incluindo denunciar conteúdos prejudiciais nas respetivas plataformas ou envolver-se ativamente na produção de contra-discurso que desafie a mensagem. Em última instância, as respostas coletivas são particularmente importantes, uma vez que o contra-discurso tende a ser mais eficaz quando vários utilizadores reforçam conjuntamente as normas do que é um comportamento aceitável (Williams, 2022).

7.4. Medidas ao Nível das Plataformas

As **plataformas digitais** desempenham um papel central na prevenção e gestão do discurso de ódio *online* (Weber et al., 2023). Os mecanismos de denúncia, os sistemas de moderação de conteúdos e as políticas de transparência podem influenciar a rapidez com que os conteúdos nocivos são respondidos (Isasi & Juanatey, 2016). As secções de comentários podem também servir como espaços de promoção de contra-discursos, e os algoritmos de recomendação podem ser desenhados para mitigar a existência de bolhas de filtro e para introduzir pontos de vista mais diversificados (Weber et al., 2023).

No entanto, na prática, os mecanismos de denúncia, moderação e remoção de conteúdos nem sempre funcionam de forma eficaz. Estudos recentes destacam limitações persistentes tanto nas ferramentas de moderação automatizadas

como nos sistemas de revisão humana, em particular na deteção de formas indiretas de discurso de ódio, incluindo conteúdos misóginos e racistas (FRA, 2023). Uma análise de 1 500 publicações nas redes sociais que já tinham sido avaliadas pelos sistemas de moderação das plataformas revelou que mais de metade continuava a ser classificada como discurso de ódio por codificadores humanos, ilustrando a dificuldade de identificar de forma consistente conteúdos prejudiciais (FRA, 2023). Por outro lado, os sistemas de moderação de conteúdos podem também remover incorretamente conteúdos legais, levantando preocupações relativamente à transparência e à liberdade de expressão. Estas inconsistências refletem os desafios que as plataformas digitais enfrentam na moderação de grandes volumes de conteúdo, incluindo a sua dependência de sistemas automatizados e a crescente pressão sobre as equipas de revisão humana. A escala, complexidade e velocidade da comunicação *online* tornam a moderação de conteúdos extremamente complexa (FRA, 2023). Plataformas como o *Twitter/X* processam centenas de milhões de publicações por dia, com a atividade a atingir, por vezes, níveis ainda mais elevados. A gestão deste volume de conteúdo é um enorme desafio, mesmo recorrendo a sistemas automatizados e moderação humana. A gestão destes conteúdos implica ainda a tomada de decisões difíceis, uma vez que as plataformas têm de equilibrar os esforços de moderação com preocupações em torno da liberdade de expressão e de censura (Isasi & Juanatey, 2016).

A investigação sugere ainda que os próprios mecanismos de denúncia podem desencorajar o envolvimento das pessoas que utilizam as redes sociais. As interfaces de denúncia são frequentemente vistas como opacas, impessoais e difíceis de utilizar, com *feedback* limitado, apoio insuficiente às vítimas e pouca transparência no que respeita às decisões de moderação (Cover et al., 2025; European Observatory of Online Hate, 2024). Alguns estudos sugerem que as taxas de denúncia continuam a ser relativamente baixas, em parte devido à desconfiança nas respostas das plataformas e à incerteza quanto à sua tomada de medidas.

Em resposta a estes desafios, algumas iniciativas europeias recentes, como a Lei dos Serviços Digitais (*Digital Services Act*), procuram reforçar a transparência, a responsabilização e a cooperação entre plataformas,

organizações da sociedade civil e autoridades policiais. As organizações terceiras e as denúncias às forças policiais podem ainda garantir o apoio adicional às vítimas, o acesso a serviços de defesa e a canais alternativos para a denúncia de conteúdos nocivos (European Observatory of Online Hate, 2024).

Apesar destas limitações, denunciar conteúdos de ódio nas redes sociais continua a ser um passo importante. Embora os procedimentos variem de plataforma para plataforma, as ferramentas de denúncia estão normalmente acessíveis através de políticas, formulários de reclamação ou secções do tipo "Denunciar um problema". Mesmo que os resultados nem sempre sejam os esperados, o recurso a estes canais formais continua a ser uma forma significativa de combater e limitar a propagação do discurso de ódio (Silva et al., 2024).

Em suma, a prevenção do discurso de ódio é uma responsabilidade coletiva. Instituições, educadores, empresas de TI, plataformas digitais, organizações da sociedade civil, autoridades responsáveis pela aplicação da lei e cidadãos, todos desempenham um papel na definição de um ambiente digital seguro e inclusivo. Importa sublinhar que uma prevenção eficaz não depende apenas de restrições ou punições, mas também da utilização de estratégias de sensibilização e educação, da promoção da responsabilidade ética e de salvaguardas institucionais, para reduzir os impactos negativos do discurso de ódio e reforçar a resiliência democrática.

Leituras recomendadas:

European Union Agency for Fundamental Rights - FRA (2023). *Online content moderation - Current challenges in detecting hate speech (Catalogue No. TK-04-23-883-EN-C).* Publications Office of the European Union. https://fra.europa.eu/en/publication/2023/online-content-moderation

Gagliardone, I., Gal, D., Alves, T. & Martinez, G. (2015). *Countering online hate speech*. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000233231

UNESCO & United Nations Office on Genocide Prevention and the Responsibility to Protect. (2023). *Addressing hate speech through education: A guide for policy-makers*. https://unesdoc.unesco.org/ark:/48223/pf0000384872.

Williams, M. (2022). *The Science of Hate: How Prejudice Becomes Hate and What We Can Do to Stop It*. Faber & Faber, Ltd

8. Enquadramento Legal e Institucional em Portugal

Para além de compreender o que é o discurso de ódio e como este se manifesta e funciona, é fundamental distinguir entre o que constitui uma potencial infração penal ao abrigo da legislação portuguesa e o que não tem este enquadramento.

A análise jurídica do discurso de ódio em Portugal tem necessariamente de começar pela análise do que se enquadra como liberdade de expressão e informação, um direito consagrado no artigo 37.º da Constituição da República Portuguesa e em outros instrumentos jurídicos internacionais, como na Declaração Universal dos Direitos Humanos (artigo 19.º), na Convenção Europeia dos Direitos do Homem (artigo 10.º), no Pacto Internacional Sobre os Direitos Civis e Políticos (artigo 19.º) e na Convenção Americana sobre Direitos Humanos (artigo 13.º, nº 5). Segundo o artigo 37.º da Constituição da República Portuguesa, "*todos têm direito de exprimir e divulgar livremente o seu pensamento pela palavra, pela imagem ou por qualquer outro meio, bem como o direito de informar, de se informar e de ser informados, sem impedimentos nem discriminações*".

Contudo, a liberdade de expressão não é um direito absoluto, entendimento sustentado pela jurisprudência portuguesa, como no Acórdão do Tribunal da Relação do Porto[15] que sublinha que "*o direito à liberdade de expressão, tal como muitos outros, não é um direito absoluto e que possa ser exercido de*

15 Acórdão do Tribunal da Relação do Porto de 7 de junho de 2023 (Proc. n.º 5551/19.0T9LSB-A.P1)

forma irrestrita ou ilimitada, uma vez que existem limites ao exercício do direito de exprimir e divulgar livremente o pensamento e as opiniões". O artigo 18.º da Constituição da República Portuguesa também reforça esta posição, informando ainda que as restrições devem apenas "*limitar-se ao necessário para salvaguardar outros direitos ou interesses constitucionalmente protegidos*", como o direito à identidade pessoal, ao bom nome e reputação, à imagem e a proteção legal contra quaisquer formas de discriminação (artigo 26.º, nº1). Estes direitos são muitas vezes colocados em causa quando falamos em incidentes de discurso de ódio, havendo uma linha ténue entre o que é considerado liberdade de expressão e o que se classifica como discurso de ódio.

