

# **Contranarrativas, Narrativas Alternativas e Discurso de Ódio na Era Digital**

**Jonas David Marinho Batista**

**Dissertação  
de Mestrado em Novos Media e Práticas Web**

**Versão corrigida e melhorada após a sua defesa pública**

**Lisboa (Agosto, 2024)**

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Novos Media e Práticas Web, realizada sob a orientação científica de Marisa Torres da Silva e a coorientação de Paulo Nuno Vicente, Departamento de Ciências da Comunicação da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa.

# Contranarrativas, Narrativas Alternativas e Discurso de Ódio na Era Digital

Jonas David Marinho Batista

## RESUMO

Na era digital, o discurso de ódio renova-se e reafirma-se como fenómeno, desafio, ameaça e problema, com manifestações e impactos reais, sendo mais difícil de apreender e combater. Neste cenário, esta dissertação pretende contribuir para a compreensão tanto do discurso de ódio na atualidade quanto das estratégias de ação implementadas para o enfrentar e mitigar, desacreditando ou desconstruindo as narrativas nas quais se baseia. Mais ainda, e alicerçada em ambas, para estimular abordagens inovadoras e proativas, fundadas na solução e na ação, em linha e fora dela, enquanto prioridade necessária, senão mesmo imperativa. Com recurso a uma metodologia crítica e reflexiva, comparativa e sistemática, exploratória e ideativa, este trabalho promove (i) o estado da arte de conceitos de referência em contexto, (ii) o mapeamento, a identificação e a análise de projetos, ferramentas e recursos-chave fundados em *contranarrativas* e *narrativas alternativas*, e (iii) a adoção de um modelo alternativo ao dominante combate direto e reativo ao discurso de ódio. Em relação à alteração de paradigma são delineadas três propostas, destacando-se a de ideação e operacionalização, em grande escala, da metodologia e dos princípios essenciais das comunicações baseadas na esperança, tendo em vista o seu desenvolvimento e a sua aplicação prática, em particular, nas e com as comunidades locais, reconhecendo-as e capacitando-as para serem a narrativa nos espaços comuns e digitais.

Palavras-chave: contranarrativas, narrativas alternativas, comunicações baseadas na esperança, discurso de ódio, era digital

# **Counter-narratives, Alternative Narratives and Hate Speech in the Digital Era**

**Jonas David Marinho Batista**

## **ABSTRACT**

In the digital era, hate speech is renewed and reaffirmed as a phenomenon, a challenge, a threat and a problem, with real manifestations and impacts, and therefore more difficult to apprehend and combat. In this backdrop, this dissertation aims to contribute to understanding both hate speech today and the action strategies implemented to confront and mitigate it, by discrediting or deconstructing the narratives on which it is based. Moreover, and based on both, to stimulate innovative and proactive approaches based on solutions and action, both online and offline, as a necessary, if not imperative, priority. Using a critical and reflective, comparative and systematic, exploratory and ideational methodology, this work promotes (i) the state of the art of reference concepts in context, (ii) the mapping, identification and analysis of key projects, tools and resources based on counter-narratives and alternative narratives, and (iii) the adoption of an alternative model to the dominant direct and reactive combat against hate speech. Regarding the paradigm shift, three proposals are outlined, highlighting the ideation and large-scale operationalisation of the methodology and essential principles of hope-based communications, with a view to their development and practical application, particularly in and with local communities, recognising and empowering them to be the narrative in common and digital spaces.

Keywords: counter-narratives, alternative narratives, hope-based communications, hate speech, digital era

## Nota prévia

A presente dissertação é escrita ao abrigo do novo acordo ortográfico e recorre à tradução livre de originais em línguas estrangeiras para língua portuguesa.

# Índice

<b>Introdução.....</b>	<b>6</b>
<b>Capítulo I. Revisão de Literatura .....</b>	<b>8</b>
<b>I. 1. Definições conceituais .....</b>	<b>8</b>
<b>I. 1.1 Discurso de Ódio.....</b>	<b>8</b>
I. 1.1.1. Discurso de Ódio na Era Digital .....	12
<b>I. 1.2. Narrativa .....</b>	<b>26</b>
I. 1.2.1 Narrativa digital.....	35
<b>I. 1.3 Contranarrativa .....</b>	<b>39</b>
<b>I. 1.4 Narrativa alternativa .....</b>	<b>44</b>
<b>I. 1.5 Contranarrativas, Narrativas Alternativas e Discurso de Ódio .....</b>	<b>46</b>
<b>I. 1.6 Comunicações baseadas na esperança .....</b>	<b>52</b>
<b>Capítulo II. Mapeamento de projetos, ferramentas e recursos .....</b>	<b>55</b>
<b>II. 1. Objetivos e questões de investigação.....</b>	<b>55</b>
<b>II. 2. Estratégias de investigação .....</b>	<b>56</b>
<b>II. 3. Projetos, ferramentas e recursos selecionados .....</b>	<b>57</b>
<b>II. 4. Análise comparativa: pontos-chave comuns.....</b>	<b>60</b>
<b>Capítulo III. Problemática e contextualização .....</b>	<b>63</b>
<b>III. 1. Contextualização do Projeto .....</b>	<b>63</b>
<b>III. 2. Processo de Ideação .....</b>	<b>65</b>
<b>Capítulo IV. Considerações finais.....</b>	<b>69</b>
<b>Referências bibliográficas .....</b>	<b>71</b>

## Introdução

O *discurso de ódio* é um fenómeno antigo que, de alguma forma, se emancipou e fortaleceu durante as múltiplas fases e diversos processos de transformação digital, ou digitalização, constituindo-se, na atualidade, como um desafio significativo e um problema de importância social (Maarouf et al., 2023). No presente, tanto o potencial quanto os desafios e perigos da Internet são inegáveis. A *era digital* afigura-se, concomitantemente, a era de *viralização do ódio* (Maarouf et al., 2023) e de *profunda polarização* (Vasist et al., 2024). A variedade de estilos, retóricas e dispositivos que assume nos espaços digitais, em particular, nas plataformas de *media* sociais, tornam o *discurso de ódio* eficaz, apelativo e, cada vez mais, árduo — mas não impossível — de detetar e de combater (Pukallus & Arthur, 2024). O já «[...] longo caminho a percorrer no que diz respeito à [sua] erradicação» (Almeida et al., 2023, p. 432) parece estender-se ainda mais e encher-se de obstáculos e perigos renovados. A vivacidade da discussão teórica e da produção académica e em outros domínios, assim como os esforços desenvolvidos, nomeadamente, sob a forma de programas de «disrupção» e «contra-mensagens» (Haider, 2022, p. 4), destacando-se as *contranarrativas* e as *narrativas alternativas* enquanto *instrumentos vitais de resistência* (Silva, 2023), são meritórios, mas insuficientes.

Posto isto, a presente dissertação pretende contribuir para a compreensão do fenómeno do *discurso de ódio* no presente e, com base nela, para o desconstruir. O trabalho visa, nessa medida, auxiliar a perceção mais profunda referente às estratégias de ação impulsionadoras de (contra)discursos positivos, construtivos e inclusivos para prevenir e mitigar o *discurso de ódio*, em linha e fora dela, salientando a sua relevância e a sua essencialidade.

Assim, o primeiro capítulo procura materializar o resultado da atividade crítica e reflexiva, de identificação, exploração e sistematização dos diversos conceitos envolvidos na pesquisa. Apresenta as definições concetuais dos termos de referência deste trabalho, a sua complexidade e a sua evolução, contextualizando-as, em particular, nos espaços digitais e nas plataformas de *media* sociais, promovendo o estado da arte.

O segundo capítulo é dedicado à investigação, analisando-se temática, comparativa e sistematicamente, oito textos-chave centrados em estratégias de combate ao *discurso de ódio*, ao radicalismo e extremismo, em linha e fora dela, e de promoção dos direitos humanos. Nele, são apresentados os resultados, destacam-se as áreas de convergência, examinam-se e sintetizam-se as ideias derivadas numa lista, não exaustiva, de 10 tópicos principais no desenvolvimento de *contranarrativas* e de *narrativas alternativas*. Mais ainda, lançam-se as bases para o processo de ideação que se segue.

O terceiro capítulo incorpora a síntese da problemática em contexto e, sobretudo, o esboço do processo de exploração e de ideação de um modelo alternativo ao combate direto e reativo ao *discurso de ódio*. A partir da adaptação e combinação de elementos pré-existentes, realizam-se três propostas tendo em vista a alteração de paradigma, o seu desenvolvimento e a sua aplicação prática.

## Capítulo I. Revisão de Literatura

### I. 1. Definições conceptuais

#### I. 1.1 Discurso de Ódio

O termo *discurso de ódio* é complexo, controverso e polissêmico, não tendo uma definição nem interpretação unívoca ou universal, coexistindo entendimentos múltiplos e abordagens multidisciplinares do fenómeno, bem como em relação às suas funções e/ou efeitos. A flexibilidade para identificar *discurso de ódio* nas suas várias manifestações cria dúvida e discordância sobre o que o constitui (Article 19, 2015). Apesar da facilidade aparente em reconhecê-lo, da frequência crescente da sua utilização e da sua prevalência, e das várias tentativas para o conceptualizar, sistematizar e fixar, «[...] persistem diferenças colossais, consoante os contextos nacionais e culturais, na definição, âmbito e critérios desta noção» (Silva et al., 2021, p. 12).

O fenómeno do *discurso de ódio* é antigo e foi, desde logo, tal como continua a ser, objeto de estudo, teorização e conceptualização, tanto nos círculos académicos quanto fora deles, nomeadamente nos âmbitos jurídico, político, regulatório e dos direitos humanos, em muito por influência de contextos nacionais e culturais, de fenómenos ou incidentes sociais e de episódios específicos, tentando cada um deles fornecer uma definição integral.

Nos enquadramentos legais, documentos e tratados, o termo *discurso de ódio* é associado a outros conceitos, confrontado com *liberdade de expressão*, mencionado de forma indireta, evitado, substituído por outras nomenclaturas ou formulações mais elaboradas — por exemplo, *discurso ofensivo, incitador* ou *discriminatório; intolerância, discriminação* ou *incitamento à violência* — ou, simplesmente, omitido, demonstrando talvez também «[...] relutância em normalizar ou legitimar o uso da expressão ‘discurso de ódio’, dada a sua condição de termo muito contestado» (Article 19, 2015, p. 17).

Na literatura académica, o conceito e/ou as questões a ele associadas começaram a ser abordados no decénio de 1980. Depois de 2010, registam um

crescimento significativo e constante, e são um *tópico tendência* na atualidade — em muito pela renovada prevalência e visibilidade do *discurso de ódio em linha*, objeto do ponto I. 1.1.1. A produção científica é, cada vez mais, vasta e não circunscrita a uma área específica, mas distribuída por campos de estudo diversos — tais como: direito, sociologia, antropologia, teoria política, história, comunicação e psicologia, entre outros (Anderson & Barnes, 2022; Montero et al., 2022; Paz et al., 2020; Tontodimamma et al., 2021) —, numa mistura eclética de definições de *discurso de ódio* (SELMA, 2019).

A *profusão de definições disjuntivas*, extremamente *amplas* ou *estreitas*, pode justificar um ceticismo geral quanto à possibilidade de uma descrição definitiva de *discurso de ódio* (Anderson & Barnes, 2022; Siegel, 2020). Da mesma forma, a complexidade leva a maioria dos autores a optar por uma definição geral (Paz et al., 2020) e outros a utilizar conceitos alternativos, tais como *discurso extremista* (Hare & Weinstein, 2009; Pohjonen & Udupa, 2017), *discurso perigoso* (Benesch, 2012), *discurso de medo* (Buyse, 2014), *linguagem ofensiva* (Warner & Hirschberg, 2012) e *linguagem abusiva* (Nobata et al., 2016; Niemann et al., 2019).

O *discurso de ódio* é considerado, simultaneamente, como tendo um caráter situacional e evolutivo, emotivo e performativo (Article 19, 2015; SELMA, 2019; Silva et al., 2021), declarado, subtil, moderado, não-emotivo e até brando (Parekh, 2006), espontâneo e organizado (Cervone et al., 2021; Herz & Molnár, 2012), com efeitos (in)diretos, consequências e interpretações variadas (SELMA, 2019; Silva, 2019; Silva et al., 2021). Portanto, *discurso de ódio* é um termo *guarda-chuva*, abrigando um número considerável de significados diversos e contestados e, por isso, é equívoco (Brown, 2017), grandemente incompreendido e diferentemente interpretado (Benesch, 2014b).

No entanto, as definições jurídico-normativas são as dominantes (Udupa et al., 2021). Nesta lógica de *lei-e-ordem*, o *discurso de ódio* é amplamente entendido como um *discurso patológico*, cujos efeitos negativos devem ser diagnosticados, prevenidos e mitigados (Udupa et al., 2021; Udupa & Pohjonen, 2019). Nessa medida, o *discurso de ódio* deve ser combatido, regulado, proibido e sancionado.

A definição avançada pela Comissão Europeia contra o Racismo e a Intolerância (em inglês, *European Commission against Racism and Intolerance — ECRI*)<sup>1</sup> destaca-se como uma das mais citadas e abrangentes, bem como, talvez, uma das mais completas no campo dos instrumentos internacionais que, pela natureza e objeto deste trabalho, aqui relevamos. A Recomendação de Política Geral n.º 15 da ECRI, de 8 de dezembro de 2015, relativa ao combate ao *discurso de ódio*, define-o como sendo:

O uso de um ou mais modos particulares de expressão — nomeadamente, a defesa, a promoção ou a incitação à discriminação, ódio ou desrespeito a uma pessoa ou grupo de pessoas, assim como qualquer forma de assédio, insulto, estereótipo negativo, estigmatização ou ameaça a essa pessoa ou grupo de pessoas, bem como qualquer forma de justificação de todos estes modos de expressão — que sejam baseados numa lista não exaustiva de características pessoais ou estatutos que incluem “raça”, cor, língua, religião ou crença, nacionalidade ou origem étnica ou nacional, bem como ascendência, idade, deficiência, sexo, género, identidade de género e orientação sexual. (ECRI, 2016, p. 16)

Ao contrário de outros instrumentos, a Recomendação n.º 15 sugere uma lista de terminologias que deviam ser ilegais e de todos os grupos vulneráveis que são frequentemente alvo de *discurso de ódio*. Dado que o *discurso do ódio* pode assumir muitas formas — não apenas palavras escritas ou faladas, mas também imagens, sinais, símbolos, vídeos, e mesmo um determinado comportamento, como gestos —, esta Diretriz define «expressão» como qualquer tipo de discurso comunicado e publicado, em qualquer formato, através de vários meios de comunicação, incluindo eletrónicos, assim como a sua distribuição e preservação (ECRI, 2016). Contudo, a Orientação exclui da definição de *discurso de ódio* qualquer *expressão*, tal como sátira, que meramente ofenda, magoe ou angustie outra pessoa, e que se encontra, como tal, protegida pelo Artigo 10.º da Convenção Europeia dos Direitos Humanos (Silva et al., 2021). Isto significa que o que constitui *discurso de ódio* tem de considerar mais do que apenas palavras e expressão. Tem de ter em conta também outros fatores, tais como: o orador,

---

<sup>1</sup> A ECRI é um organismo de monitorização de direitos humanos do Conselho da Europa, especializado em questões relacionadas com o racismo, a discriminação racial, a xenofobia, o antisemitismo e a intolerância, em 46 países da Europa, incluindo Portugal.

o contexto, e se a expressão se enquadra ou não na categoria de «discurso perigoso». Isto é, a linguagem especificamente relacionada com o incitamento à violência em massa ou a expressões que têm uma probabilidade significativa de «[...] estimular ou ampliar a violência de um grupo contra outro, dadas as circunstâncias em que [foram] feitas ou disseminadas» (Benesch, 2012, p. 1).