A compreensão destes limites constitucionais permite entender de que forma o ordenamento jurídico português enquadra este fenómeno.

Juridicamente, o **Código Penal português prevê o discurso de ódio como um tipo legal autónomo**, nomeadamente o crime de discriminação e incitamento ao ódio e à violência (artigo 240.º do Código Penal).

> **Artigo 240.º - Discriminação e incitamento ao ódio e à violência**
>
> 1 - Quem:
>
> a) Fundar ou constituir organização ou desenvolver atividades de propaganda que incitem ou encorajem à discriminação, ao ódio ou à violência contra pessoa ou grupo de pessoas em razão da sua origem étnico-racial, origem nacional ou religiosa, cor, nacionalidade, ascendência, território de origem, religião, língua, sexo, orientação sexual, identidade ou expressão de género ou características sexuais, deficiência física ou psíquica; ou
>
> b) Participar nas organizações referidas na alínea anterior, nas atividades por elas empreendidas ou lhes prestar assistência, incluindo o seu financiamento;
>
> é punido com pena de prisão de 1 a 8 anos.
>
> 2 - **Quem, publicamente, por qualquer meio destinado a divulgação**, nomeadamente através da apologia, negação ou banalização grosseira de crimes de genocídio, guerra ou contra a paz e a humanidade:
>
> a) **Provocar atos de violência contra pessoa ou grupo de pessoas por causa da sua origem** étnico-racial, origem nacional ou religiosa, cor, nacionalidade, ascendência,

território de origem, religião, língua, sexo, orientação sexual, identidade ou expressão de género ou características sexuais, deficiência física ou psíquica;

b) **Difamar ou injuriar pessoa ou grupo de pessoas por causa da sua origem** étnico-racial, origem nacional ou religiosa, cor, nacionalidade, ascendência, território de origem, religião, língua, sexo, orientação sexual, identidade ou expressão de género ou características sexuais, deficiência física ou psíquica;

c) **Ameaçar pessoa ou grupo de pessoas por causa da sua origem** étnico-racial, origem nacional ou religiosa, cor, nacionalidade, ascendência, território de origem, religião, língua, sexo, orientação sexual, identidade ou expressão de género ou características sexuais, deficiência física ou psíquica; ou

d) **Incitar à discriminação, ao ódio ou à violência contra pessoa ou grupo de pessoas por causa da sua origem** étnico-racial, origem nacional ou religiosa, cor, nacionalidade, ascendência, território de origem, religião, língua, sexo, orientação sexual, identidade ou expressão de género ou características sexuais, deficiência física ou psíquica;

é punido com pena de prisão de 6 meses a 5 anos.

3 - Quando os crimes previstos nos números anteriores forem cometidos através de sistema informático, o tribunal pode ordenar a eliminação de dados informáticos ou conteúdos.

Analisando-o, verifica-se que existem requisitos necessários para que um comportamento seja considerado como discurso de ódio, pelo que apenas são criminalizados incidentes específicos, atendendo ao potencial dano causado. Assim, os **seguintes pressupostos[16] devem ser considerados**:

1. A criminalização **não afeta condutas que decorram em privado**. A tipificação do ilícito penal **exige que a conduta punível se realize no espaço público e envolva qualquer meio destinado à divulgação**, o que supõe o uso do discurso verbal, o panfleto, a grafitagem, a afixação de cartazes, a utilização da imprensa e de sítios web, bem como a colocação de mensagens na internet fora do âmbito de grupos fechados.
2. Constitui pressuposto da prática do crime que o uso público dos referidos meios de divulgação pelo agente se **destine a fazer "a apologia, negação**

16 Diário da República (n.d.). *Crime de incitamento ao ódio e à violência*. Lexionário. https://diariodarepublica.pt/dr/lexionario/termo/crime-incitamento-ao-odio-a-violencia

ou banalização grosseira de crimes de genocídio, guerra ou contra a paz e a humanidade”.

3. É exigível que o uso dos meios de divulgação destinados a fazer a apologia ou a negação de crimes contra a paz e a humanidade **tenham um efeito ou resultado discriminatório concreto**, traduzido na provocação de atos de violência, na prática dos crimes de injúria ou difamação, na ameaça e no incitamento à violência ou ódio contra “pessoa ou grupo de pessoas por causa da sua raça, cor, origem étnica ou nacional, ascendência, religião, sexo, orientação sexual, identidade de género ou deficiência física ou psíquica”.

Além do mais, o artigo 240.º apresenta uma **lista fechada, ainda que extensa, de motivações baseadas em preconceito que podem fundamentar o crime** de discriminação e incitamento ao ódio e à violência, sendo necessária a violação de pelo menos um destes para estarmos perante um crime desta natureza. Esta conclusão é importante, pois mostra que **outras motivações baseadas em preconceito não previstas no artigo não poderão ser punidas criminalmente**, como o ódio motivado por razões políticas ou associado a subculturas.

Isto acontece porque, segundo o princípio da legalidade penal previsto no artigo 1.º do Código Penal, os indivíduos só podem ser punidos criminalmente por comportamentos “*descritos e declarados passíveis de pena por lei anterior ao momento da sua prática*”, não sendo “*permitido o recurso a analogia para quantificar um facto como crime*”. Estes comportamentos podem, no entanto, ser criminalmente puníveis através de outros crimes, como a difamação (artigo 180.º) e a injúria (artigo 181.º), artigos do mesmo código, não se atendendo, assim, à sua potencial motivação discriminatória.

Após o breve enquadramento legal do discurso de ódio em Portugal, são apresentados, a título expositivo, dois casos reais. O objetivo é ilustrar o que pode ou não ser considerado discurso de ódio ao abrigo da lei portuguesa.

Exemplo 1[17]

- **Contexto:** Dois arguidos, o AA e o BB, foram condenados no tribunal de 1ª Instância, por publicarem no *Twitter/X* comentários como: "*Prostituição forçada das gajas do Bloco [partido político de esquerda]*", "*Tudo, tipo arrastão*", "*Concordo. Incluam as do PCE, MRRP, MAS e PS [partidos políticos de esquerda]* ".

- **Decisão da 1ª Instância:** O tribunal entendeu que as expressões não eram humorísticas, mas sim dirigidas a mulheres de esquerda, com o objetivo de ofender, humilhar e associá-las à prostituição. Como tal, condenou o AA a 2 anos e 10 meses de prisão efetiva e o BB a 1 ano e 8 meses de prisão, suspensa por 2 anos com regime de prova e pagamento de 750€ à APAV [Associação Portuguesa de Apoio à Vítima].

- **Recursos:** Insatisfeito com os resultados da sentença da 1ª instância, o AA interpôs recurso, alegando que as frases não eram sérias, mas sim humorísticas; que não quis ofender todas as mulheres, apenas as ativistas políticas; que houve erro na apreciação da prova; e que deveria ter direito à pena suspensa. Já BB alegou que não ficou provado que o titular da conta era BB; que não houve perícia informática; que houve violação do princípio in *dubio pro reo* e mencionou a necessidade de adequação da medida da pena.

- **Decisão da Relação:** Considerou improcedentes ambos os recursos e manteve as sentenças da 1ª Instância.

- **Justificação da Decisão da Relação:** O tribunal confirmou a autoria das publicações, concluindo que as contas de AA e de BB pertenciam efetivamente aos arguidos, com base em elementos como mensagens,

17 Acórdão do Tribunal da Relação de Lisboa, de 05/12/2024, processo n.º 1633/22.0T9LSB.L1-5. https://www.dgsi.pt/jtrl.nsf/33182fc732316039802565fa00497eec/f9ef3628cf2e149080258bf80055a9c9?OpenDocument&Highlight=0,discrimina%C3%A7%C3%A3o,e,incitamento,ao,%C3%B3dio

emails e fotografias, e validou o recurso à prova indireta através de presunções naturais. Rejeitou a tese de que as expressões tinham caráter humorístico, afirmando que frases como "prostituição forçada das gajas do Bloco" são objetivamente degradantes e dirigidas a um grupo específico de mulheres de esquerda, preenchendo o crime previsto no artigo 240.º, n.º 2, alínea b) do Código Penal. Esclareceu, ainda, que não é necessário ofender todas as mulheres, bastando visar um grupo definido por sexo e ideologia, e concluiu que os arguidos atuaram com dolo direto. Por fim, manteve a pena de prisão efetiva aplicada a AA, sublinhando o seu extenso historial criminal, a proximidade temporal entre o fim da liberdade condicional e os factos e a ausência de arrependimento ou empatia demonstrada em julgamento.

Exemplo 2[18]

- **Contexto:** A arguida publicou no jornal Público um artigo de opinião onde utilizou expressões generalizantes e depreciativas sobre africanos e ciganos, associando-os a comportamentos desviantes, incivilidade, violência e inferioridade cultural. O Movimento SOS Racismo apresentou queixa, considerando que o texto constituía crime de discriminação e incitamento ao ódio (artigo 240.º do Código Penal).

- **Decisão do Juiz de Instrução:** O Ministério Público arquivou o inquérito, entendendo que o texto estava protegido pela liberdade de expressão. O assistente, em resposta, requereu a abertura de instrução.

- **Decisão da Instrução:** O juiz de instrução rejeitou o requerimento de abertura de instrução, por considerar que o requerimento não descrevia factos suficientes para integrar o crime, o texto da arguida constituía mera opinião, protegida pela liberdade de expressão e que não estavam preenchidos os elementos do artigo 240.º do Código Penal.

18 Acórdão do Tribunal da Relação de Lisboa, de 06/07/202, processo n.º 5551/19.0T9LSB.L1-5. https://www.dgsi.pt/jtrl.nsf/33182fc732316039802565fa00497eec/53d43a27fb12dafc802587480047ae12?OpenDocument&Highlight=0,discrimina%C3%A7%C3%A3o,e,incitamento,ao,%C3%B3dio

- **Recurso:** Não satisfeito com a decisão, o assistente recorreu à Relação, alegando que o requerimento descrevia casos suficientes e concretos e que as expressões da arguida eram objetivamente discriminatórias. Acrescentou ainda que o juiz de instrução confundiu liberdade de expressão com discurso de ódio e que o despacho violava o artigo 287.º do Código de Processo Penal e o artigo 240.º do Código Penal. Face a isto, pediu a revogação do despacho e a abertura de instrução ao Tribunal da Relação.

- **Decisão da Relação:** A Relação deu provimento ao recurso, revogou o despacho e determinou a abertura da instrução, considerando o requerimento do assistente formalmente válido e contendo factos suficientes. O tribunal salientou que a liberdade de expressão não legitima discursos de natureza discriminatória, sublinhando que o juiz de instrução fez uma interpretação excessivamente ampla desse direito fundamental.

Em suma, o quadro jurídico português procura equilibrar a proteção da liberdade de expressão com a necessidade de prevenir a discriminação, o ódio e a violência. Assim como ilustrado nos exemplos apresentados, esta distinção nem sempre é simples, especialmente em ambientes digitais, onde o discurso de ódio pode ser apresentado como humor, opinião ou comentário político. Nem todas as manifestações ofensivas ou discriminatórias constituem uma infração penal ao abrigo da lei portuguesa, uma vez que é necessário que sejam cumpridos critérios jurídicos específicos para que a conduta seja abrangida pelo artigo 240.º do Código Penal. Simultaneamente, o sistema jurídico reconhece que a liberdade de expressão não pode ser utilizada para legitimar discursos que prejudiquem a dignidade, a segurança e os direitos de indivíduos ou grupos. Compreender estes limites legais é, por isso, essencial para identificar o discurso de ódio, reconhecer as suas potenciais consequências e promover uma participação mais informada e responsável nos espaços públicos e *online*.

Cofinanciado pela União Europeia

Conclusão

O discurso de ódio *online* não é um fenómeno marginal. Reflete e influencia dinâmicas sociais mais amplas, incluindo a desigualdade, a polarização, a desinformação e a deterioração das normas sociais. Tem impacto nas vítimas, nas pessoas que assistem a estas situações e no funcionamento da sociedade, ao normalizar a exclusão, ao dessensibilizar o público e ao colocar em causa as normas democráticas partilhadas. Este manual destaca a importância de analisar o ódio numa perspetiva integrada e contextualizada, explorando a forma como este funciona e se manifesta, nomeadamente através da desumanização, de narrativas de ameaça e de linguagem codificada. Além disso, fornece ferramentas para reconhecer a escalada do ódio, analisando-o à luz de contextos sociais, políticos e digitais específicos.

O discurso de ódio não surge isoladamente, desenvolvendo-se no seio de ecossistemas moldados pela polarização, desinformação e tensões políticas e sociais. O reconhecimento destas dinâmicas é essencial para respostas proporcionais e responsáveis.

As sociedades democráticas dependem da liberdade de expressão, do debate aberto e do pluralismo. A proteção destes princípios deve andar de mãos dadas com a salvaguarda da dignidade humana, da igualdade e da não discriminação. Ferramentas jurídicas, mecanismos institucionais e iniciativas de sensibilização do público desempenham todos um papel, mas a prevenção depende também da educação, do pensamento crítico e da responsabilidade coletiva.

Em última análise, prevenir e combater o discurso de ódio não se resume apenas a restringir conteúdos prejudiciais. Pelo contrário, implica reforçar a coexistência democrática e o respeito pelos direitos humanos. Ao promover o reconhecimento informado, o envolvimento ético e um discurso público inclusivo, as sociedades podem reduzir as consequências negativas deste fenómeno e reforçar os valores que as sustentam.

Cofinanciado pela União Europeia

Referências

Allport, G. W. (1954). *The Nature of Prejudice*. Addison-Wesley.

Almeida, P., & Pereira, J. (2026). Anticiganismo no Facebook: discursos de ódio e racismo quotidiano em Portugal. *Etnográfica. Revista do Centro em Rede de Investigação em Antropologia*.