De forma resumida, *discurso de ódio* consiste em «[...] incitar o ódio ou perpetuar estereótipos sobre [indivíduos ou] grupos específicos com base em características como raça, género ou orientação sexual» (ECRI, 2016, como citado em Weber et al., 2020, p. 2). Por conseguinte, uma das características centrais do *discurso de ódio* é visar minorias ou comunidades política e historicamente marginalizadas, nomeadamente mulheres, pessoas afrodescendentes, ciganas e LGBTI+ (Article 19, 2015; Silva et al., 2021).

O *discurso de ódio* pode ser *plural*, mas nem tudo é *discurso de ódio*, dependendo de muitos elementos e podendo assumir muitas formas. Algumas expressões podem ser meramente avaliativas, ofensivas, vexantes, repugnantes ou inaceitáveis, mas não se configurarem como *discurso de ódio*. Para ser considerado *de ódio*, um discurso ou expressão tem de conter alguns elementos constantes, nomeadamente: *intenção* (magoar ou ofender ou intimidar, com propósito, hostilidade ou desprezo); *incitamento* (instigação à violência ou difusão de ideias preconceituosas); e *efeitos específicos* (dano, discriminação ou violência). Além destas, o contexto em que o discurso tem lugar, o papel e o estatuto de quem o profere, o meio de comunicação no qual ocorre e o tipo de audiência a que chega são igualmente determinantes para a avaliação do discurso como sendo de ódio (Di Girolamo, 2021; SELMA, 2019). Por fim, e ainda que, em certos casos, possa constituir um crime (Nielsen, 2002), *discurso de ódio* é diferente de *crime de ódio*: o *discurso de ódio* pode ser referido como *crime de ódio* quando assume a forma de conduta que é, em si, uma ofensa criminal (ECRI, 2016).

Tal como a sua definição, os efeitos ou danos causados pelo *discurso de ódio* convocam igualmente visões distintas. Ainda assim, pode dizer-se que o *discurso de ódio* «[...] cria, reforça e perpetua a opressão» (Richardson-Self, 2018, como citado em Silva et al., 2021, p. 15), produz uma «[...] atmosfera virulenta de “dupla vitimização”» (Cohen-Almagor, 2011, p. 2), institui «[...] hierarquias sociais baseadas na desigualdade e dominação» (Paz et al., 2020, p. 2), incentiva e pretende justificar a discriminação

(Parekh, 2012), é um «[...] veneno de ação lenta» que mina o «bem comum» e ameaça a «paz social» (Waldron, 2012, p. 4), toda a sociedade (ECRI, 2016), senão mesmo a própria democracia (Davids, 2018). Medo, intimidação, assédio, abuso e discriminação são alimentados pelos emissores, recetores, mensagens, canais e interações envolvidas no *discurso de ódio* (Leets, 2002; Matsuda et al., 1993; Whillock & Slayden, 1995, como citado em Paz et al., 2020). No fundo, o *discurso de ódio* produz danos em e por si mesmo (Nielsen, 2002), fazendo exatamente o que a sua terminologia sugere: «[...] semeia ódio e dano» (Davids, 2018, p. 11) ou, numa das suas piores consequências aparentes, *violência* (Benesch, 2014a), subliminar, psicológica e/ou física.

A complexidade, controvérsia e polissemia do termo *discurso de ódio* e a (in)definição das suas funções e/ou efeitos são agravados na, e pela, *era digital*, da mesma maneira que o são a sua preponderância e evidência.

#### I. 1.1.1. Discurso de Ódio na Era Digital

As manifestações de intolerância e discriminação relativamente a grupos marginalizados não foram criadas na, nem pela, *era digital*. Mas foi precisamente com a internet e em linha que o *discurso de ódio* ganhou renovada complexidade, prevalência e visibilidade, tal como, de alguma forma, se multiplicou, legitimou e/ou normalizou, tornando-se mais difícil de apreender e combater (Castaño-Pulgarín et al., 2021; SELMA, 2019). Por outras palavras, os novos espaços digitais de comunicação e interação, de partilha *livre* de pensamentos, opiniões e emoções — de forma *anónima*,  *mascarada* ou, mesmo, *identificada* — são também terrenos férteis de produção e disseminação de *discurso de ódio*, promovendo a sua rápida propagação, multiplicação, viralização e banalização no dia a dia.

A *era digital*, ou *revolução digital*, ou *era da informação*, teve início por volta de 1980, está em curso e é caracterizada pela rápida e contínua inovação tecnológica, da mudança gradual do analógico para o digital até à (quase) omnipresença da tecnologia e dos dispositivos eletrónicos e digitais na vida quotidiana. A Internet, o seu sistema-chave *World Wide Web*<sup>2</sup>, as Tecnologias de Informação e Comunicação (TIC), os meios,

---

<sup>2</sup> Criada por Tim Berners-Lee, em 1989.

ferramentas e dispositivos de acesso e transmissão — em particular, banda larga, sem fios (*wi-fi*) e móvel; computadores portáteis, *telemóveis inteligentes* (vulgo, *smartphones*) e *tablets* — são elementos-chave da transformação digital, ou digitalização, uma operação não linear, de múltiplas fases e diversos processos.

As redes locais com uma natureza *privada* e geográfica estabelecida, deram lugar ao primeiro exemplo de uma rede comercial e global nos anos 90 e início do século XX: a *Web 1.0*, *tradicional* ou *estática*, como também é conhecida, fundada na *World Wide Web*, na linguagem HTML<sup>3</sup>, no endereço URI ou URL<sup>4</sup>, e no protocolo HTTP<sup>5</sup>. No fundo, uma teia (*net*) de ficheiros de hipertexto que podia ser visualizada por sistemas de navegação (*browsers* [CERN, 2023]). À semelhança dos meios de comunicação de massa, a primeira fase da *Web* era centrada na distribuição e leitura de informação (*read-only-web*), funcionando de *um-para-muitos*. A produção de conteúdos era muito embrionária, as oportunidades de interação eram muito raras e os internautas, ainda espécie de *espetadores/consumidores*, apenas conseguiam consultar os conteúdos de páginas estáticas, receber e descarregar informações, mas não influenciar os conteúdos disponíveis.

O paradigma da utilização da internet muda com a *Web 2.0*<sup>6</sup>. Na sua segunda fase, e atual, a *Web* incorpora atributos da *Web* participativa, colaborativa, interativa e social, permitindo o envolvimento de múltiplas fontes na criação e disseminação de informação, numa comunicação de *muitos-para-muitos*. Nos últimos 15 a 20 anos, a *Web* estruturada deu lugar à *Web* semiestruturada, as páginas agregam a interação entre sítios *Web* e utilizadores, a conectividade social, o conteúdo gerado pelo próprio utilizador e a partilha (*read-write-share*). Os utilizadores são produtores, editores e disseminadores de narrativas mediatizadas, que são vistas por milhões de pessoas em quase todo o mundo, num lapso de tempo cada vez mais curto. O acesso à internet através de dispositivos móveis e portáteis, as plataformas de *media* sociais e as

---

<sup>3</sup> Acrónimo de Linguagem de Marcação de Hipertexto (*HyperText Markup Language*) que define estruturalmente um conteúdo numa página da internet (*Web page*).

<sup>4</sup> Siglas de Identificador Uniforme de Recursos (*Uniform Resource Identifier*), ou Localizador Uniforme de Recursos (*Uniform Resource Locator*), ou seja, um endereço *Web* único utilizado para identificar cada recurso na *Web*.

<sup>5</sup> Abreviatura de Protocolo de Transferências de Hipertexto (*HyperText Transfer Protocol*) que permite a recuperação, transferência e visualização de recursos ligados a partir de toda a *Web*.

<sup>6</sup> Termo criado pelo empresário norte-americano Tim O'Reilly, em 2005.

aplicações são inovações-chave que impulsionaram o crescimento exponencial da *Web 2.0*, da interatividade e utilidade em linha, assim como a criação, estabelecimento e domínio das cinco *gigantes tecnológicas*, conhecidas na Europa como GAFAM — Google (Alphabet), Apple, Facebook (Meta), Amazon, e Microsoft —, e na América como FAANG<sup>7</sup> — Facebook (Meta), Amazon, Apple, Netflix, e Google (Alphabet).

A próxima fase de evolução — isto é, a *Web 3.0*<sup>8</sup> — já está em marcha, desde meados do decénio de 2010, e adivinha-se potencialmente tão disruptiva quanto a anterior. A nova etapa estende as tecnologias da *Web 2.0* e associa-as às tecnologias de dados interligados entre si (*linked data*) e à *Web Semântica (Semantic Web)* para criar uma forma mais sofisticada de entender, interpretar e disponibilizar dados, assim como uma experiência de utilização em linha mais personalizada e intuitiva. Os conceitos centrais de construção são descentralização, abertura e maior utilidade. E entre as suas tecnologias fundadoras contam-se: Internet das Coisas (*Internet of Things [IoT]*), computação em nuvem (*cloud computing*), inteligência artificial (*AI*), incluindo aprendizagem automática (*machine learning*) e *blockchain*, realidade virtual (*VR*) e aumentada (*AR*), robótica e automação, e manufatura aditiva (*additive manufacturing*), como impressão 3D.

A internet e as plataformas de *media* sociais são parte inestimável da vida quotidiana (Cohen-Almagor, 2021). A sua penetração e quase ubiquidade é avassaladora (Kilvington, 2021; O'Regan, 2018). No início de 2024, o número de utilizadores de internet ultrapassou 5,3 mil milhões — 66,2% da população mundial —, destes mais de 5 mil milhões utilizam plataformas de *media* sociais — 62,3% da população mundial. O *Facebook* lidera com mais de 3 mil milhões de utilizadores ativos, segue-se-lhe o *YouTube* com quase 2,5 mil milhões, o *WhatsApp* e o *Instagram* com 2 mil milhões, e o *TikTok* com 1,5 mil milhões (Kemp, 2024a). Cada utilizador passa, em média, 6 horas e 40 minutos por dia na internet, 2 horas e 23 minutos dos quais nas plataformas de *media* sociais (Kemp, 2024a). Em Portugal há 8 milhões e 840 mil de utilizadores de internet — 86,4% da população total — e, destes, 7 milhões e 430 mil estão nas plataformas de

---

<sup>7</sup> Termo popularizado pelo apresentador do programa de televisão *Mad Money*, da CNBC, Jim Cramer, em 2013.

<sup>8</sup> Termo criado pelo engenheiro de *software* e um dos principais protagonistas do desenvolvimento da tecnologia *blockchain*, Gavin Wood, em 2014.

*media* sociais. O *YouTube* é a plataforma com maior número de utilizadores — 7 milhões e 430 mil —, segue-se-lhe o *Facebook* — 5 milhões e 950 mil — o *Instagram* — 5 milhões e 800 mil — e o *LinkedIn* — 4 milhões e 900 mil (Kemp, 2024b). Cada utilizador passa, em média, 7 horas e 30 minutos por dia na internet, 2 horas e 23 minutos dos quais nas plataformas de *media* sociais (Kemp, 2024b).

A rápida popularização das tecnologias digitais e da internet transformaram radicalmente vários aspetos, em particular, a forma como a informação e o conhecimento são comunicados e disseminados (Di Girolamo, 2021; Kilvington, 2021). A digitalização da informação, elemento central e indispensável da *era digital*, numa rede evolutiva e complexa, simultaneamente local e global, genérica e personalizável, revolucionou também as condições sociais de interação e comunicação (Castells, 2007), de participação na vida pública (Miranda et al., 2022) e do comportamento (Kilvington, 2021). A Internet, enquanto sistema tecnológico, é um produto social culturalmente informado, lançando «[...] as bases para a criação de redes autodirigidas de organização, ação coletiva e construção de significado» (Castells, 2001, p. 55). No fundo, a Internet é «[...] um gigante social, legal e moral» (Alkiviadou, 2023).

A comunicação em linha apresenta quatro diferenças-chave da comunicação não em linha: anonimato, invisibilidade, imaginação dissociativa e reação rápida (Kilvington, 2021). Do mesmo modo, a publicação digital de informação e comunicação inclui vários aspetos que a diferenciam das formas tradicionais. De entre os quais se destacam: a escala, amplitude e instantaneidade; a ausência de decisão editorial por uma pessoa que não o autor; e a concentração de um parte muito substancial de todo o discurso na internet num grupo muito pequeno de plataformas ou intermediários<sup>9</sup> (O'Regan, 2018). Isto é, entidades privadas que «[...] têm a capacidade de decidir se e como moldar a expressão em linha» (Citron & Norton, 2011, p. 1456).

As dissemelhanças e os fatores da comunicação em linha e da publicação digital fazem também da *Web 2.0* um «[...] veículo indiscriminado para todas as formas de discurso» (O'Regan 2018, p. 3), incluindo o *discurso de ódio*, «[...] encorajam a

---

<sup>9</sup> Leia-se intermediários da Internet, isto é, «[...] entidades privadas que alojam ou indexam conteúdos em linha» (Citron & Norton, 2011, p. 1438), tais como os motores de busca, as plataformas de *media* sociais e os fornecedores de serviço de Internet (*Internet Service Providers* — ISP).

desinibição que exacerba o ódio em linha» (Kilvington, 2021, p. 261), e funcionam como um «[...] mecanismo multiplicador de forças», que amplia o seu alcance (Daniels, 2008, p. 132).

Todas as inovações e tecnologias têm benefícios e desvantagens, um *lado brilhante* e um *lado obscuro*. A *era digital*, a internet e as plataformas de *media* sociais não são exceção. A *era digital* da informação, interatividade, simplicidade, rapidez e flexibilidade, é paradoxalmente a época da desinformação, da *pós-verdade*, do viés algorítmico e do *discurso de ódio* em linha.