Almeida, P., Pereira, J., & Candido, D. (2023). Online hate speech on social media in Portugal: extremism or structural racism? *Social Identities, 29*(5), 419–435. https://doi.org/10.1080/13504630.2024.2324277

Arce-García, S., Said-Hung, E., & Montero-Díaz, J. (2025). Unmasking coordinated hate: Analysing hate speech on Spanish digital news media. *New Media & Society, 27*(10), 5848–5868. https://doi.org/10.1177/14614448241259715

Åkerlund, M. (2021). Dog whistling far-right code words: the case of 'culture enricher' on the Swedish web. *Information, Communication & Society, 25*(12), 1808–1825. https://doi.org/10.1080/1369118x.2021.1889639

Assimakopoulos, S., Baider, F.H., & Millar, S. (2017) Online hate speech in the European Union: a discourse-analytic perspective. Springer Nature

Bahador, B. (2023). Monitoring hate speech and the limits of current definition. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech research* (pp. 291-298). Berlin https://doi.org/10.48541/dcr.v12.17

Baider, F. (2022). Covert hate speech, conspiracy theory and anti-semitism: Linguistic analysis versus legal judgement. International Journal for the Semiotics of Law-Revue, 35, 2347–2371. https://doi.org/10.1007/s11196-022-09882-w

Baider, F. (2023). Accountability Issues, Online Covert Hate Speech, and the Efficacy of Counter-Speech. *Politics and Governance, 11*(2), 249-260. https://doi.org/10.17645/pag.v11i2.6465

Baider, F., & Constantinou, M. (2020). Covert hate speech. *Journal of Language Aggression and Conflict, 8*(2), 262–287. https://doi.org/10.1075/jlac.00040.bai

Bastos, C. (2019). Luso-tropicalism debunked, again. Race, racism, and racialism in three Portuguese-speaking societies. In W. Anderson, R. Roque, & R. Ventura Santos (Eds.), Luso-tropicalism and its discontents: the making and unmaking of racial exceptionalism (pp. 243–264). Berghahn. https://doi.org/978-1-78920-113-0

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). Boston Review. https://doi.org/10.7551/mitpress/9780262012973.003.0002

Benesch, S. (2023). Dangerous speech. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), Challenges and perspectives of hate speech research (pp. 185-197). Berlin https://doi.org/10.48541/dcr.v12.11

Benesch, S., Ruths, D., Dillon K. P., Saleem, H. M. & Wright, L. (2016). *Counterspeech on Twitter/ X: A Field Study*. Dangerous Speech Project. https://dangerousspeech.org/counterspeech-on-Twitter/ X-a-field-study

Bicchieri, C., Muldoon, R. & Sontuoso, A. (2023) "Social Norms", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/archives/win2023/entries/social-norms/

Bilewicz, M. (2025). How Appraisal Model Allows to Distinguish Intergroup Conspiracy Theories from Other Forms of Hate Speech. *Psychological Inquiry, 35*(3-4), 216–222. https://doi.org/10.1080/1047840x.2024.2442919

Bilewicz, M., & Soral, W. (2020). Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology, 41*(S1), 3–33. https://doi.org/10.1111/pops.12670

Bliuc, A.-M., Betts, J. M., Vergani, M., Bouguettaya, A., & Cristea, M. (2024). A theoretical framework for polarization as the gradual fragmentation of a divided society. *Communications Psychology, 2*(1). https://doi.org/10.1038/s44271-024-00125-1

Borinca, I., Van Assche, J., Gronfeldt, B., Sainz, M., Anderson, J., & Taşbaş, E. H. O. (2023). Dehumanization of outgroup members and cross-group interactions. *Current Opinion in Behavioral Sciences, 50*, 101247. https://doi.org/10.1016/j.cobeha.2023.101247

Bozhidarova, M., Chang, J., Ale-Rasool, A., Liu, Y., Ma, C., Bertozzi, A. L., Brantingham, P. J., Lin, J., & Krishnagopal, S. (2023). Hate speech and hate crimes: a data-driven study of evolving discourse around marginalized groups. 2023 IEEE International Conference on Big Data (BigData), 3107–3116. https://doi.org/10.1109/bigdata59044.2023.10386312

Bradshaw, S. (2024). *Disinformation and identity-based violence: Analysis & new insights*. Stanley Center for Peace and Security.

Burnham, S. L. F., Arbeit, M. R., & Hilliard, L. J. (2022). The Subtle Spread of Hateful Memes: Examining Engagement Intentions Among Parents of Adolescents. *Social Media + Society, 8*(2). https://doi.org/10.1177/20563051221095100

Calderón, F. H., Balani, N., Taylor, J., Peignon, M., Huang, Y.-H., & Chen, Y.-S. (2021). Linguistic Patterns for Code Word Resilient Hate Speech Identification. *Sensors, 21*(23), 7859. https://doi.org/10.3390/s21237859

Calderón, C. A., Holgado, P. S., Gómez, J., Barbosa, M., Qi, H., Matilla, A., Amado, P., Guzmán, A., López-Matías, D., & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities. *Humanities and Social Sciences Communications, 11*(1). https://doi.org/10.1057/s41599-024-03899-1

Cardoso, C., Moreno, J., Narciso, I., Couraceiro, P., & dos Santos, J.G.B. (2025). Portuguese General Elections 2025 – Information and Disinformation on Social Media. OberCom – Observatório da Comunicação. https://doi.org/10.5281/zenodo.15781756

Cardoso, M., Ribeiro, E., & Batista, F. (2025). Portuguese Far-Right Discourse on Social Media: Insights from Topic Modeling. In 14th Symposium on Languages, Applications and Technologies (SLATE 2025). Open Access Series in Informatics (OASIcs), 135, 12:1-12:16, Schloss Dagstuhl – Leibniz-Zentrum für Informatik. https://doi.org/10.4230/OASIcs.SLATE.2025.12

Carvalho, P., Caled, D., Silva, M. J. (2025). The Thin Line Between Conspiracy Theories and Opinion: Why Humans and AI Struggle to Differentiate Them. *International Journal of Communication, 19*, 565–591. https://ijoc.org/index.php/ijoc/article/view/22182

Carvalho, P., Caled, D., Silva, C., Batista, F., & Ribeiro, R. (2023). The expression of hate speech against Afro-descendant, Roma, and LGBTQ+ communities in YouTube comments. *Journal of Language Aggression and Conflict, 12*(2), 171-206. . https://doi.org/10.1075/jlac.00085.car

Carvalho, P., Cunha, B., Santos, R., Batista, F., & Ribeiro, R. (2022, June). Hate speech dynamics against African descent, Roma and LGBTQI communities in Portugal. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 2362-2370).

Carvalho, P. & Guerra, R. (2023). "D3.2/D3.3 annotation guidelines OHS & OCS," Knowhate Project—CERV-2021-EQUAL (101049306), Private/Sensitive—Limited Under the Conditions of the Grant Agreement, ISCTE-Inst. Univ. de Lisboa, Lisbon, Portugal, Tech. Rep., May 2023.

Castelo, C. (2021) https://cesa.rc.iseg.ulisboa.pt/afroport/portuguese-non-racism-on-the-historicity-of-an-invented-tradition/.

Čehajić-Clancy, S., & Bilewicz, M. (2021). Moral-exemplar intervention: A new paradigm for conflict resolution and intergroup reconciliation. *Current Directions in Psychological Science*, *30*(4), 335-342. https://doi.org/10.1177/09637214211013001

Chung, Y.-L., Abercrombie, G., Enock, F., Bright, J., & Rieser, V. (2024). Understanding Counterspeech for Online Harm Mitigation. *Northern European Journal of Language Technology, 10*(1). https://doi.org/10.3384/nejlt.2000-1533.2024.5203

Conselho da Europa (2016). *REFERÊNCIAS: Manual para o combate do discurso de ódio online através da educação para os direitos humanos* (Ed. revista). Direção-Geral da Educação. https://www.dge.mec.pt/sites/default/files/ECidadania/educacao_Direitos_Humanos/documentos/referencias_manual_para_o_combate_do_discurso_de_odio_online.pdf.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice". *Journal of Personality and Social Psychology*, *88*(5), 770–789. https://doi.org/10.1037/0022-3514.88.5.770

Couperus, S., Rensmann, L., & Tortola, P. D. (2023). Historical legacies and the political mobilization of national nostalgia: Understanding populism's

relationship to the past. *Journal of Contemporary European Studies, 31*(2), 253–267. https://doi.org/10.1080/14782804.2023.2207480

Cover, R., Beckett, J., Brevini, B., Lumby, C., Simcock, R., & Thompson, J. D. (2025). Reporting online abuse to platforms: Factors, interfaces and the potential for care. *Convergence: The International Journal of Research into New Media Technologies, 32*(1), 142–158. https://doi.org/10.1177/13548565251324508

Dare to be Grey. (2024). *Reporting models of online hate*. European Observatory of Online Hate. https://eooh.eu/articles/reporting-models-online-hate-speech

de Coster, S., Veilleux-Lepage, Y., Amarasingam, A., & Abbas, T. (2024). Uncovering the Bias and Prejudice in Reporting on Islamist and Non-Islamist Terrorist Attacks in British and US Newspapers. *Perspectives on Terrorism, 18*(3). https://doi.org/10.19165/2024.3034

de Roos, M. S., & Caon, G. (2026). The radicalisation loop: A layered linguistic model of disengagement and re-engagement in an incel forum. *Current Psychology, 45, 358*. https://doi.org/10.1007/s12144-025-08944-z

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In *The SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 3–29). London, England: Sage.

Dreißigacker, A., Müller, P., Isenhardt, A., & Schemmel, J. (2024). Online hate speech victimization: consequences for victims' feelings of insecurity. *Crime Science, 13*(4). https://doi.org/10.1186/s40163-024-00204-y

European Union Agency for Fundamental Rights - FRA (2023). *Online content moderation - Current challenges in detecting hate speech* (Catalogue No. TK-04-23-883-EN-C). Publications Office of the European Union. https://fra.europa.eu/en/publication/2023/online-content-moderation

EUAFR - European Union Agency for Fundamental Rights (2025). Fundamental rights report – 2025. https://fra.europa.eu/en/publication/2025/fundamental-rights-report-2025

European Commission Against Racism and Intolerance. (2025). *Sixth report on Portugal* [Report]. Council of Europe.

European Institute for Gender Equality. (2025). Country profile for Portugal. https://eige.europa.eu/gender-based-violence/countries/portugal?language_content_entity=en

Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why we hate. *Emotion Review*, *10*(4), 309-320. https://doi.org/10.1177/1754073917751229

Fonseca, A., Pontes, C., Moro, S., Batista, F., Ribeiro, R., Guerra, R., Carvalho, P., Marques, C., & Silva, C. (2024). Analyzing hate speech dynamics on Twitter/X: Insights from conversational data and the impact of user interaction patterns. *Heliyon, 10*(11), e32246. https://doi.org/10.1016/j.heliyon.2024.e32246

Fortuna, P., & Nunes, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Computing Surveys, 51*(4), 1–30. https://doi.org/10.1145/3232676

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Psychology Press.

Gaertner, S. L., Dovidio, J. F., Guerra, R., Hehman, E., & Saguy, T. (2016). A common ingroup identity: Categorization, identity, and intergroup relations. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (2nd ed., pp. 433–454). Psychology Press.

Gagliardone, I., Gal, D., Alves, T. & Martinez, G. (2015). *Countering online hate speech*. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000233231

Gallacher, J. D., Heerdink, M. W., & Hewstone, M. (2021). Online Engagement Between Opposing Political Protest Groups via Social Media is Linked to Physical Violence of Offline Encounters. *Social Media + Society, 7*(1), 1-16. https://doi.org/10.1177/2056305120984445

Geetanjali, & Kumar, M. (2025). Exploring hate speech detection: challenges, resources, current research and future directions. *Multimedia Tools and Applications, 84*, 38423–38459. https://doi.org/10.1007/s11042-025-20716-2

Gennaro, G., Derksen, L., Abdelrahman, A. *et al.* (2025). Counterspeech encouraging users to adopt the perspective of minority groups reduces hate speech and its amplification on social media. *Scientific Reports,* 15, 22018. https://doi.org/10.1038/s41598-025-05041-w

Ghenai, A., Noorian, Z., Moradisani, H., Abadeh, P., Erentzen, C., & Zarrinkalam, F. (2025). Exploring hate speech dynamics: The emotional, linguistic, and thematic impact on social media users. *Information Processing & Management, 62*(3), 104079. https://doi.org/10.1016/j.ipm.2025.104079

Greene, J. (2025). *Malinformation*. EBSCO Research Starters. https://www.ebsco.com/research-starters/information-technology/malinformation#full-article

Guerra, R., Carvalho, P., Marques, C., Carmona, M., Sarroeira, R., Batista, F., Ribeiro, R., Fonseca, A., Moro, S., & Silva, C. (2025). Unpacking online hate speech in Portuguese social media: a social-psychological and linguistic-discursive approach. *Humanities and Social Sciences Communications, 12*, 1709. https://doi.org/10.1057/s41599-025-05392-9

Harel, T. O., Jameson, J. K., & Maoz, I. (2020). The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict. *Social Media + Society, 6*(2), 1-10. https://doi.org/10.1177/2056305120913983

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and social psychology review*, *10*(3), 252-264. https://doi.org/10.1207/s15327957pspr1003

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65* (1), 399-423. https://doi.org/10.1146/annurev-psych-010213-115045

Hassan, G., Rabah, J., Madriaza, P., Brouillette-Alarie, S., Borokhovski, E., Pickup, D., Varela, W., Girard, M., Durocher-Corfa, L., & Danis, E. (2022). PROTOCOL: Hate online and in traditional media: A systematic review of the evidence for associations or impacts on individuals, audiences, and communities. *Campbell Systematic Reviews, 18*(2), e1245. Portico. https://doi.org/10.1002/cl2.1245

Herek, G. M. (2004). Beyond "homophobia": Thinking about sexual prejudice and stigma in the twenty-first century. *Sexuality Research and Social Policy*, *1*(2), 6–24. https://ssrn.com/abstract=1142860

Homolar, A., & Löfflmann, G. (2021). Populism and the Affective Politics of Humiliation Narratives. *Global Studies Quarterly, 1*(1), 1–11. https://doi.org/10.1093/isagsq/ksab002

Inwood, O., & Zappavigna, M. (2023). Conspiracy Theories and White Supremacy on YouTube: Exploring Affiliation and Legitimation Strategies in YouTube Comments. *Social Media + Society, 9*(1), 1-16. https://doi.org/10.1177/20563051221150410

Isasi, A. C., & Juanatey, A. G. (2016). *Hate speech in social media: A state-of-the-art review*. Ajuntament de Barcelona. https://ajuntament.barcelona.cat/bcnvsodi/wp-content/uploads/2017/02/Informe_discurso-del-odio_ES-en-GB.pdf

Karantzas, G. C., & Simpson, J. A. (2026). The Perpetration of Dehumanization: A Systematic Review. *Current Opinion in Psychology, 70,* 102309. https://doi.org/10.1016/j.copsyc.2026.102309

Katz-Wise, S. L., & Hyde, J. S. (2012). Victimization experiences of lesbian, gay, and bisexual individuals: A meta-analysis. *Journal of Sex Research, 49*(2-3), 142–167. https://doi.org/10.1080/00224499.2011.637247

Kearns, C., Sinclair, G., Black, J., Doidge, M., Fletcher, T., Kilvington, D., Liston, K., Lynn, T., & Rosati, P. (2022). A Scoping Review of Research on Online Hate and Sport. *Communication & Sport, 11*(2), 402–430. https://doi.org/10.1177/21674795221132728