As tecnologias digitais estabeleceram outras alterações fundamentais na equação do *discurso de ódio*:

[...] i) a aceleração dos processos de disseminação dos discursos, ii) o alargamento transfronteiriço das narrativas, ampliando o espaço e o tempo dos ataques, iii) o crescimento potencial das audiências, iv) a diversificação dos atores humanos e não humanos envolvidos no processo, v) a sofisticação das estruturas de produção e consumo, vi) a formação de comunidades virtuais entre propagadores de ódio, vii) o anonimato criado por estruturas tecnológicas, como a navegação VPN, viii) os espaços virtuais pouco regulados, como a *deep web*. (Di Fátima & Miranda, 2022, pp. 14-15)

Assim, o entusiasmo inicial, referente ao potencial de democratização, emancipação e libertação, parece cada vez mais excessivamente otimista (Turner, 2018) e eclipsado por uma explosão de linguagem repleta de ódio e ira violentos à escala global (Udupa et al., 2021), potenciando uma eventual *globalização do ódio* (Perry & Olsson, 2009).

A internet contém o melhor e o pior da humanidade (Cohen-Almagor, 2011), sendo uma «bênção mista», tanto proporcionando e agilizando a educação, a investigação e o debate, quanto potenciando algumas condutas socialmente destrutivas (Citron & Norton, 2011), reafirma crenças e valores, fere e subordina (Di Fátima & Miranda, 2022). A internet tornou mais fácil, rápido e barato difundir o ódio (Kilvington, 2021), já que ocasiona espaços virtuais, nacionais e internacionais, de *discurso de ódio* (Tsesis, 2001, p. 863), tornando-o uma corrente dominante (Citron & Norton, 2011) e

concorrendo para a sua *naturalização* (Casquilho-Martins et al., 2022; Kilvington, 2021; SELMA, 2019; Silva et al., 2021). Todavia, o problema não é a internet, mas a utilização negativa que dela se pode fazer (Cohen-Almagor, 2011), ainda que, mais ou menos, influenciada ou condicionada por legisladores e intermediários.

A presença do *discurso de ódio* em linha é tão antiga quanto a própria internet (Siegel, 2020). Os produtores de ódio (*haters*) identificaram, igual e rapidamente, as potencialidades e as fragilidades da comunicação em linha, utilizando-as para seu proveito (Perry & Olsson, 2009). As preocupações relativas à mobilização de grupos de ódio, em particular, racistas, através da criação de sítios *Web* destinados a difundir mensagens de ódio, surgiram em meados de 1990 (Lee-Won et al., 2020).

A história do *discurso de ódio* na *World Wide Web* terá começado a 11 de janeiro de 1995, quando o defensor da supremacia da raça branca norte-americano, Don Black, criou o Stormfront, uma espécie de «supermercado de ódio» em linha que deu voz a muitas formas de antissemitismo e racismo (Cohen-Almagor, 2011, p. 4). Entretanto, os grupos de ódio garantiram uma presença forte e bem-mobilizada na *Web*, utilizando estratégias persuasivas de propagação e legitimação de mensagens contra minorias sociais e étnicas (Lee-Won et al., 2020), em todas as suas camadas, desde a mais *obscura* (*Dark Web*) até à mais à *superfície* (*Surface Web*). Para tal, recorrem tanto a sítios *Web* explicitamente de ódio (*overt Web sites*) como aquele, quanto a formas mais sofisticadas e de alcance mais abrangente, nomeadamente os «sítios *Web* camuflados» (*cloaked Web sites*) — ou seja, que disfarçam propositadamente o seu propósito odioso sob informação imparcial ou fatural (Daniels, 2008) —, e às plataformas de *media* sociais. O [www.martinlutherking.org](http://www.martinlutherking.org) será, talvez, o exemplo mais infame de um «sítio *Web* camuflado» (SELMA, 2019).

As oportunidades de disseminação de *discurso de ódio* proliferam, como nunca, com a internet (Cohen-Almagor, 2011; Daniels, 2008) e, em particular, com e nas plataformas de *media* sociais. Nelas, o ódio é «[...] ubíquo, interativo e multimédia» (Di Fátima et al., 2023, p. 13) e, por isso, podem constituir-se como «[...] palcos virtuais de ódio» (Kilvington, 2021, p. 256), concedendo ao *discurso de ódio* uma crescente visibilidade (Siegel, 2020).

As plataformas de *media* sociais têm por base sistemas *Web 2.0* orientados para os utilizadores, uma natureza participativa de criação-partilha de conteúdo gerado pelos próprios (*user-generated content*), permitem o envolvimento de múltiplas fontes na geração e disseminação de mensagens, levando a que as expressões de ódio mediadas digitalmente tomassem uma nova direção (Lee-Won et al., 2020), e tornando o *discurso de ódio* num problema sério e crescente que afeta as plataformas dominantes e marginais (Lupu et al., 2023).

As plataformas de *media* sociais, em particular, as dominantes, utilizam algoritmos que geram bolhas de filtro (*filter bubbles*) e favorecem a criação de câmaras de reverberação (*eco chambers*), propagando e amplificando o discurso, incluindo o de ódio, inúmeras vezes através de partilhas, comentários e reações dos utilizadores (Daniels, 2008; Di Fátima & Miranda, 2022; Kilvington, 2021; Matamoros-Fernández, 2017, 2018; Pukallus & Arthur, 2024), e aumentando a probabilidade de o ódio se tornar uma parte aceitável do discurso em linha (Citron & Norton, 2011). Ao apelarem conscientemente para *inflamar as emoções dos seguidores* através de quaisquer meios necessários, infamar o exogrupo, infligir danos permanentes e irremediáveis e, em última instância, para a conquista, as plataformas de *media* sociais tornam o ódio numa *arma* e o *discurso de ódio* num *estratagema, artifício* ou *truque de guerra na era digital* (Kirk & Schill, 2021), assemelhando-se a um *campo de batalha* de comunicação digital (Pukallus & Artur, 2024).

Porém, a responsabilidade do aumento do *discurso de ódio* em linha não é exclusiva das plataformas de *media* sociais, devendo apreciar-se outros fatores que estão além delas (Kilvington, 2021). De entre esses elementos, refiram-se: condições inatas e situacionais de utilizadores, como autores e destinatários, que afetam o comportamento de ambos, por exemplo, os graus de influência e empatia, de e com perpetradores (Brown, 2018); os estados de espírito individuais e coletivos (Farrington et al., 2015); e os contextos envolventes de interação, exposição e reprodução (Cheng et al., 2017).

A tendência de agravamento do *discurso de ódio* e a consciencialização dos danos que pode causar, tanto em linha quanto fora dela, *pressionaram* e *impeliram* Facebook, Instagram, Twitter — agora X —, YouTube e, até mesmo, TikTok, a *aliar-se à causa*.

Portanto, estas plataformas encetaram — ou foram levadas a fazê-lo — um processo, ainda inacabado ou em mutação constante, de estabelecimento, operacionalização e robustecimento das suas políticas, estratégias ou intervenções, recursos e sistemas de moderação para procurar *curvar* o *discurso de ódio*. Primeiro, de forma reativa e recorrendo a técnicas *humanas* (*comercial content moderation*). Depois, mais (pro)ativa e conjuntamente, com especialistas, entidades públicas, outros públicos interessados e com os próprios utilizadores, disponibilizando-lhes ferramentas de denúncia, assim como com instrumentos mais sofisticados *não humanas* — algoritmos e *design* de interface/padrão (*interface/default design*).

*Facebook, Instagram, Twitter/X, YouTube e TikTok* avançaram com as suas próprias definições de *discurso de ódio*, criaram novas políticas dedicadas e integraram-nas tanto nos seus *Termos e Serviços* ou *Padrões da Comunidade*<sup>10</sup>, quanto em *relatórios de transparência* ou de *aplicação das regras da comunidade*<sup>11</sup>. Com isto, determinaram de forma, mais ou menos, clara, a proibição do *discurso de ódio* e as ações aplicáveis à sua utilização. Estas práticas de moderação — ou de sanção/penalização — passaram a incluir: advertências; supressão de algumas funcionalidades; eliminação de conteúdos; desativação de contas, temporária ou por períodos variáveis, e remoção definitiva de contas — isto é, *desplataformizar* (*deplatforming* [Batista, 2022]). As GAFAM, ou FAANG, «[...] tornaram-se legisladores, juízes, júris e executores das suas próprias políticas de discurso de ódio na Internet» (Mirrlees, 2021, p. 93).

No entanto, o entendimento das plataformas sobre *discurso de ódio* é divergente ou heterogéneo — o grau de detalhe dos conceitos, termos e exemplos varia — e os métodos utilizados não são expostos com minúcia, sugerindo uma transparência parcial, sempre controlada pelas próprias, e a ausência de um trabalho cooperativo no *combate* ao *discurso de ódio* (Batista, 2022).

A intrincada teia de regras encobertas e obscuridade de procedimentos de moderação resulta igualmente do facto do conteúdo gerado pelos utilizadores (*user-*

---

<sup>10</sup> Vide «[Discurso de Incentivo ao Ódio](#)» da Meta, «[Conduta de propagação de ódio](#)» do X, «[Política de Incitamento ao ódio](#)» do YouTube e «[Combater o discurso e o comportamento de ódio](#)» do TikTok.

<sup>11</sup> Vide «[Community Standards Enforcement Report](#)» do Facebook e do Instagram, «[Aplicação das Regras](#)» do X, «[Aplicação das regras da comunidade do YouTube](#)», e «[Relatório de aplicação das Diretrizes da Comunidade](#)» do TikTok.

*generated content*), incluindo *discurso de ódio*, ser um bem gerador de rendimento, ou mesmo a maior *moeda de troca*, e uma grande responsabilidade corporativa das plataformas de *media* sociais (Roberts, 2018). Nestes termos, a moderação de conteúdo é uma prática comercial, o seu critério-chave é a monetização, a par com a maximização de rendabilidade, e o seu papel «[...] é fundamentalmente uma questão de proteção da marca da empresa» (Roberts, 2018, secção *Commercial content moderation and the logic of opacity*, par. 3).

O número de utilizadores, a sua participação e, por conseguinte, a monetização de anúncios podem ser afetadas pela moderação de conteúdos tóxicos ou perigosos, incluindo de *discurso de ódio* (Madio & Quinn, 2023). Assim, as motivações de moderação estão relacionadas tanto com a pressão regulatória — da qual são exemplos, o «Código de Conduta da União Europeia para a luta contra os discursos ilegais de incitação ao ódio»<sup>12</sup> e o «Regulamento dos Serviços Digitais (RSD)»<sup>13</sup> da Comissão Europeia, e a lei *NetzDG (Network Enforcement ACT)* aprovada na Alemanha, em 2017 — , quanto com a pressão de anunciantes (Madio & Quinn, 2023). A prevalência do *discurso de ódio* nas plataformas e a incapacidade de criarem espaços seguros motivou um êxodo de anunciantes do YouTube em 2018, conhecido como «Adpocalypse». Em 2020, mais de mil empresas, entre as quais a Unilever, Levi's, Ben and Jerry e Coca-Cola, suspenderam o investimento publicitário no *Facebook*, no seguimento da campanha *#StopHateForProfit*. Dois anos mais tarde, o GroupM, empresa líder mundial no investimento em meios, classificou o *Twitter* como uma «plataforma de alto risco», depois de Elon Musk assumir o controlo e flexibilizar as políticas de moderação de conteúdos, levando à suspensão de compra de anúncios ou abandono de muitas marcas de luxo, como a Balenciaga (Madio & Quinn, 2023). Mais recentemente, o *X (ex-Twitter)* conheceu novo boicote de grandes anunciantes, incluindo a Walt Disney Co. e a Apple Inc., depois de Elon Musk endossar uma publicação antissemita, no dia 16 de novembro de 2023.

---

<sup>12</sup> Criado em 2016 e acordado, primeiro, com *Facebook*, *Microsoft*, *Twitter* — abandonou o Código em maio de 2023 — e *YouTube*, e depois também com *Instagram*, *Snapchat*, *Dailymotion*, em 2018, *Jeuxvideo.com*, em 2019, *TikTok*, em 2020, *LinkedIn*, em 2021, *Rakuten Viber* e *Twitch*, em 2022.

<sup>13</sup> O «Digital Services Act (DSA)» foi publicado no Jornal Oficial da União Europeia a 27 de outubro de 2022 e entrou em vigor a 17 de fevereiro de 2024.

As dúvidas e críticas estendem-se às duas formas dominantes de moderação de conteúdos nas plataformas: (1) automatizada e (2) humana. A moderação humana acontece na *sombra*, é realizada por prestadores de serviços subcontratados (*outsourcing*) por empresas de trabalho temporário, numa rede transfronteiriça, com baixas remunerações, estando ligada a precariedade e exploração laboral, a problemas de saúde mental, designadamente transtorno de stress pós-traumático, e a progressiva diminuição de empatia por exposição recorrente a conteúdos brutais, sobretudo gráficos, como de *discurso de ódio* (Chen, 2014; Mirrlees, 2021; Roberts, 2018, 2019; Sinnreich, 2020). A moderação automatizada, delegada em algoritmos, potencialmente opacos, propriedade de empresas privadas e não imputáveis de responsabilidade, é também obscura, podendo aumentar o número de falsos positivos e negativos, e corresponder virtualmente à delegação de todas as funções legislativas, judiciais e executivas numa única empresa (Gillespie et al., 2020). Assim, as normas e instituições democráticas, a soberania e a autodeterminação são implicitamente ameaçadas (Sinnreich, 2018), abrindo-se espaço para a *autocracia*, *corporatocracia* e *monocultura* (Gillespie et al., 2020).

Ao influenciarem o acesso a plataformas que são cada vez mais centrais para trabalhar, viver e prosperar em conjunto, ambos os tipos de moderação servem como um ato de despolitização (Roberts, 2018) e têm potencial para impactar o tecido social das sociedades (Gillespie et al., 2020). Ao aplicarem um processo de moderação por níveis, em que o derradeiro é *desplataformizar* (*deplatforming*), as plataformas de *media* sociais poderão, de algum modo, estar a mudar o que é considerado aceitável (*social*) e inaceitável (*antisocial*) e, assim, «[...] a consolidar o seu poder enquanto *designers*, operadores e governantes do ecossistema digital» (Van Dijck et al., 2021, p. 1), e talvez das próprias sociedades.