Kentmen-Cin, C. (2025). Hate Speech on Social Media: A Systemic Narrative Review of Political Science Contributions. *Social Sciences, 14*(10), 610. https://doi.org/10.3390/socsci14100610

Kteily, N., & Bruneau, E. (2017). Backlash: The politics and real-world consequences of minority group dehumanization. *Personality and Social Psychology Bulletin, 43*(1), 87–104. https://doi.org/10.1177/0146167216675334

Mackie, D. M., & Smith, E. R. (2015). Intergroup emotions. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, & J. A. Simpson (Eds.), *APA handbook of personality and social psychology, Vol. 2. Group processes* (pp. 263–293). American Psychological Association. https://doi.org/10.1037/14342-010

Madriaza, P., Hassan, G., Brouillette-Alarie, S., Mounchingam, A. N., Durocher-Corfa, L., Borokhovski, E., Pickup, D., & Paillé, S. (2025). Exposure to hate in online and traditional media: A systematic review and meta-analysis of the impact of this exposure on individuals and communities. *Campbell Systematic Reviews, 21*, e70018. https://doi.org/10.1002/cl2.70018

Magu, R., & Luo, J. (2018). Determining Code Words in Euphemistic Hate Speech Using Word Embedding Networks. *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*. https://doi.org/10.18653/v1/w18-5112

Mannocci, L., Mazza, M., Monreale, A., Tesconi, M., & Cresci, S. (2024). Detection and Characterization of Coordinated Online Behavior: A Survey. arXiv preprint arXiv:2408.01257. https://doi.org/10.48550/arXiv.2408.01257

Mansur, Z., Omar, N., Tiun, S., & Alshari, E. M. (2024). A normalization model for repeated letters in social media hate speech text based on rules and spelling correction. *Plos One, 19*(3), e0299652. https://doi.org/10.1371/journal.pone.0299652

McDonald, R. I., & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, *3*, 147-151. http://dx.doi.org/10.1016/j.cobeha.2015.04.006

Miranda, S., Gouveia, C., Di Fátima, B., & Antunes, A. C. (2023). Hate speech on social media: behaviour of Portuguese football fans on Facebook. *Soccer & Society, 25*(1), 76–91. https://doi.org/10.1080/14660970.2023.2230452

Montesinos-Cánovas, E., García-Sánchez, F., García-Díaz, J. A., Alcaraz-Mármol, G., & Valencia-García, R. (2023). Spanish hate-speech detection in football. *Procesamiento del Lenguaje Natural*, 71, 15-27.

Morales, E., Hodson, J., O'Meara, V., Gruzd, A., & Mai, P. (2025). Online toxic speech as positioning acts: Hate as discursive mechanisms for othering and belonging. *New Media & Society,* 1-19. https://doi.org/10.1177/14614448251338493

Mossou, S., & Lane, A. (2018). *Anti-migrant hate speech*. Quaker Council for European Affairs. https://www.qcea.org/wp-content/uploads/2018/06/Hate-Speech-Report_final.pdf

Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, *19*(4), 2131-2167. https://doi.org/10.1093/jeea/jvaa045

Murib, Z. (2022). Don't Read the Comments: Examining Social Media Discourse on Trans Athletes. *Laws, 11*(4), 53. https://doi.org/10.3390/laws11040053

National Coordinator for Counterterrorism and Security (NCTV). (2024). *Memes as an online weapon. An analysis into the use of memes by the far right*. https://english.nctv.nl/publications/reports/2024

Neves, S., Borges, J., Ferreira, M., Correia, M., Sousa, E., Rocha, H., Silva, L., Allen, P., & Vieira, C. P. (2023). A literature review on violence and discrimination against trans people in Portugal: Are we still living in a dictatorship? *Sexualities*, 28(1–2), 349–364. https://doi.org/10.1177/13634607231197059

Papcunová, J., Martončik, M., Fedáková, D., Kentoš, M., Bozogáňová, M., Srba, I., Moro, R., Pikuliak, M., Šimko, M., & Adamkovič, M. (2021). Hate speech operationalization: a preliminary examination of hate speech indicators and their structure. *Complex & Intelligent Systems, 9*(3), 2827–2842. https://doi.org/10.1007/s40747-021-00561-0

Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *SAGE Open, 10*(4), 1-15. https://doi.org/10.1177/2158244020973022

Pickles, J. (2020). Sociality of hate: The transmission of victimization of LGBT+ people through social media. *International Review of Victimology, 27*(3), 311–327. https://doi.org/10.1177/0269758020971060

Pontes, C., Fonseca, A., Moro, S., Batista, F., Ribeiro, R., Marques, C., Carvalho, P., Silva, C., & Guerra, R. (2024). Unveiling Patterns of Hate Speech in the Portuguese Sphere: A Social Network Analysis Approach. *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 70–81. https://doi.org/10.1007/978-3-031-73997-2_7

Prabhu, R., & Seethalakshmi, V. (2025). A comprehensive framework for multi-modal hate speech detection in social media using deep learning. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-025-94069-z

Said-Hung, E., Moreno-López, R., & Mottareale-Calvanese, D. (2023). Promotion of hate speech by Spanish political actors on Twitter. *Policy & Internet, 15*(4), 665–686. Portico. https://doi.org/10.1002/poi3.353

Sanchez-Mazas, M., & Licata, L. (2015). Xenophobia: Social psychological aspects. *International Encyclopedia of the Social & Behavioral Sciences*, *25*, 802-807. https://doi.org/10.1016/B978-0-08-097086-8.24031-2

Silva, C., & Carvalho, P. (2023). When can compliments and humour be considered hate speech? A perspective from target groups in Portugal. *Comunicação e sociedad*e, 43, e023006. https://doi.org/10.17231/comsoc.43(2023).4135

Smith, L. G. E., Thomas, E. F., Bliuc, A.-M., & McGarty, C. (2024). Polarization is the psychological foundation of collective engagement. *Communications Psychology, 2*(1). https://doi.org/10.1038/s44271-024-00089-2

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior, 44*(2), 136–146. Portico. https://doi.org/10.1002/ab.21737

Sousa, Y., & Cabecinhas, R. (2025). Estereótipos Sociais. *Psicologia Social, Comunicação e Cultura*, *69*. https://doi.org/10.21814/uminho.ed.157.6

Šrol, J., Čavojová, V., & Ballová Mikušková, E. (2022). Finding Someone to Blame: The Link Between COVID-19 Conspiracy Beliefs, Prejudice, Support for Violence, and Other Negative Social Outcomes. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.726076

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), Reducing prejudice and discrimination, (pp. 23-46). Lawrence Erlbaum Associates Publishers.