Além disso, e apesar de muitos serem *eliminados*, vários milhões de conteúdos de ódio continuam a ser produzidos e disseminados diariamente nas plataformas de *media* sociais (Lupu et al., 2023). Da mesma forma, a remoção depende das plataformas, é muito variável e há conteúdos que são repostos. Entre janeiro e março de 2024, *Facebook* e *Instagram* «tomaram medidas» em relação a 16 milhões de conteúdos de *discurso de ódio*, dos quais 293 mil acabaram por ser restaurados e 96,45% por diligência

própria, estimando a prevalência de «violações de discurso de ódio» entre 0,02% e 0,03% (Meta, 2024). No mesmo período, o *YouTube* removeu mais de 8 milhões e 295 mil vídeos, a quase totalidade dos quais por «detecção automática» e 1,9% por «incitação ao ódio ou abusivo», estimando a «taxa de visualização de conteúdo em violação» entre 0,12% e 0,13% — 41 625 dos vídeos removidos foram repostos (Google, 2024). Por sua vez, o *TikTok* eliminou quase 167 milhões de vídeos, dos quais 8,8% por «discurso e comportamento de ódio» e 13,2% de «organizações ou indivíduos violentos ou odiosos», tomou a iniciativa em 97,7% das ocorrências e repôs mais de 6 milhões de vídeos (*TikTok*, 2024). De acordo com o último relatório publicado, em 2021, o *Twitter* removeu 11 milhões de conteúdos, destes 2 milhões e 900 mil por «conduta de propagação de ódio» (*Twitter*, 2023).

Já os dados registados pela Comissão Europeia, na sétima, e última, avaliação divulgada sobre o «Código de Conduta da União Europeia para a luta contra os discursos ilegais de incitação ao ódio» nas plataformas de *media* sociais dominantes, em coordenação com autoridades nacionais, demonstra que, em 2022, 64,4% das notificações foram revistas num prazo de 24 horas — uma diminuição significativa face a 2021 (81%) e a 2020 (90,4%) — e que 63,6% do conteúdo foi removido — um pouco acima de 2021 (62,5%), mas aquém de 2020 (71%) —, tendo a maioria das plataformas, à exceção do *YouTube*, removido menos conteúdo de *discurso de ódio* do que em 2021. Ainda assim, a qualidade da resposta às notificações dos utilizadores melhorou — mais 6,1% por comparação com o ano anterior (Reynders, 2022).

A visibilidade crescente do *discurso de ódio* na internet e, particularmente, nas plataformas de *media* sociais dominantes colocou-o no centro das agendas académicas, legais, políticas e mediáticas. Uma centralidade que poderá ter enviesado a perceção da prevalência ou omnipresença do fenómeno em linha, contrariando as evidências empíricas que sugerem que o *discurso de ódio* é, na realidade, bastante raro no conjunto (Alkividou, 2023; Siegel, 2020). Do mesmo modo, esta crescente atenção originou várias definições de *discurso de ódio* em linha, afastando a resolução eventual do debate em torno da sua conceptualização (Siegel, 2020), e tornou persistentes as interrogações sobre o que o distingue ou torna *especial* (Udupa, 2021), complexificando a equação.

O *discurso de ódio* em linha, tal como o não conectado, convoca várias conceptualizações. Todavia, numa visão abrangente, as diferentes definições académicas, legais e das próprias plataformas de *media* sociais de *discurso de ódio* em linha têm traços-comuns. Estes são: o discurso é público ou dirigido a um membro do grupo, incita a más ações além do próprio discurso em si, causa dano e não tem uma finalidade redentora; o conteúdo da mensagem exprime ódio; o orador pretende prejudicar ou provocar uma má ação; o alvo é um grupo, ou um indivíduo membro de um grupo; e o contexto torna possível uma resposta violenta. Apesar de não formarem uma definição única, estas linhas-comuns podem ser úteis para identificar um *discurso* como sendo *de ódio* (Sellars, 2016).

O *discurso de ódio* proferido em linha ou fora dela é, relativamente, semelhante, mas o primeiro tem um conjunto de características particulares, inerentes ao próprio *medium* digital e ao contexto, que o tornam substancialmente diferente na sua natureza (Pizzi, 2022; SELMA, 2019). A permanência, itinerância, invisibilidade, anonimato — seja ele, percebido, sob pseudónimo ou total —, transnacionalidade, facilidade de acesso, dimensão da audiência, comunidade, e instantaneidade, que caracterizam a própria internet, são, igual e potencialmente, distintivas do *discurso de ódio* em linha (Brown, 2018; Gagliardone et al., 2015; Kilvington, 2021; Pizzi, 2022; SELMA, 2019).

Além disso, o *discurso de ódio* em linha é mutável, desenvolve linguagens e técnicas retóricas próprias, (in)diretas, explícitas e implícitas, e envolve um número considerável de atores, por vezes, com funções (in)diferenciadas de autor, difusor, vítima (Pizzi, 2022). Nos espaços digitais, as formas mais comuns de *discurso de ódio* — religioso, racista e étnico, político, de género e desencadeado por eventos terroristas — expressam-se através de léxicos e figuras de retórica negativos e de subjugação, observação vicária, humor, estereótipos negativos, especulação, comparação, comentários degradantes, calúnia, difamação, revolta, sarcasmo, ameaça, assédio sexual, humilhação, hostilidade, obscenidade, insultos e desumanização (Castaño-Pulgarín et al., 2021).

Por outro lado, o *discurso de ódio* em linha visa, principalmente, e também em Portugal, afrodescendentes, migrantes, comunidades muçulmanas, judaicas, ciganas e LGBTI+, bem como, em particular, mulheres (Almeida et al., 2023; Carvalho et al., 2024;

Casquilho-Martins et al., 2022; Cohen-Almagor, 2011; Costa, 2021; FRA, 2023; Miranda et al., 2022; Silva et al., 2021; Sousa de Almeida, 2022). Afora estes, e em particular, nas plataformas de *media* sociais predomina uma estratégia direcionada de ataques de ódio ou *rusgas* (*raids*) coordenados por grupos bem organizados contra utilizadores com um número considerável de seguidores (*followers*) ou figuras proeminentes, tais como bloguistas (*bloggers*), celebridades e jornalistas. Isto poderá explicar a presença importante do fenómeno nos meios de comunicação de massas e o desvio na perceção da sua prevalência (Siegel, 2020).

A ausência de características únicas e identificadoras, as subtilezas de linguagens e as diversas definições tornam o *discurso de ódio* difícil de identificar, apreender e combater (Carvalho et al., 2024; Castaño-Pulgarín et al., 2021; Pukallus & Arthur, 2024; SELMA, 2019), mas não o tornam inofensivo. Ainda que a prevalência em linha possa ser *menor* e menos *galopante* do que se intui (Alkiviadou, 2023; Siegel, 2020), o *discurso de ódio* em linha pode afetar o comportamento dentro e fora dela, causar efeitos diretos e indiretos, a curto e longo prazo, e ter consequências reais, tanto nos indivíduos quanto nas sociedades (Citron & Norton, 2011; Daniels, 2008; Kilvington, 2021; SELMA, 2019; Siegel, 2020).

O *discurso de ódio* em linha pode ter um efeito significativo sobre o bem-estar individual e infligir danos psicológicos e emocionas — tais como: medo, tensão, aumento da ansiedade, distúrbios de sono, baixa autoestima, sentimentos de solidão e isolamento, de medo e insegurança, de inferioridade e depressão —, podendo gerar despersonalização e condicionar as relações intergrupais. Adicionalmente, o *discurso de ódio* em linha tem potencial para ocasionar um custo social mais vasto. Pode reforçar atitudes e comportamentos de ódio, produzir terrenos férteis de polarização e radicalismo, desvirtuar a perceção e compreensão da realidade, normalizar a discriminação, a intolerância, o estigma e o ódio, criar um clima mais amplo de medo e hostilidade, minando a igualdade, a dignidade e a cidadania, logo, a coesão social. Na sua forma mais agressiva, o *discurso de ódio* em linha pode incitar à violência física, a crimes de ódio e, no limite, matar (Citron & Norton, 2011; Foxman & Wolf, 2013; Kilvington, 2021; SELMA, 2019; Siegel, 2020; Vasist et al., 2024).

A violência de monges budistas *nacionalistas* contra a comunidade muçulmana Rohingya, em Myanmar; a emergência do grupo da direita radical “Hindus da Internet” (*Internet Hindus*), na Índia; os protestos de rua do Pegida, movimento alemão e xenófobo contra a islamização do ocidente, por toda a Europa; as investidas contra migrantes africanos acusados de vender *comida falsa* nas suas lojas de conveniência informais (*spaza shops*), na África do Sul; o ataque, transmitido em direto, a duas mesquitas em Christchurch, na Nova Zelândia; o assalto ao Capitólio dos Estados Unidos da América; e, mais recentemente, a invasão do Congresso Nacional, Supremo Tribunal e Palácio da Alvorada em Brasília, capital federal do Brasil, exemplificam a manifestação e impacto reais do *discurso de ódio* em linha (Kirk & Schill, 2021; Lupu et al., 2023; Udupa et al., 2021). A estes pode acrescentar-se a ascensão dos movimentos da direita radical e *neonacionalistas* na América do Norte e na Europa, incluindo em Portugal — o partido Chega, de André Ventura —, acionada e fundada na utilização das plataformas digitais como *arma* (Matamoros-Fernández & Farkas, 2021) e num novo léxico de discurso de exclusão e de beligerância em linha, através de uma retórica inflamada de piadas *racializadas*, intimidação e *mentiras cheias de factos* (*fact-filled untruths* [Udupa et al., 2021]).

A internet e, particularmente, as plataformas de *media* sociais podem não ser *cidadelas* do *discurso de ódio* (Siegel, 2020). Os utilizadores de diferentes países podem perceber o *discurso de ódio* de maneira diferente (Tontodimama et al., 2021), ter dificuldade em distinguir o que é *profano* ou *odioso* em linha (Weber et al., 2020) ou não reconhecerem algumas das suas formas (Schmid et al., 2024). Porém, a toxicidade do ódio nos espaços digitais está a aumentar (EOOH, 2023), um número crescente de pessoas narra já ter sido exposto a *discurso de ódio* e já se ter sentido atacado ou ameaçado nos espaços digitais, nomeadamente nas plataformas de *media* sociais (Alkiviadou, 2023; Castaño-Pulgarín et al., 2021; Siegel, 2020), e os utilizadores-alvo podem ser especialmente afetados por ele (Schmid et al., 2024).

A exacerbação do *discurso de ódio* e dos prejuízos que pode causar, quer em linha quer fora dela, transformou-se numa preocupação crescente, pública e política, particularmente no último decénio, mobilizando autoridades públicas, entidades da sociedade civil e pessoas singulares para o desenvolvimento de estratégias de ação

contra o *discurso de ódio*, nomeadamente, sob a forma de *contranarrativas* e *narrativas alternativas*.

## I. 1.2. Narrativa

A *narrativa* «[...] está aí, como a vida» (Barthes, 1977, p. 79), ou seja, a *narrativa*, tal como a vida que é a sua matéria, está sempre presente, porque é uma imagem desta última. A sua ubiquidade e expressões múltiplas justificam a relevância do conceito teórico de *narrativa* para todas as disciplinas implicadas na experiência humana. Do mesmo modo que *discurso de ódio*, a *narrativa* é um conceito de carácter ambíguo, complexo e polissémico, que admite uma grande variedade de abordagens, conceções e enunciações, não existindo uma definição única universalmente aceite (Latour et al., 2017; Ryan, 2007, 2017; Santana, 2017; Schlegel, 2021). A par de *discurso de ódio*, a *narrativa* é um termo vulgarizado, ocorrendo como sinónimo de *história*, *fábula* e *discurso* (Babo, 2017; Glazzard, 2017; Ryan 2007, 2017). Logo, a pergunta «Que é uma narrativa?» inspirou, sugere e suscitará respostas diversas e, mais ou menos, opostas, em função do posicionamento adotado.

A *narrativa* vem a ser estudada há mais de dois milénios, tendo sido objeto do primeiro escrito conhecido de estudos literários, a «Poética» de Aristóteles (à volta de 335 a.C.). A distinção entre duas das componentes essenciais da narrativa, «*mythos*» — «[...] a ordenação de acontecimentos numa sequência» — e «*ethos*» — «[...] a utilização de personagens para vivenciar os acontecimentos e comentá-los» —, foi realizada por Aristóteles (Glazzard, 2017, p. 7). Apesar da ampliação e do aperfeiçoamento das ideias de Aristóteles por uma sucessão de teóricos ao longo dos séculos, a *narrativa* só foi sujeita a uma análise empírica e teórica aprofundada, no início do século XX, com o aparecimento de um novo movimento de teoria e crítica literária: o *formalismo russo* (Glazzard, 2017). Já na segunda metade do século XX, a teoria literária passou a ser dominada pelo *estruturalismo*. A disciplina de estudos literários que emergiu e se confirmou, primeiro, em França e, depois, na Europa e na América, deu origem a um método baseado na linguística para analisar a *narrativa* «[...] como uma estrutura análoga a uma língua, com normas, convenções e padrões» (Glazzard, 2017, p. 8). As abordagens *formalista* e *estruturalista* começaram a ser codificadas mais tarde, ainda

no século XX, na *narratologia*, «[...] uma subdisciplina que desenvolveu um conjunto de termos e conceitos precisos para compreender as estruturas narrativas, tanto a nível temático quanto técnico» (Glazzard, 2017, p. 8).

Desde o decénio de 1980, a *narrativa* surgiu no contexto das *ciências humanas* e, em pouco mais de dez anos, transformou-se em matéria de interesse de um vasto número de novas investigações, não apenas como um novo objeto de pesquisa, mas enquanto «[...] uma nova abordagem teórica, de um novo género de filosofia da ciência» (Brockmeier & Harré, 2003, p. 525). Não obstante a teorização literária e linguística, o uso do termo *narrativa* trivializou-se, tornou-se exagerado e, por isso, o «[...] seu potencial conceitual e analítico tende, algumas vezes, a tornar-se pouco claro» (Brockmeier & Harré, 2003, p. 526).

Dito de outra forma, a transição para as *humanidades* tornou a *narrativa* ubíqua, passando a encontrar-se em toda a parte, da política à medicina, dos estudos científicos ao direito, sem esquecer a ciência cognitiva, o discurso cultural, a indústria de *software* e os *media* digitais<sup>14</sup>, debilitou substancialmente o seu significado e aumentou o risco de dissolver-se em *crença, valor, experiência, interpretação* ou, simplesmente, *conteúdo* (Ryan, 2002, 2005, 2007). A utilização livre da terminologia originou que «[...] poucas palavras tiveram tanto uso e sofreram tanto abuso quanto *narrativa* e o seu sinónimo parcial, *história*» (Ryan, 2007, p. 22), assim como, talvez também, e noutra âmbito, *discurso de ódio*.

Todavia, existem, ou persistem, duas abordagens dominantes: (1) a *narrativa* como ação discursiva (Narratologia) e (2) a *narrativa* como estrutura ou esquema cognitivo de perceção do mundo (Estudos Culturais [Bamberg, 2007a; Ryan, 2005, 2017]). A *narratologia* reclama a *narrativa* como seu território exclusivo e restringe-a à sua dimensão técnica — um *texto* e uma *história* —, estando demasiado focada na definição e descrição conceptual, secundarizando a questão crucial do papel da *narrativa* na vida social e mental (Ryan, 2017). Por oposição, os *Estudos Culturais* separam as *histórias* dos *textos*, permitindo que representações mentais, individuais ou coletivas, que não dependem de qualquer *texto* específico, possam ser consideradas

---

<sup>14</sup> Isto é, os *media* «[...] que precisam do computador para a sua produção e visualização» (Page & Thomas, 2011, p. 1).

narrativas, estendendo demasiado o conceito e testando a sua utilidade (Ryan, 2017). Ainda que a primeira peque por *defeito*, devendo *alargar-se*, e a segunda por *excesso*, precisando de estreitá-lo, ambas devem ser complementares na resposta à pergunta «Que é uma narrativa?» (Ryan, 2017).

A existência de classificações à primeira vista bem determinadas não invalida a dificuldade em definir com precisão *narrativa*, existindo, pelo menos, quatro razões para tal (Brockmeier & Harré, 2003). Em primeiro lugar, as formas e os estilos da *narrativa* apresentam grande diversidade, incluindo formas específicas, híbridas e mistas, géneros e subgéneros, muito além da literatura e da cinematografia tradicionais, parecendo inumeráveis<sup>15</sup>. Além disso, elementos ou estruturas da *narrativa* (enredo, espaço, tempo e ação) permeiam grande parte «[...] dos outros tipos de discurso, tais como textos científicos, jurídicos, históricos ou religiosos» (Brockmeier & Harré, 2003, p. 528). A estas duas somam-se a questão da definição das vozes autorais — a *narrativa* não é apenas uma invenção individual, nem uma descrição objetiva de acontecimentos —, e a onnipresença da *narrativa*, isto é, a sua existência natural na imaginação, no pensamento e na ação (Brockmeier & Harré, 2003; Ryan, 2017).

Em sentido amplo, a *narrativa* é um *texto* — enquanto o uso deliberado e estruturado de signos (linguagem, imagem, talvez música e as suas várias combinações) — que transmite uma *história* — uma *representação mental* evocada na mente do recetor, constituída por personagens, cenários e acontecimentos (Ryan, 2017). Enquanto *representação mental*, a *história* pode ser separada dos signos e recodificada noutros signos, como o demonstram os fenómenos de recontar, traduzir e adaptar. Embora uma *narrativa* exija um *texto* feito de signos materiais (um *discurso*) e um significado específico (uma *história*), a *narratividade* do *texto* é investida na *história*, porque certos tipos de *textos*, tais como os filosóficos, jurídicos ou científicos, não visam contar *histórias*, mas apresentar outras formas de *discurso* (Ryan, 2017). Nestes termos, a *narrativa* é «[...] um texto que traz à lembrança uma história que, inversamente, é uma

---

<sup>15</sup> A *narrativa* é apresentada na forma de, ou como, monólogos ou diálogos, histórias verídicas ou literárias, reais ou fictícias, textos orais ou escritos, incluindo conversação, argumentação e percepção, anedotas e boatos, mitos e contos de fadas, notícias e relatos históricos, casos jurídicos e explicações científicas, teatro e bailado, música e diversas artes visuais, texto mediado eletronicamente e jogo de computador (Brockmeier & Harré, 2003; Ryan, 2017).

representação mental criada enquanto resposta direta às pistas dadas pelo texto» (Ryan, 2017, p. 518).

A *narrativa*, enquanto *representação mental*, consiste num *mundo* (cenário) povoado por indivíduos ou agentes inteligentes (personagens), que participam, ou estão envolvidos, em ações e acontecimentos (eventos, enredo), através dos quais se transformam ou alteram o estado do mundo (dimensão temporal [Ryan, 2002, 2005, 2017]). Assim, a *narrativa* tem princípio, meio e fim, uma condição (re)conhecida como o «arco narrativo» (Ryan, 2017). A *representação mental (história)* pode ser sugerida em vários graus e o número e importância das transformações que afetam o mundo da *história* são uma questão de juízo subjetivo. Logo, o conjunto de todas as *narrativas* é um *conjunto indefinido (fuzzy set)* e a *narratividade* uma *propriedade escalar (scalar property)* que permite vários níveis de concretização e, portanto, a existência de formas marginais<sup>16</sup> e prototípicas<sup>17</sup> de *narrativa* (Ryan, 2017).

O grau de *narratividade* de um *texto*, ou seja, a medida em que a *representação mental* evocada por um *texto* cumpre a definição de *história*, pode ser determinado através da aferição de oito condições, organizadas em três dimensões semânticas e uma dimensão formal e pragmática (Ryan, 2007). Enquanto o número de requisitos cumpridos determina o grau de *narratividade* de um *texto*, a proeminência das quatro dimensões precisa a tipologia semântica dos textos narrativos, oferecendo um conjunto de ferramentas para responder à pergunta «Este *texto* é uma *narrativa*?» ou «Qual o grau de *narratividade* deste *texto*?». As oito condições e respetivas dimensões são:

#### **Dimensão espacial**

(1) A narrativa deve ser sobre um mundo povoado por seres individuais.

#### **Dimensão temporal**

(2) Este mundo deve estar situado no tempo e sofrer transformações significativas.

(3) As transformações devem ser causadas por eventos físicos não habituais.

---

<sup>16</sup> Por exemplo, crónicas históricas, anais e diários, receitas, poesia lírica e sonhos (Ryan, 2017).

<sup>17</sup> As que são (re)conhecidas por todas as pessoas como histórias, nomeadamente contos de fadas ou lendas, tragédia, *suspense* ou mistério, narrativas naturais e ficcionais — romance (Ryan, 2017).

### **Dimensão mental**

(4) Alguns dos participantes nos eventos devem ser agentes inteligentes que têm uma vida mental e reagem emocionalmente aos estados do mundo.

(5) Alguns dos acontecimentos têm de ser ações intencionais desses agentes.

### **Dimensão formal e pragmática**

(6) A sequência de acontecimentos tem de formar uma cadeia causal unificada e conduzir ao desfecho.

(7) A ocorrência de pelo menos alguns dos acontecimentos tem de ser afirmada como um facto para o mundo da história.

(8) A história tem de comunicar algo significativo para a audiência. (Ryan, 2007, p. 29) [Negritos do original]

Ou, por outras palavras, as *histórias*:

- são representações da vida, não a própria vida (transgredida pelos sonhos).
- são acontecimentos únicos e não repetíveis *ad eternum* (transgredida pelas receitas).
- devem envolver personagens individualizadas (transgredida pela poesia lírica e receitas).
- não são registos automáticos de tudo o que acontece num período de tempo arbitrariamente determinado, concentram-se em acontecimentos que provocam alterações significativas ao estado do mundo em questão, envolvendo seleção e organização de elementos, processo designado como *criar enredo* (*emplotting* — transgredida por diários, anais e crónicas).
- são contadas de uma posição retrospectiva, o que fornece uma visão abrangente dos acontecimentos e das suas consequências (transgredida pelos diários). (Ryan, 2017, p. 519)

Posto isto, a *narratividade*, como propriedade nuclear da *narrativa*, vai além de *enredar personagens*, *refletir experiência* ou *criar interesse*, sendo «[...] a capacidade de convocar à imaginação um mundo, os indivíduos que o habitam e os acontecimentos que o fazem evoluir e que têm importância afetiva para os seus habitantes» (Ryan, 2017,

p. 525). Do mesmo modo, «[...] quanto mais vivamente um mundo narrativo se imprime na imaginação do leitor, do ouvinte ou do espectador, e quanto mais todas as suas partes criam um todo coerente, mais elevada é a narratividade do texto que o exhibe» (Ryan, 2017, p. 526). Quanto maior o grau de narratividade da *narrativa*, mais apelativa e persuasiva é considerada (Schlegel, 2021).

Em suma, a *narrativa* é um *texto* que transmite uma *história*; o *texto* é o uso deliberado e estruturado de signos; e a *história* é uma *representação mental* em resposta às pistas deixadas pelo *texto* (Ryan, 2017). Coisas diferentes são o *discurso* — uma representação particular dos acontecimentos da *história* codificada em signos materiais (Ryan, 2005) — e a *fábula* — «[...] uma série de acontecimentos lógicos e cronologicamente relacionados que são causados ou vividos pelos atores [agentes que realizam ações]» (Bal & Van Boheemen, 2009, pp. 5-6).

Em outros termos, a *narrativa* pode ser «[...] a transmissão e interpretação, lógica e internamente coerente, de personagens e eventos relacionados. A transmissão e interpretação em conjunto dão significado à história, relacionando acontecimentos isolados numa história coletiva maior» (Latour et al., 2017, p. 62). Estrutura (princípio, meio e fim; antes e depois), personagens (principais/protagonistas, heróis ou heroínas e inimigas/antagonistas, vilões ou vilãs), contexto (cultural, social e histórico, mecanismos e regras respetivos), relações (entre personagens, positivas ou negativas) e significado (histórias individual e coletiva) são elementos presentes, frequente e maioritariamente implícitos, em cada *narrativa* (Latour et al., 2017).

A *narrativa*, enquanto estrutura cognitiva abstrata, tem um carácter camaleónico, uma capacidade de migração adaptativa, pode concretizar-se em suportes expressivos diversos, verbais, icónicos e verbo-icónicos (Reis & Lopes, 2007, como citados em Santana, 2017), não está adstrita a qualquer *medium* particular, podendo ser evocada por muitos *media* e tipos de signos, mas cada *medium* tem recursos expressivos variados e, por conseguinte, produz uma manifestação concreta de *narrativa* (Ryan, 2002, 2006, 2007).

No fundo, a *narrativa* é mutável, além de diferir de cultura para cultura, evolui de acordo com as inovações tecnológicas, (re)adapta-se às propriedades, recursos, ou

possibilidades, de cada *medium*<sup>18</sup> e, longe de ser estritamente condicionada por elas, joga com elas (Ryan, 2006), pode aproveitá-las, total ou parcialmente, ignorá-las ou ir além delas (Ryan, 2005). Na mesma medida, a natureza, peculiaridades e recursos de cada *medium* oferecem possibilidades narrativas diferentes e diversas, o seu poder de representação é variável (Ryan 2006; Schlegel, 2021). Por outro lado, o potencial narrativo sofre consequências na migração de um *medium* para outro, sendo preenchido e atualizado diferentemente (Ryan, 2006).

A *narrativa* domina, textual e discursivamente, o campo dos *media* (Peixinho & Araújo, 2017). Os *media*, analógicos e digitais, são grandes fazedores de narrativas, ficcionais ou factuais, grandes, pequenas e micro — ou até de narrativas alternativas e pós-narrativas —, informam e desinformam, produzem e reproduzem, preceitos, crenças e estereótipos, moldam o mundo, os seus habitantes e a representação de um e dos outros (Peixinho & Araújo, 2017).

A *narrativa* é uma máquina precisa, com mecanismos próprios e funções determinadas, de «textualização do mundo e da experiência», de «construção do humano» e de «ordenação do tempo para engendrar sentido» (Babo, 2017). A experiência temporal é marcada, organizada e esclarecida pela narratividade (Kearney, 2012), e as vivências individuais e coletivas reconstruídas em cenas narrativas (Bruner, 1998, como citado por Peixinho & Araújo, 2017).

A *narrativa* integra a natureza humana, é omnipresente na vida humana, e a partilha de narrativas é um impulso inato (Schlegel, 2021). A humanidade cria e apresenta significados, alcança conhecimento, estrutura a ação e ordena as experiências por meio da *narrativa* (Brockmeier & Harré, 2003; Kearney, 2012; Gulbrandsen & Just, 2013). Os seres humanos são e estão sujeitos à *narrativa* que, assim, se assume como um processo fazedor e revelador do mundo, individual e coletivo (Brockmeier & Harré, 2003; Kearney, 2012).

A *narrativa* varia de acordo com as circunstâncias, a sua forma e o seu conteúdo «[...] são dependentes do tempo e do espaço específicos em que a história é contada»

---

<sup>18</sup> *Medium* enquanto «uma categoria que faz verdadeiramente a diferença quanto às histórias que podem ser evocadas ou contadas, como são apresentadas, porque são comunicadas e como são vividas» (Ryan, 2005, p. 11).

(Bamberg & Wipff, 2020). Além disso, a *narrativa* é construída de acordo com o(s) propósito(s) do(s) narrador(es), apresenta uma descrição particular de passado, presente e futuro de acordo com estratégias, objetivos e expectativas próprias (Ummah, 2021), podendo manifestar uma construção do mundo que as sirva e orientar o comportamento (Carthy et al., 2020). Ao reproduzir e promover uma certa posição ou conjunto de valores, a *narrativa* «[...] é mais do que mera descrição; não só ilustra como explica (...) ao fazê-lo, a narrativa gera inevitavelmente significado, fazendo uma afirmação moral, por mais subtil que seja» (Copeland, 2019, como citado por Schlegel, 2021, p. 197). Ao mesmo tempo, a *narrativa* está sujeita ao pré-conhecimento, ideias e crenças pré-concebidas da(s) audiência(s), e a sua influência depende da conexão de uma miríade de outros fatores (Schlegel, 2021).

No entanto, a *narrativa*, factual ou fictícia, tem poder de persuasão, influencia e molda percepções, atitudes, valores e comportamentos de indivíduos e grupos (Schlegel, 2021), podendo servir propósitos positivos e negativos, de humanização e desumanização, de amor e de ódio (Carthy et al., 2020; Kearney, 2012; Latour et al., 2017). Ao salientar determinados aspetos ou pormenores de uma história e minimizar ou omitir outros, a *narrativa* pode influenciar a interpretação de factos e acontecimentos, conduzindo a uma perspetiva distorcida e a um entendimento da realidade incompleto ou enviesado. Ao apresentar uma interpretação particular sobre acontecimentos e questões, a *narrativa* tem potencial para promover certas ideias e ideologias, enquanto suprime outras, e influenciar crenças sociais e moldar visões políticas, culturais e históricas. Ao promover certos estereótipos e preconceitos, a *narrativa* pode moldar a percepção sobre indivíduos e grupos, reforçá-los e perpetuá-los, criar uma falsa percepção sobre pessoas ou grupo de pessoas e influenciar atitudes e comportamentos em relação a eles (Carthy et al., 2020; Kearney, 2012; Latour et al., 2017).

As *narrativas dominantes (master narratives)* — ou posições ideológicas predominantes (Bamberg, 2007b) — não só são guiões culturais, fornecem o rumo mundano e necessário dos acontecimentos, a inteligibilidade da vida quotidiana, mas também têm uma tendência para normalizar e naturalizar, restringindo e delimitando a capacidade de agir e, aparentemente, a ação (Bamberg & Andrews, 2004). As *narrativas*

*dominantes* estão frequentemente enraizadas em tradições históricas e culturais e são formadas por estruturas de poder. São amplamente aceites, definem o que é socialmente aceitável e culturalmente possível, moldam a forma como as sociedades compreendem o mundo e o seu lugar nele, influenciam o papel e a posição de cada indivíduo e grupo na sociedade (Latour et al., 2017).