Stephan, W. G., & Stephan, C. W. (2016). Intergroup Threats. The Cambridge Handbook of the Psychology of Prejudice, 131–148. https://doi.org/10.1017/9781316161579.00

Taha, R., & Sailofsky, D. (2025). 'But she's not even trans!': A rhetorical analysis of 'liberal feminist' defences of Imane Khelif amid Olympic transvestigations. *International Review for the Sociology of Sport*. https://doi.org/10.1177/10126902251371324

Theofilopoulos, T. (Ed.). (2024). *Online hate speech patterns in media platforms' comments sections: Cyprus, France, Greece, Italy* [Report]. Symplexis. https://cesie.org/media/chase-online-hate-speech-patterns.pdf

Tindale, C. W. (2007). *Fallacies and Argument Appraisal*. Cambridge University Press. https://doi.org/10.1017/cbo9780511806544

Udupa, S. (2025). *Extreme speech and hate hide in everyday chats*. Ludwig-Maximilians-Universität München. https://www.lmu.de/en/newsroom/news-overview/news/extreme-speech-and-hate-hide-in-everyday-chats-7805fcd7.html

Uyheng, J., & Carley, K. M. (2020). Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. *Journal of Computational Social Science, 3*(2), 445–468. https://doi.org/10.1007/s42001-020-00087-4

UNESCO & United Nations Office on Genocide Prevention and the Responsibility to Protect. (2023). *Addressing hate speech through education: A guide for policy-makers*. https://unesdoc.unesco.org/ark:/48223/pf0000384872.

Valentim, J. P. (2011). Social psychology and colonialism: Luso-tropicalism as a social representation in the context of contemporary Portuguese society. Em J. P. Valentim (Ed.). Societal approaches in social psychology (pp.179-194). Berne: Peter Lang.

van Dijk, T. A. (1992). Discourse and the Denial of Racism. *Discourse & Society, 3*(1), 87–118. https://doi.org/10.1177/0957926592003001005

van Dijk, T. A. (2023). (Anti)Racist discourse. The Routledge Handbook of Discourse Analysis, 244–260. https://doi.org/10.4324/9781003035244-20

van Prooijen, J.-W. (2021). Injustice Without Evidence: The Unique Role of Conspiracy Theories in Social Justice Research. *Social Justice Research, 35*(1), 88–106. https://doi.org/10.1007/s11211-021-00376-x

van Prooijen, J.-W., & Douglas, K. M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies, 10*(3), 323-333. https://doi.org/10.1177/1750698017701615

Vergani, M., Perry, B., Freilich, J., Chermak, S., Scrivens, R., Link, R., Kleinsman, D., Betts, J., & Iqbal, M. (2024). Mapping the scientific knowledge and approaches to defining and measuring hate crime, hate speech, and hate incidents: A systematic review. *Campbell Systematic Reviews, 20*(2). Portico. https://doi.org/10.1002/cl2.1397

Wagoner, B., Jørgensen, M. S., & Pahuus, K. (2026). Conspiracy theories through the lens of collective memory. *Current Opinion in Psychology, 68*, 102227. https://doi.org/10.1016/j.copsyc.2025.102227

Wang, M., Ma, S., Li, N., Zhang, P., Li, C., Gu, N., & Lu, T. (2026). Echoes of Norms: Investigating Counterspeech Bots' Influence on Bystanders in Online Communities. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (pp. 1-23).

Wardle, C. (2024). *A conceptual analysis of the overlaps and differences between hate speech, misinformation and disinformation*. United Nations, Department of Peace Operations and Office of the Special Adviser on the Prevention of Genocide. https://peacekeeping.un.org/sites/default/files/report_-_a_conceptual_analysis_of_the_overlaps_and_differences_between_hate_speech_misinformation_and_disinformation_june_2024.pdf

Weber, I., Vandebosch, H., Poels, K., & Pabian, S. (2023). Features for Hate? Using the Delphi Method to Explore Digital Determinants for Online Hate Perpetration and Possibilities for Intervention. *Cyberpsychology, Behavior, and Social Networking, 26*(7), 479–488. https://doi.org/10.1089/cyber.2022.0195

Williams, M. (2022). *The Science of Hate: How Prejudice Becomes Hate and What We Can Do to Stop It*. Faber & Faber, Ltd

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology, 60,* 93-117. https://doi.org/10.1093/bjc/azz049

Wodak, R. (2015). The Politics of Fear: What Right-Wing Populist Discourses Mean. London: Sage. https://doi.org/10.4135/9781446270073

Zendeli, E. (2017). The right to education as a fundamental human right. *Contemporary Educational Researches Journal, 7*(4), 158–166. https://doi.org/10.18844/cerj.v7i4.2718

Zhang, C. (2018). *WeChatting American Politics: Misinformation, Polarization, and Immigrant Chinese Media* (TOW Reports). Tow Center for Digital Journalism. https://www.cjr.org/tow_center_reports/wechatting-american-politics-misinformation-polarization-and-immigrant-chinese-media.php/

Cofinanciado pela
União Europeia

Anexos

A - Reconhecer o Discurso de Ódio: Indicadores Linguísticos, Simbólicos e Contextuais

Este anexo complementa as dimensões analíticas apresentadas na secção 4, centrando-se no reconhecimento prático do discurso de ódio em contextos do mundo real. Reúne uma série de **marcadores linguísticos, simbólicos e contextuais** que podem indicar a presença de expressões de ódio ou discriminatórias, incluindo tanto formas amplamente utilizadas como específicas de cada contexto. Estes marcadores não devem ser entendidos como indicadores fixos ou exaustivos, mas sim como elementos em evolução cujo significado depende do uso, da intenção e do contexto. Os exemplos apresentados visam apoiar a sensibilização e a interpretação, ajudando os leitores a identificar como o discurso de ódio pode manifestar-se em diferentes ambientes comunicativos.

Tabela A1. Indicadores Gerais de Discurso de Ódio em Várias Plataformas

CATEGORIA DA PLATAFORMA	TIPO DE MARCADOR	DESCRIÇÃO	EXEMPLO ILUSTRATIVO
PLATAFORMAS DE PARTILHA DE VÍDEOS (*YOUTUBE, TWITCH, DISCORD, ROBLOX*)	Linguagem de ódio transcrita ou legendada	Expressões de ódio que aparecem em títulos de vídeos, descrições, legendas, legendas geradas automaticamente ou transcrições manuais de falas usadas para análise.	Comentários depreciativos dirigidos a um grupo nas legendas de vídeos ou nas legendas automáticas.
	Ódio motivado pelo envolvimento / desencadeado por comentários	Discurso de ódio que surge nas secções de comentários como reação ao conteúdo do vídeo, títulos ou enquadramento da	Comentários de ódio ou desumanizantes dirigidos a grupos mencionados no vídeo,

		pessoa influenciadora; muitas vezes amplificado através de respostas, "gostos" ou comentários repetidos.	intensificados através de comentários após a publicação.
	Simbolismo visual	Logótipos, bandeiras, esquemas de cores ou memes visuais que sinalizam identidade ideológica ou exclusão.	Utilização de símbolos extremistas ou de imagens de bandeiras históricas.
REDES SOCIAIS (*X, INSTAGRAM, TIKTOK, FACEBOOK*)	*Hashtags, Emojis* & Códigos numéricos	Comunicação codificada usando *hashtags, emojis*, números ou combinações que sinalizam afiliação ideológica ou narrativas de exclusão.	#1488, 🐀, ou combinações de *emojis* repetidamente utilizados em grupos de ódio.
	Combinação de elementos visuais e texto	Combinações de imagens e legendas que transmitem sarcasmo ou superioridade.	*Meme* que contrasta "nós" contra "eles".
	Nomes de utilizador, nomes de grupos & descrições/biografias	Nomes de conta ou descrições de perfil que contenham referências extremistas, termos codificados de ódio ou *slogans* excludentes relevantes para o conteúdo publicado.	Nomes de utilizador incluindo "Reconquista", "Portugueses Primeiro", "1143", "*Blood & Honour*" "*IncelsExit*" ou títulos

			semelhantes codificados.
PLATAFORMAS DE MENSAGENS INSTANTÂNEAS (*TELEGRAM, WHATSAPP*, GRUPOS DO *DISCORD*)	Mídia reencaminhado	Circulação de vídeos editados, *stickers* ou *memes* que reforçam estereótipos.	*GIFs* que sugerem desumanização.
FÓRUNS & BLOGUES	Enquadramento narrativo	Publicações extensas que recorrem ao pânico moral, à pseudociência ou a narrativas conspirativas.	"Eles controlam os meios de comunicação social e os bancos."
	Assinaturas & avatares dos utilizadores	Elementos de identidade visual ou citações que revelam uma perspetiva ideológica.	Imagens históricas do fascismo.
NOTÍCIAS & SECÇÕES DE COMENTÁRIOS DA IMPRENSA	Enquadramento & padrões de reação	Comentários que incitam ao ódio em reação a temas noticiosos específicos.	"Da próxima vez, é melhor que a polícia não vá... que se matem uns aos outros. "Se deportarmos os africanos, a criminalidade diminuirá 40 % em Portugal!"