As *narrativas dominantes* podem ser benéficas em alguns contextos, por exemplo, em casos de reabilitação de vícios ou de recuperação de doenças têm potencial para serem utilizadas como um mecanismo para lidar com a situação ou uma fonte de esperança (Carthy et al., 2020). Mas podem igualmente servir funções mais sinistras, assumindo-se como problemáticas ou opressivas — por exemplo, narrativas que perpetuam a desigualdade e a discriminação, como colonialismo, supremacia branca, patriarcado e heteronormatividade. As *narrativas dominantes* exploradas, recicladas, instrumentalizadas e apresentadas por grupos como a *Al-Qaeda* ou o autoproclamado Estado Islâmico (EI), coletivos neonazis e da direita radical, para sustentar ideologias extremistas violentas, estão entre as formas mais efetivas de persuasão e de mudança de atitude (Carthy et al., 2020).

As *narrativas*, em particular as dominantes, têm a possibilidade de ser, simultaneamente, explícitas e implícitas, poderosas e penetrantes, mas não são fixas nem imutáveis. Considerando a sua ubiquidade, acelerada pela *era digital*, na vida humana, e o seu poder persuasivo aparente, as *narrativas* têm o potencial para assumir, simultaneamente, um papel crucial na reprodução, reforço e legitimação, assim como na prevenção, oposição e modificação de crenças, preconceitos e estereótipos, de formas de discriminação, radicalismo e extremismo, de estruturas de poder sociais e políticas desiguais (Latour et al., 2017; Manchola-Castilho, 2002; Schlegel, 2021). Ou, recentrando o tema deste trabalho, as narrativas podem alimentar e desnutrir o *discurso de ódio*, nomeadamente enquanto *contranarrativas* e/ou *narrativas alternativas*, promovendo os direitos humanos.

### I. 1.2.1 Narrativa digital

O desenvolvimento tecnológico, desde a invenção da imprensa ao acesso público às mais recentes ferramentas de inteligência artificial, teve e tem, como terá, um impacto significativo na *narrativa*, revolucionando a forma como é criada, partilhada e experienciada (Bakhtiary & Behzadi, 2023). A evolução da tecnologia digital fomentou a revolução, nomeadamente através de melhorias multimédia, envolvimento de utilizador, personalização e colaboração, introduzindo alterações significativas na teoria e na prática *narrativa* (Page & Thomas, 2011).

Os aspetos essenciais da *narrativa* — tempo, espaço, personagens e eventos — mantêm-se no digital, mas adquirem novas características, apresentam novos comportamentos e são frequentemente reconfigurados radicalmente (Peixinho & Araújo, 2017; Rettberg & Rettberg, 2010; Ryan, 2006). As tecnologias do computador e da internet desafiaram a *narrativa* linear tradicional e potenciaram novas possibilidades de *narrativa* interativa e não-linear (Couldry, 2008; Kearney, 2012) ou multissequencial e multiforme, com estruturas caleidoscópicas e metamórficas (Murray, 2003). A *narrativa* libertou-se do suporte específico e distribuiu-se por outras plataformas e dispositivos, tais como *telemóveis inteligentes* e *tablets*, redefinindo as relações entre espaço e tempo (Canavilhas et al., 2017), escritor/narrador, leitor/audiência e *media* (Rettberg & Rettberg, 2010). As antigas narrativas foram substituídas por novas, ou renovadas, com múltiplas vozes, enredos e contextos multimediáticos, entrelaçando-se e justapondo-se (Kearney, 2012). As narrativas digitais são mais simuladas, emergentes, participativas e simultâneas do que, respetivamente, representativas, escritas, recetivas e retrospectivas (Ryan, 2006). A *era digital* é também a era da *textualidade ubíqua* (Rettberg & Rettberg, 2010).

As primeiras experiências com *narrativa* não-linear iniciaram-se na literatura impressa e não digital no decénio de 1960, destacando-se entre as mais conhecidas «Rayuela» ou «Hopscotch» de Julio Cortázar, «Se Numa Noite de Inverno um Viajante» de Italo Calvino e «Fogo Pálido» de Vladimir Nabokov (Rettberg & Rettberg, 2010). As formas iniciais de *narrativa digital* surgiram no mesmo período da interface de linha de comandos (*Command Line Interface — CLI*), como o *robô de conversação em linha* (*chatbot*) ELIZA (1966), utilizado nos primeiros computadores, e de aventuras baseadas

em texto ou jogos de ficção interativa, como «Colossal Cave Adventure» (1976). Todavia, só desde os anos de 1980, com a generalização da utilização do computador pessoal, evoluiu amplamente a *narrativa digital* num processo estendido por vários decénios, e ainda em curso, reflexo do desenvolvimento da tecnologia, da computação em rede e dos meios de comunicação, assim como das práticas literárias e artísticas contemporâneas (Page & Thomas, 2011; Torres et al., 2023).

A introdução do hipertexto e o desenvolvimento da *World Wide Web* no decénio de 1990 originaram formas narrativas nativas do ambiente digital, nomeadamente ficção hipertextual, narrativas mediadas por computador e literatura baseada na rede, que vieram completar os hipertextos dos anos de 1980, publicados em disquetes e CD-ROM pela Eastgate Systems (Page & Thomas, 2011; Torres et al., 2023). «Invisible Seattle: The novel of Seattle» (1983), «Afternoon, A Story» (1987), «Victory Garden» (1992), «Patchwork Girl» (1995), «253» (1997), «The Unknown» (1999) e «Disappearing Rain» (2001) são alguns exemplos nestes domínios. Entretanto, uma gama de recursos multimodais cada vez mais sofisticada foi integrada nos textos digitais com o desenvolvimento de *software* multimédia, como o Flash da Adobe<sup>19</sup>, enriquecendo a hiperficção de elementos visuais, como gráficos e animações, e permitindo a utilização destes como parte da *interface* interativa — «These Waves of Girls» (2001) e «Inanimate Alice» (em curso) exemplificam-no (Page & Thomas, 2011).

O advento da *Web 2.0*, nos finais do decénio de 1990, influenciou tanto a criação quanto a receção de narrativas (Page & Thomas, 2011). A produção, edição e disseminação simplifica-se, abre-se a qualquer utilizador com competências técnicas relativamente diminutas e expande-se por (micro)blogues, fóruns de discussão, *wikis* e plataformas de *media* sociais. Uma proliferação de comunidades de narração de histórias (*storytelling communities*), como o *Center for Digital Storytelling*, exploram os recursos textuais em linha. Os jogos eletrónicos simulam a *vida* — por exemplo, «The Sims» (2000) —, incorporam elementos do mundo real na *narrativa digital* — nomeadamente, os jogos de realidade alternativa (*Alternate Reality Games* — ARG), como «I Love Bees» (2004) — e integram a inteligência artificial — refira-se «Façade» (2005) —, esbatendo a fronteira entre ficção e realidade. As plataformas de *media*

---

<sup>19</sup> O Adobe Flash conheceu o seu fim a 12 de janeiro de 2021.

sociais originam histórias novas e híbridas de múltiplos modos e gêneros, numa narração predominantemente visual, criativa, colaborativa, imediata e fragmentada, também em atualizações de estados, publicações e comentários (Page & Thomas, 2011). A excessiva saturação do ambiente informativo em linha dota a *narrativa digital* de alguns atributos próprios que se materializam na pressão para (1) misturar texto, som, vídeo e imagem estática, (2) limitar a sua duração, (3) uniformizar formatos, esquemas ou sequências, e (4) antecipar audiências involuntárias e inoportunas (Couldry, 2008). A expansão do desenvolvimento de novas formas narrativas continua, e continuará, a par com a inovação tecnológica (Page & Thomas, 2011).

No decénio de 2010, a propagação de narrativas, em série e colaborativas, nas plataformas de *media* sociais é seguida pela emergência de narrativas interativas em forma de videojogos, documentários (*i-docs*) e peças jornalísticas, narrativas de realidade virtual e aumentada, assim como narrativas geradas por inteligência artificial (*AI*). Os formatos emergentes incluem também narrativas para abordar, e com potencial para representar, expressar e compreender (Murray, 2003), questões sociais altamente complexas e potencialmente fraturantes, tais como racismo, colonialismo, herança cultural e desinformação (Koenitz et al., 2023; Perkis & Taveter, 2022; Silva et al., 2023). A título de exemplo, refiram-se: «Heavy Rain» (2010), «Life is Strange» (2015), «Bad News» (2018) e «Starlight Stadium» (2023); «Collapsus» (2010), «Bear 71» (2012) e «Hearts and mind: the interrogations» (2016); «Snow Fall» (2012); «Clouds Over Sidra» (2015); «Spheres» (2018); «Sunspring» (2016), e «1 the Road» (2018).

Na atual sociedade, em rede e globalizada, «[...] as narrativas digitais incluem histórias em jogos de computador, literatura eletrónica, realidade virtual e aumentada, *chatbots* e aplicações *web* e móveis, bem como histórias que circulam nas redes sociais ou são geradas por inteligência artificial» (Torres et al., 2023, p. 12). Com a mediação dos computadores nos seus contextos em rede, o processo narrativo ganhou um novo agente, os algoritmos, cada vez mais participante na formação de narrativas (Torres et al., 2023). Além de narrador(es) e audiência(s), as narrativas digitais passaram a ser negociadas também pelos sistemas que as influenciam, permitem e restringem (Torres et al., 2023). Se o funcionamento orientado por algoritmos é a característica fundamental dos sistemas digitais, a *interatividade*, enquanto categoria genérica que abrange uma

grande variedade de relações entre um utilizador e um *texto*, é a propriedade mais relevante e distintiva da *narrativa digital* (Ryan, 2006).

Os *media* digitais têm uma capacidade singular de remediar, múltiplas possibilidades de (re)transmitir e (re)transformar (Couldry, 2008). Assim, as narrativas da atualidade são «[...] voláteis, instáveis, dinâmicas e imprevisíveis» (Monteiro & Carvalhais, 2023, p. 62), multimediáticas ou transmediáticas, interativas e colaborativas. São configuradas por relatos provenientes de uma multiplicidade de fontes e plataformas, híbridos e fragmentados, descontínuos e dispersos, efémeros e caleidoscópicos, fatuais e fictícios, informativos e de entretenimento, locais e universais, públicos ou privados, que as estende para vários lados, formando uma teia virtual que enreda todos na produção de significado e configura as representações, seletivas e parciais, do mundo (Motta, 2017) cada vez mais desterritorializado (Kearney, 2012).

No cenário digital híbrido de proliferação e saturação, a realidade é permanentemente (re)construída pelos *media* e utilizadores (Amaral, 2019), numa multiplicidade de ecrãs, num *agora* infinito no qual tudo está potencialmente presente (Gulbrandsen & Just, 2013) e em universos oscilantes (Peixinho & Araújo, 2017), com impacto social e implicações ideológicas (Amaral, 2019). Somos participantes ativos no processo de criação de significado, adotamos, mas também adaptamos, influenciamos e, portanto, remodelamos formas consensuais de ver a realidade, ou seja, praticamos a remediação, e construímos instintivamente as nossas próprias representações distintas dessa realidade, somos *bricoleurs*, por via de e como processos de narração (Deuze, 2006). Os numerosos métodos misturados, manipulados e convergentes através dos quais geramos e consumimos informação estão a alterar rapidamente a forma como nos interligamos, interagimos e damos sentido às nossas vidas, pelo meio de narrativas (Deuze, 2006).

Ainda que os *media*, interação, acessibilidade e envolvimento da *narrativa* tradicional e da *narrativa* digital sejam diferentes, o poder de seduzir e inspirar é partilhado por ambas (Bakhtiary & Behzadi, 2023). Todavia, devido aos seus recursos e práticas, ubiquidade, alcance e mobilidade, a influência da segunda, acedida através da internet e das plataformas de *media* sociais, é imensa (Latour et al., 2017; Ummah, 2021). A expansão da *narrativa* digital produz desafios relevantes, como a hibridização

de ficção e facticidade (Peixinho & Araújo, 2017), a saturação informativa (Bakhtiar & Behzadi, 2023), o condicionamento de hierarquias de voz e ação (Couldry, 2008), a polarização de visões do mundo e a propagação de discriminação (Latour et al., 2017).

Os reptos significativos das *narrativas digitais* problemáticas ou opressivas, mas também, e em particular, dominantes, estimularam o desenvolvimento e a elaboração de estratégias alternativas para as desafiar e contestar igualmente em linha, nomeadamente, sob a forma de *contranarrativas* e de *narrativas alternativas*, primeiro, no contexto do extremismo e do terrorismo violento, bem como, depois, no âmbito do *discurso de ódio*.

### I. 1.3 Contranarrativa

O termo *contranarrativa* é relativamente recente, controverso e obscuro, acolhe várias perspetivas ambíguas, não existindo um entendimento comum quanto à sua concetualização, nem uma definição única e comumente aceite (Briggs & Feve, 2013; Carthy et al., 2020; Carthy & Sarma, 2023; Glazzard, 2017; Tuck & Silverman, 2016; Van Eerten et al., 2017). Apesar de ser frequentemente utilizado nos círculos académicos e políticos, está concetualmente subdesenvolvido, tem uma teoria apenas parcialmente articulada, ainda confusa e imprecisa (Carthy et al., 2020; Glazzard, 2017). Do mesmo modo, a investigação sobre o seu impacto e eficácia, designadamente no domínio que o operacionalizou e generalizou, está por sintetizar e completar (Carthy et al., 2020; Haider, 2022; Silverman et al., 2016). Todavia, a *contranarrativa* não é «uma causa perdida», pode ter um potencial considerável, mas precisa de ser melhorada (Glazzard, 2017, pp. 3 e 16), e pode sê-lo (Schlegel, 2021), até porque desenvolvê-la «[...] é uma alternativa necessária» (Tuck & Silverman, 2016, p. 4).

A ideia de aplicar *contranarrativa* em resposta a retóricas hegemónicas problemáticas ou opressivas tem já algum tempo, sendo investigada e praticada, desde os anos de 1990, nas áreas da psicologia do desenvolvimento, da sociologia, da ciência política, da educação, dos estudos culturais e dos meios de comunicação social (Carthy et al., 2020; Carthy & Sarma, 2023; Latour et al., 2017). A abordagem tem surgido sob a forma de narrativas «contrafactuais», para contestar narrativas históricas inexatas, e de

«contra-história» em relação ao autismo, infertilidade e deficiência, papéis de gênero e ecologia, história e impacto do colonialismo, origem racial e relações inter-raciais, tal como tem sido aplicada igualmente, de forma mais ampla, à prevenção da violência e à promoção da paz (Carthy & Sarma, 2023; Latour et al., 2017). Além disso, a utilização de «contrapropaganda», como estratégia de combate à propagação de ideias comunistas, interessou igualmente investigadores e decisores políticos no decénio de 1950 (Carthy & Sarma, 2023). Contudo, o conceito conheceu particular desenvolvimento e aplicação no início do século XXI, no domínio da prevenção e do combate, ou desconstrução, do terrorismo e do extremismo violento, assim como dos seus processos de propaganda, envolvimento, radicalização, recrutamento e retenção, direta ou indiretamente, em linha e fora dela. A assunção lógica de que ao papel crucial da *narrativa* na comunicação extremista e nos processos de radicalização poderia equivaler o apoio da *narrativa* para preveni-los e combatê-los foi o ponto de partida de um processo em curso que, no último decénio, motivou igualmente um aumento da investigação nas duas áreas (Schlegel, 2021).

Os Estados Unidos da América e o Reino Unido registam a grande maioria dos esforços envidados ou patrocinados por governos para desenvolver *contranarrativas*, nomeadamente para contestar o apelo da «narrativa singular» da *Al-Qaeda*<sup>20</sup>, após os ataques terroristas de 11 de setembro de 2001 e da campanha militar que se lhes seguiu, nomeada «*The Global War on Terrorism*» ou, vulgo, «*War on Terror*» (Schmid, 2014). Neste âmbito, a primeira articulação registada de *contranarrativa* surgiu no desenvolvimento da Estratégia Contraterrorista Britânica (*Counter-Terrorism Strategy — CONTEST* [Glazzard, 2017]), arquitetada por Sir David Omand, em 2003, o qual, dois anos depois, observava que é urgente «[...] uma contranarrativa que ajude os grupos expostos à mensagem terrorista a dar sentido ao que estão a ver à sua volta» (Omand, 2005, p. 109). Desde então, a adoção de *contranarrativa* para prevenir e lutar contra a *narrativa extremista*, especialmente a jihadista/radical islâmica, mobilizou grupos de reflexão, governos, Organizações da Sociedade Civil (OSC) ou Não Governamentais (ONG), organismos multilaterais, órgãos de comunicação social e, até, plataformas de

---

<sup>20</sup> Sobre a «narrativa singular» da *Al-Qaeda* vide, por exemplo, Schmid (2014).

*media* sociais (Briggs & Feve, 2013; Glazzard, 2017; RAN, 2015; Schmid, 2014; Ummah, 2021).

A mobilização ocorreu, em particular, também no Reino Unido — com grande destaque para o Institute for Strategic Dialogue (ISD) e a Quilliam Foundation — e nos Estados Unidos da América — designadamente, o Consortium for Strategic Communication (CSC) e a Research and Development Corporation (RAND) —, mas também na Europa — refiram-se o International Centre for Counter-Terrorism (ICCT), a Against Violent Extremism Network (AVE) e a Radicalisation Awareness Network (RAN) — e para lá do Ocidente — mencionem-se a Qatar International Academy for Security Studies (QIASS) e o Hedayah, centro internacional de excelência para lutar contra o extremismo violento, sediado em Abu Dhabi (Glazzard, 2017; Schmid, 2014).

A adoção e prática, de alguma forma generalizada, de *contranarrativa* deriva, especialmente, da presença, eficácia e proliferação de grupos e conteúdos extremistas na internet e nas plataformas de *media* sociais. A título de exemplo, refiram-se a presença da *Al-Shabaab*, a rede terrorista afiliada da *Al-Qaeda*, no (ex-)Twitter, do movimento de direita radical English Defence League (EDL) no Facebook e da comunidade defensora da supremacia da raça branca no Stormfront (Briggs & Feve, 2013; RAN, 2015; Tuck & Silverman, 2016; Van Eerten et al., 2017). Do mesmo modo, decorre como solução potencial, alternativa pró-ativa e «suave», aos mecanismos e às práticas tradicionais, reativas, restritivas e negativas, operadas tanto por medidas legislativas e políticas, quanto por «moderação» ou «filtragem», de identificação e punição, isto é, de suspensão, bloqueio e eliminação de páginas, perfis, contas e utilizadores, de alguma forma, de silenciamento e de censura (Benesch, 2014a; Briggs & Feve, 2013; Haider, 2022; RAN, 2015; Tuck & Silverman, 2016; Van Eerten et al., 2017).

A diversidade de atores correlaciona-se com a multiplicidade de abordagens, fazendo de *contranarrativa* um termo abrangente que compreende «[...] uma vasta gama de atividades com diferentes objetivos e táticas, desde a diplomacia pública e as comunicações estratégicas do governo até às campanhas orientadas para desacreditar as ideologias e ações de extremistas violentos» (Briggs & Feve, 2013, p. 6). O conjunto complexo de estratégias e iniciativas, posto em prática, em vários países, tanto da base para o topo quanto do topo para a base, de em linha para fora dela, do local para o

global (REACT, 2019, p. 7; Tuck & Silverman, 2016, p. 4), pode ser sintetizado num «espectro de contra-mensagens» ou «de contranarrativas» com três tipos principais, a saber:

- (1) Comunicações Estratégicas do Governo — ações para fazer passar a mensagem sobre o que o governo está a fazer, a sua política e lógica, incluindo atividades de sensibilização do público[-alvo], de estabelecimento de relações com os principais círculos eleitorais e públicos, bem como refutar a desinformação;
- (2) Narrativas Alternativas — ações que visam minar as narrativas extremistas violentas, concentrando-se no que somos "a favor" e não no que somos "contra" [aqui objeto do ponto seguinte, l. 1.4];
- (3) Contranarrativas — ações para desconstruir, desacreditar e desmistificar diretamente as mensagens extremistas violentas. (Briggs & Feve, 2013, pp. 2 e 6; RAN, 2015, p. 4).

Assim, as *contranarrativas* procuram desconstruir, desacreditar e enfraquecer, de um modo mais direto e explícito, as narrativas extremistas, através de um conjunto vasto e variado de objetivos e mensagens, com ideologia, lógica, factos ou humor, paródia e sátira, emoção, teologia e exposição de hipocrisias, mentiras e inverdades, para desde semear a dúvida até desvincular ou desradicalizar (Briggs & Feve, 2013, p. 16; RAN, 2015, p. 4). De entre os propósitos e comunicações contam-se: desgastar o enquadramento intelectual e dismantelar a ideologia extremista, destacar o seu impacto negativo, expor as suas incongruências ou contestar os seus objetivos e eficácia geral, como também tentar desacreditar ou deslegitimar os seus mensageiros (Briggs & Feve, 2013, p. 16; Tuck & Silverman, 2016, p. 4).

As *contranarrativas* destinam-se a diferentes públicos-alvo — indivíduos, grupos e redes no círculo do radicalismo, como simpatizantes, apoiantes passivos ou ativos —, mas devem ser direcionadas a um grupo-alvo específico e promovidas por mensageiros credíveis, nomeadamente ex-extremistas violentos ou sobreviventes do extremismo violento, ativistas comunitários e académicos religiosos — e não necessariamente pelos governos que, devido à sua credibilidade diminuta como promotores, devem assumir um papel indireto e de facilitadores (Briggs & Feve, 2013, p. 16).

As campanhas «Say No To Terror», de autor(es) desconhecido(s), iniciada em 2009, «Trojan T-Shirt», da EXIT Alemanha, de 2011, e «Rechts gegen Rechts», do Gesellschaft Demokratische Kultur (ZDK) alemão, de 2014, os projetos «Abdullah X», de autor anónimo, criado no Reino Unido, em 2014, «Extreme Dialogue», do Institute of Strategic Dialogue (ISD), lançado no Canadá, em 2015, e «Moonshoot», criado no Reino Unido, em 2015, assim como a iniciativa «One2One», do ISD, de 2015, exemplificam<sup>21</sup> o espectro de *contranarrativas* executadas para prevenir e combater, ou desconstruir, o terrorismo e o extremismo violento (RAN, 2015; Van Eerten et al., 2017). Além destas, mencione-se ainda os projetos «Counter@ct: prevenção e combate à radicalização online», promovido pela Associação Portuguesa de Apoio à Vítima (APAV), entre 2019 e 2021, e «Brasileiras Não Se Calam», iniciado em julho de 2020.

A popularidade, o grau de aceitação e de consenso, o número substancial de guias, relatórios e estudos de como desenvolver as narrativas e torná-las mais efetivas, produzidos, sobretudo, no ocidente, contrastam com a escassez de evidências do impacto e da eficácia das *contranarrativas*, sendo ambos contestados (Carthy et al., 2020; Carthy & Sarma, 2023; Glazzard, 2017; Haider, 2022; Schlegel, 2021). Apesar de serem cada vez mais reconhecidas como ferramentas de prevenção ou de combate ao terrorismo e ao extremismo violento (Schlegel, 2021; Silverman et al., 2016), do seu potencial vital nestes domínios (Reynolds & Tuck, 2016), e de terem a capacidade de, pelo menos, diminuir alguns fatores de risco (Carthy et al., 2020), as *contranarrativas* comportam perigos, podem induzir resistência (Carthy & Sarma, 2023) e produzir efeitos secundários não pretendidos (Rieger et al., 2018), imprevisíveis ou agravantes (Carthy et al., 2020) ou ser percebidas apenas como sendo reativas ou defensivas (Schmid, 2014).

Portanto, as *contranarrativas* devem ser desenvolvidas com cautela (Carthy & Sarma, 2023), empregar «[...] os recursos estéticos e afetivos da narrativa» (Glazzard, 2017, p. 16), com narratividade engrandecida para maiores efeitos persuasivos (Frischlich et al., 2017, 2018, como citados por Schlegel, 2021) ou, talvez, integrar a ficção (Schlegel, 2021) e, porventura, sejam mais úteis no contexto da prevenção (Van Eerten et al., 2017). Mais ainda, as *contranarrativas* e o seu potencial eventual não

---

<sup>21</sup> Para outros exemplos de *contranarrativas* vide, e.g., «[Coleção de Abordagens e Práticas](#)» da RAN.

resolverão *per se* a radicalização e o extremismo violento. Têm de ser entendidas como um instrumento de um conjunto mais amplo de medidas, estratégias e processos, tais como ação política e legal, literacia mediática digital e educação para os direitos humanos, que envolvem vários elementos e agentes numa missão comum, em linha e fora dela (Haider, 2022; Latour et al., 2017; Schlegel, 2021; Silverman et al., 2016; Ummah, 2021).

#### I. 1.4 Narrativa alternativa

O conceito de *narrativa alternativa* não é novo, mas apenas integrou recentemente a série de estratégias e ações, executadas em linha e não em linha (*vide* ponto I. 1.3), de prevenção e de combate, ou desconstrução, do terrorismo e do extremismo violento, sendo ainda objeto de debate «[...] sobre o seu potencial concetual, a sua verdadeira distinção de contranarrativa e a natureza exata da sua relação com o discurso dominante» (Carthy, 2021, p. 4).

Ainda que, na prática, se possam sobrepor (Bamberg & Wipff, 2020), confundir ou, mesmo, coexistir (Latour et al., 2017), ou ser referenciadas coletivamente (Schlegel, 2021), *contranarrativa* e *narrativa alternativa* são termos distintos (Bamberg & Wipff, 2020) e informados por quadros teóricos diferentes e divergentes (Carthy, 2021). Enquanto o primeiro tem por base uma estratégia de *resposta-mudança* (*response-changing strategy* [Carthy, 2021, p. 5]), a teoria e a evidência da contra-argumentação, uma força ilocutória fundamental de contrariar (Bamberg & Wipff, 2020), o segundo alicerça-se «[...] mais em quadros de reforço da resiliência (*e.g.*, prevenção e pensamento crítico)» (Carthy, 2021, p. 5) e é identificado pelo conteúdo (Bamberg & Wipff, 2020).

A *contranarrativa* procura desconstruir, deslegitimar e enfraquecer outra narrativa de forma direta, como visto com maior detalhe no ponto I. 1.3. Já a *narrativa alternativa* procura minar os seus pressupostos dominantes, diferenciando-se totalmente da sua lógica, discurso, argumentos centrais e influência (Carthy, 2021). Centra-se no que a sociedade é «a favor» (Briggs & Feve, 2013; RAN, 2015), procura mostrar várias perspetivas, propõe uma versão ou interpretação diferente «[...]

ênfatizando alternativas positivas que ultrapassam a imagem negativa das narrativas que pretendem combater, sem as reforçar ao focarem-se nelas» (Latour et al., 2017, p. 13).

A *narrativa alternativa* baseia-se em pressupostos opostos à *narrativa dominante*, procura projetar uma história diferente e positiva, fundada na moral, em valores e normas sociais partilhados, sobretudo, no Ocidente, tais como: democracia, Estado de direito, pluralismo, respeito pelos direitos humanos, liberdade, igualdade, tolerância, abertura, não discriminação, educação e paz para todos (Briggs & Feve, 2013; RAN, 2015; Schmid, 2014). Ao realçar os valores partilhados, as causas comuns e a solidariedade, a *narrativa alternativa* tenta influenciar simpatizantes, mas não apoiantes ativos, dos fundamentos extremistas, ou apoiar a união da maioria silenciosa contra o apelo do extremismo violento (Briggs & Feve, 2013, p. 12). O objetivo é relegar a narrativa dominante problemática, ou seja, retirar-lhe a importância ou o lugar na sociedade (Carthy, 2021).

As *narrativas alternativas* assumem igualmente diversas formas — nomeadamente, iniciativas, projetos, campanhas, movimentos e/ou intervenções — e podem ser projetadas por muitos atores diferentes dos governos — que, pela credibilidade diminuta, devem facilitar a sua conceção e apresentação a partir da base, utilizá-las nas declarações públicas de políticos e reforçá-las através de políticas e práticas governamentais — e da sociedade civil — desde redes inter-religiosas e intercomunitárias de ativistas influentes a artistas populares (Briggs & Feve, 2013, p. 12). Tal como podem ter (co)financiamento público, privado ou comunitário e desenvolver-se aos níveis local, nacional e/ou transnacional (Briggs & Feve, 2013, p. 12).

Os projetos «Radical Middle Way (RMW)», criado no Reino Unido na sequência dos ataques terroristas em Londres, a 7 de julho de 2005, e «My Fellow American», lançado pela Unity Productions Foundation (UPF), nos Estados Unidos da América, em 2011, assim como o movimento «Dare to be Grey», fundado em 2016 por um grupo de estudantes da Universidade de Utrecht, nos Países Baixos, exemplificam-nas (Briggs & Feve, 2013; Van Eerten et al., 2017).