Tabela A2. Marcadores Contextuais: Exemplos de Portugal

Esta tabela apresenta termos, códigos e referências que têm sido associados a discursos extremistas ou relacionados com o ódio no contexto português. A inclusão destes termos tem fins estritamente de investigação, análise e prevenção, não implicando qualquer tipo de aprovação. A presença de qualquer marcador, por si só, não é suficiente para classificar um conteúdo como discurso de ódio; é necessária uma análise contextual e cumulativa.

SÍMBOLO / MARCADOR	CONTEXTO DE UTILIZAÇÃO EM PORTUGAL	NOTA ILUSTRATIVA
CRUZ CELTA (REFERÊNCIA TEXTUAL) **(E.G., LOGO CLUBE ÉTNICO PORTUGUEZ)**	Referenciada em nomes de utilizador, nomes de grupos, *slogans* ou descrições escritas associadas a grupos neonazis ou nacionalistas brancos.	Aparece em publicações ou biografias combinadas com *slogans* excludentes, como "Portugal aos Portugueses", muitas vezes sinalizando um alinhamento ideológico.
CÓDIGO NUMÉRICO "1143" 1143	Simboliza a independência portuguesa e é usado em nomes de utilizador, *hashtags*, títulos de grupos ou *slogans* por grupos nacionalistas ou identitários (e.g., referências ao «Grupo 1143»).	Funciona como um identificador nacionalista codificado; requer coocorrência contextual com narrativas excludentes para ser sinalizado.
CÓDIGOS "14 – 88"	Abreviações numéricas que aparecem em *hashtags*, nomes de utilizador ou comentários associados à ideologia da supremacia branca.	Sinal ideológico forte (e.g., *hashtag* "#1488").

"ESCUDO IDENTITÁRIO"	Símbolo do movimento identitário juvenil português. Utilizado em nomes de grupos, *slogans* de campanha ou textos relacionados com protestos.	Panfletos contra o que chamam de "ideologia de género" e o seu "perigo".
"PATRIOTS88" OU *PATRIOTS NETWORK*/ PATRIOTAS PELA EUROPA	Nomes de canais, *hashtags* ou descrições que façam referência a redes extremistas nacionalistas ou transnacionais.	Página @PatriotsNet_Off X e conferência da *Patriots Foundation* em Portugal.
GRUPO "*BLOOD AND HONOUR* PORTUGAL"	Um grupo neonazi que organiza eventos internacionais, como concertos, que servem essencialmente como locais de radicalização, recrutamento e financiamento das suas atividades, incluindo a produção de propaganda.	Um ataque contra o elenco da peça "Amor é um fogo que arde sem se ver", à porta do teatro A Barraca, em Lisboa, terá sido perpetrado pelo grupo neonazi violento Blood & Honour.
"*INCEL*" / TERMINOLOGIA RELACIONADA COM O FENÓMENO *INCEL*	Utilizado principalmente em espaços online e na cultura digital jovem, em particular nas redes sociais, fóruns e plataformas de jogos. Associado a narrativas misóginas e antifeministas.	Referências a "celibatários involuntários" ("*incels*"), em que alguns utilizadores manifestam hostilidade para com as mulheres e o feminismo, frequentemente através de linguagem codificada, *memes*, *emojis* ou humor irónico. Exemplos de *emojis*: 💯 – Associado à "regra 80/20", uma teoria

		distorcida propagada por grupos *incel* que sugere que 80% das mulheres estão interessadas apenas em 20% dos homens. 💥 – Um "comprimido vermelho explosivo", ou seja, alguém que é um *incel* radicalizado. 🫘 – Utilizado por comunidades *incel* para autoidentificação e para rotular outros que partilham as mesmas opiniões.
SIMBOLISMO/IDEOLOGIA DO "COMPRIMIDO VERMELHO" (*RED PILL*), "COMPRIMIDO AZUL" (*BLUE PILL*) E DO "COMPRIMIDO PRETO" (*BLACK PILL*)	Usado em comunidades "*incel*" *online* de adolescentes e jovens adultos (e.g., *TikTok, Discord*, fóruns da manosfera) para promover visões de mundo antifeministas e hierárquicas em termos de género.	Vídeos do *TikTok* ou *YouTube* e *youTubers* afirmando que os homens devem "acordar" do feminismo. 💊 Comprimido Azul: significa permanecer ignorante em relação ao "mundo real" vivido pelos *Incels*. 💊 Comprimido Vermelho: significa "despertar" para o mundo real, onde as mulheres têm uma vantagem e a opressão feminina é um mito. 💊 Comprimido Preto: significa que aderem às ideias subjacentes ao

		"Comprimido Vermelho", mas não acreditam que a sociedade vá mudar ou que a vida dos *Incels* possa melhorar.
MANOSFERA / IDEOLOGIA ANTIFEMINISTA 💊 **(UTILIZADO PARA INDICAR A PERTENÇA À "MANOSFERA")**	Usado em comentários no *YouTube, TikTok* e grupos do *Telegram* para deslegitimar o ativismo feminista, retratar as mulheres como manipuladoras ou moralmente inferiores e para promover violência real ou simbólica contra as mulheres, mas sempre de natureza sexual.	Citações como "Mulheres que namoram não saem à noite", usadas em páginas do *TikTok* portuguesas. Seguir ou ser fã de influenciadores como Numeiro, cuja filosofia de vida está interligada com ideias misóginas espalhadas online.
EMOJIS MISÓGINOS, INSULTOS E TERMOS CODIFICADOS / REFERÊNCIAS A "ADOLESCÊNCIA" (SÉRIE DA *NETFLIX*)	Insultos sexistas, gírias ou frases ou símbolos aparentemente "engraçados" ou inocentes que circulam em plataformas direcionadas aos jovens.	🍆🌭🍌🌽: uma beringela, um cachorro-quente, uma banana ou uma espiga de milho são os emojis usados para se referir ao órgão sexual masculino. 🌷🌮👄🐱: o órgão feminino é "representado" por uma flor, um taco, uma peça de sushi ou um gato.

***Fontes**: Global Project Against Hate and Extremism (2025); Sic Notícias; Público; Expresso; Diário de Notícias; Notícias ao Minuto; UN Women; nit; HM Government.*

COOPERHATE

COMBATE AO DISCURSO DE ÓDIO

O projeto Cooperhate, ao abrigo do Acordo de Subvenção n.º 101213938, é financiado pela União Europeia. As visões e opiniões expressas neste documento são, no entanto, da exclusiva responsabilidade dos autores e não refletem necessariamente as da União Europeia.

A União Europeia não pode ser responsabilizada por elas.

Cofinanciado pela União Europeia