A *narrativa alternativa* é também contestada. Foi considerada uma abordagem «pouco popular», porque comprometida com o *status quo*, tal como foi sugerido que o

seu impacto para disputar a narrativa extremista é limitado, uma vez que não produz mudanças substanciais no quadro mais geral (Carthy, 2021, p. 4). Além disso, tem potencial para produzir reações mistas (Carthy, 2021, p. 5) e para ter igualmente efeitos negativos de resistência e de «bumerangue» ou «tiro pela culatra» (*boomerang* ou *backfire effects*) [Carthy, 2021, p. 12]. Por outro lado, a avaliação do seu impacto e eficácia é notoriamente difícil e, sempre, provisória (Carthy, 2021, p. 8).

Apesar disso, ou também por isso, a *narrativa alternativa*, a sua teoria e a sua prática têm sofrido transformações sucessivas e progressivas — por exemplo, passaram a integrar a Teoria da Mudança (*Theory of Change* [ToC]) e as diretrizes preconizadas pelo Modelo GAMMMA +<sup>22</sup> —, fazendo evoluir também o seu impacto e a sua eficácia eventuais no contexto da prevenção e do combate, ou desconstrução, do terrorismo e do extremismo violento (Carthy, 2021).

### I. 1.5 Contranarrativas, Narrativas Alternativas e Discurso de Ódio

A prevalência e visibilidade do *discurso de ódio*, particularmente, em linha, os seus efeitos e as suas consequências negativas, mobilizaram, entretanto, e continuam a mobilizar, a *contranarrativa* e a *narrativa alternativa* como estratégias de combate, concretizadas sob diversas formas, por diferentes atores, e com resultados discutíveis. A relação entre os três elementos é complexa e o debate em torno dela é permanente e, especialmente, informado pela dicotomia *discurso de ódio/liberdade de expressão*.

Os programas para combater o *discurso de ódio*<sup>23</sup>, mormente em linha, desenvolvidos por muitos atores governativos, da sociedade civil e privados, incluindo as plataformas de *media* sociais, podem ser sintetizados em duas categorias principais: (1) «disrupção» e (2) «contra-mensagens» (Haider, 2022, p. 4). A primeira visa prevenir e proibir a difusão de *discurso de ódio*, incluindo vários mecanismos específicos, como preceitos legislativos e políticos, bem como práticas de moderação, filtragem, remoção e bloqueio de conteúdos e utilizadores (Haider, 2022, p. 4). A segunda, por outro lado, exige uma abordagem de envolvimento e relaciona-se com comunicações estratégicas

---

<sup>22</sup> Sobre o Modelo GAMMMA +, *vide* Ritzmann et al. (2019).

<sup>23</sup> Tal como são utilizados para combater a desinformação (aqui apenas abordado em relação com o tema principal) e as *narrativas extremistas* (*vide* pontos I 1.3 e I 1.4).

que se opõem ao *discurso de ódio* e/ou que fornecem mensagens alternativas para influenciar as crenças e as atitudes das pessoas a ele expostas e diminuir o apelo das suas mensagens (Haider, 2022, p. 4). Ambas apresentam prós e contras.

As «contra-mensagens», onde se inscrevem o «contradiscurso», a *contranarrativa* e a *narrativa alternativa*, são motivadas pelas desvantagens, inconvenientes e efeitos indesejáveis, ou perigos, da disrupção, surgindo como uma resposta, ou «remédio», preferível, para combater o *discurso de ódio* com «mais discurso» (Gagliardone et al., 2015; Gelber, 2012; Lepoutre, 2020; Mathew et al., 2019). O recurso a «mais discurso» para remediar o discurso prejudicial é uma ideia familiar no pensamento democrático-liberal há quase um século (Mathew et al., 2019). Apesar de conter também riscos, este ideário alcançou uma significativa adesão de várias instituições e organizações, envolvidas na prevenção e combate ao *discurso de ódio*, nos últimos dois decénios.

O «contradiscurso» é, somente, uma «resposta» ou um «comentário» direto de um «contra-interlocutor» que, como o próprio prefixo na aceção de oposição sugere, procura enfrentar e neutralizar o *discurso de ódio*, ou «perigoso», as suas intenções e os seus potenciais efeitos prejudiciais, incluindo a violência (Bahador, 2021; Benesch, 2014b; Benesch et al., 2016; Mathew et al., 2019). A réplica adota diferentes abordagens, tons e estratégias que podem ser sintetizadas em oito, algumas das quais bem descritas na investigação dedicada à persuasão e comunicação, e não exclusivas. Elas são: (1) indicar e corrigir afirmações e perceções, nomeadamente com factos; (2) apontar contradições e deturpações; (3) alertar para as consequências em linha e fora dela; (4) afiliação com o orador inicial ou o grupo-alvo; (5) denunciar o *discurso de ódio*; (6) utilizar imagens e vídeos; (7) desacreditar o *discurso de ódio* por meio do humor ou sarcasmo, e (8) utilizar um tom, particular e notavelmente, positivo ou linguagem hostil (Benesch et al., 2016, p. 17; Mathew et al., 2019, pp. 371-373).

Da mesma forma, a intervenção do «contra-interlocutor» pode resultar em cinco tipos de respostas: «[...] [1] pedido de desculpas ou retratação, [2] eliminação do comentário ou da conta, [3] a criação de mais *discurso de ódio*, [4] uma conversa sustentada e civilizada e [5] provocar mais contradiscurso da[s] audiência[s]» (Benesch et al., 2016, pp. 31-33). Estas réplicas, tanto positivas quanto negativas, são um

indicador do (in)sucesso do «contradiscorso» em relação ao(s) autor(es) e público(s) (Benesch, 2016, p. 31). O «contradiscorso» de líderes influentes, de várias fontes em uníssono e de refutação de falsos rumores (Benesch, 2014b, p. 17), assim como o «contradiscorso» apoiado (Gelber, 2012), do Estado (Lepoutre, 2019) e de figuras públicas (Article 19, 2015), estão entre as formas mais auspiciosas para prevenir e combater o *discurso de ódio*, a intolerância e a discriminação.

O combate ao *discurso de ódio* com «contradiscorso» é mais rápido, versátil e reativo, gerado de qualquer local e em qualquer língua, por quase todas as pessoas, na maioria dos casos pelos próprios utilizadores, sem quaisquer instituições ou leis, e preserva os preceitos do espaço público livre e aberto ao debate (Mathew et al., 2019, p. 370). As vantagens fazem do «contradiscorso» um meio efetivo (Benesch et al., 2016; Mathew et al., 2019; Schieb & Preuss, 2016) ou, pelo menos, promissor (Schäfer et al., 2023) e fomentador de um certo otimismo (Siegel, 2020) ou, na generalidade, preferível à supressão de discurso (Gagliardone et al., 2015; Lepoutre, 2019), mas utilizá-lo tem um preço (Schäfer et al., 2023), já que pode ser insuficiente, senão mesmo ilusório (Gelber, 2012).

Por outras palavras, a réplica por «contra-interlocutor(es)» tem o potencial para tanto atenuar o alcance, a visibilidade e as consequências ou danos do *discurso de ódio* (Fonseca et al., 2024; Schäfer et al., 2023; Siegel, 2020) quanto para reproduzi-lo, amplificá-lo e solidificá-lo, ser menos recordável ou memorável (Ernst et al., 2017) e ter efeitos polarizadores (Schäfer et al., 2023). Além disso, pode impactar consideravelmente uma determinada audiência, mas o seu impacto depende em muito do número de oradores de ódio que a constituem (Schieb & Preuss, 2016); bloquear danos se reunidas várias competências e capacidades (Lepoutre, 2019), assim como ter algum efeito nas audiências em determinadas circunstâncias (Bahador, 2021). Resumidamente, o potencial do «contradiscorso» é condicional e o conhecimento sobre os seus possíveis danos colaterais é ainda diminuto (Siegel, 2020).

A *contranarrativa* e a *narrativa alternativa*, por sua vez, são categorias mais amplas, confundem-se, sobrepõe-se e coexistem. Podem aplicar-se «[...] a quase todos os conteúdos destinados a combater o racismo ou o *discurso de ódio*, e independentemente do meio de comunicação utilizado, desde vídeos, *hashtags*,

*memes*, publicações nas plataformas de *media* sociais ou artigos em blogues ou jornais» (REACT, 2019, p. 9 [itálicos nossos]). Na maioria dos casos, a utilização por diversos atores é motivada pela dupla funcionalidade de ambas: lutar contra o ódio e a desinformação enquanto convidam a(s) audiência(s) a examinar diferentes opiniões, práticas e pontos de vista (REACT, 2019, p. 9).

A *contranarrativa* começa sempre por ser um «contraponto», um «frente a frente» (*vis-à-vis*) contra algo ou uma reação a um acontecimento ou narrativa e tem propensão para ser progressiva (Bamberg & Wipff, 2020; REACT, 2019). Todavia, diferencia-se exclusivamente pela força ilocutória de contrariar e transformar os pressupostos de fundo que, tipicamente, apoiam uma outra narrativa ou visão em geral, da qual se separa e à qual se pretende opor, para procurar alterar normas e crenças de um conjunto mais vasto de destinatários (Bamberg & Wipff, 2020; Benesch et al., 2016). Enquanto o «[...] contradiscurso ataca e nega o sentimento original de uma mensagem, mas a reformula no processo», a *contranarrativa* oferece quadros alternativos (Poole et al., 2021, p. 1437).

A ambição de responder e possivelmente mudar as narrativas negativas, problemáticas e opressivas com poder discursivo, como as de ódio, cada vez mais comuns e presentes, fundam a utilização da *contranarrativa* em oposição ao *discurso de ódio* e à discriminação e violência que pretende multiplicar, justificar ou difundir (REACT, 2019; Latour et al., 2017). Por conseguinte, a *contranarrativa* é uma «medida de mitigação» (*de-escalating measure* [Tekiroglu et al. 2022]) ou um «instrumento de bloqueio ou desafio» (Latour et al., 2017) que congrega todas as ações de comunicação (Schieb & Preuss, 2016), habitualmente de um para muitos e de muitos para muitos (Benesch et al., 2016), destinadas a, de forma frontal, refutar o *discurso de ódio*, a contrariar e desconstruir desinformação, preconceitos e estereótipos, representações e comportamentos negativos, a apelar e desafiar preocupações e ansiedades, a emancipar vozes e perspetivas menosprezadas e a amplificar narrativas positivas (Article 19, 2015; SELMA, 2019; REACT, 2019). Para tal, a *contranarrativa* recorre a informação detalhada e evidências verosímeis, fontes diversas e credíveis, razões ponderadas e convincentes, (contra-)argumentos verdadeiros e factuais (Chung et al., 2019; Latour et al., 2017; REACT, 2019; Schieb & Preuss, 2016). Ao fazê-lo, a *contranarrativa* não só

permite aos grupos minoritários falarem diretamente (REACT, 2019), como preserva o direito à *liberdade de expressão* e tem potencial «[...] para alterar os pontos de vista dos que odeiam e dos espectadores, encorajando a troca de opiniões e a compreensão mútua, e pode ajudar a desanuviar a conversa» (Chung et al., 2019). Tal justifica o interesse crescente de governos, sociedade civil, empresas privadas e/ou indivíduos para as orquestrarem.

O conceito de *narrativa alternativa* é muito próximo ao de *contranarrativa*, mas representa uma abordagem vagamente distinta (REACT, 2019), porquanto propõe uma versão ou interpretação diferente, centrando-se «[...] naquilo que se defende, mais do que naquilo que se combate» (Latour et al., 2017, p. 81). Em vez de desconstruírem e deslegitimarem outra(s) narrativa(s), direta e abertamente, muitas vezes em reação a acontecimentos isolados, as *narrativas alternativas* trabalham para produzir um enquadramento singular que apoie a mudança pretendida, evidenciando opiniões, visões e perspetivas diversas, positivas, inclusivas, construtivas e progressistas (Latour et al., 2017). Em particular, abordam um tema ausente do discurso corrente e contam uma história diferente da do *discurso de ódio*, discriminatório ou dos *media* dominantes, sobre assuntos como diversidade, migração e interculturalidade (REACT, 2019).

As *narrativas alternativas* visam «[...] ocupar o espaço público, dando espaço e voz a uma posição alternativa (coerente e estruturada) no discurso público» (Latour et al., 2017, p. 82). Logo, enriquecem-no, da mesma maneira que engrandecem o diálogo, compondo a arena pública com mais protagonistas, acontecimentos, pontos de vista e informações, e «[...] dando espaço a processos de emancipação individual e social» (Latour et al., 2017, p. 67). A sua finalidade derradeira é «[...] fomentar o pensamento crítico de toda a gente» (Latour et al., 2017, p. 86) para, assim, enfraquecer ou anular o *discurso de ódio*.

As *contranarrativas* e as *narrativas alternativas* proporcionam respostas adaptadas a cada grupo destinatário e contexto, compreendendo-os e relacionando-se com as suas emoções e necessidades (Latour et al., 2017). As *contranarrativas* procuram prevenir o envolvimento de simpatizantes efetivos e potenciais, bem como apoiar a mudança de perspetivas e de comportamentos de apologistas do *discurso de ódio* (Latour et al., 2017). As *narrativas alternativas*, por sua vez, visam alcançar todas as

peças, incluindo os protagonistas e defensores do *discurso de ódio* para que possam conhecer novas narrativas (Latour et al., 2017). Ambas desempenham um papel importante nas estratégias de emancipação de perpetradores e alvos do *discurso de ódio* (Latour et al. 2017).

Do mesmo modo, as *narrativas*, contrárias ou alternativas, devem ter como premissa fundamental os princípios dos direitos humanos e os valores democráticos, tais como liberdade, igualdade, tolerância e respeito pela diferença. A sua construção obedece a três regras, a saber:

- 1 — Não pode incluir ódio, violência ou discriminação
- 2 — Fomenta a igualdade, o respeito e a solidariedade entre indivíduos e grupos sociais
- 3 — Promove a compreensão da igualdade de dignidade de todos os seres humanos; promove o pensamento crítico, o diálogo justo e transmite informação correta. (Latour et al., 2017, p. 83)

Deste modo, as *contranarrativas* e as *narrativas alternativas* mobilizadas no combate contra o *discurso de ódio* e a violência têm particular utilidade para:

- Bloquear, enfraquecer ou desafiar a autoridade das narrativas, das estruturas sociais e das instituições políticas violentas e de ódio
- Organizar, estruturar e divulgar conhecimento e informação de diferentes modos, desafiando o enquadramento político dominante de exclusão
- Aumentar o consenso sobre legislação específica, políticas inclusivas e desenvolvimentos sociais. (Latour et al., 2017, p. 88)

As *contranarrativas* e as *narrativas alternativas* são apenas um dos instrumentos disponíveis para o combate, não substituindo a participação política, a ação legal e o trabalho no domínio da juventude e da educação dos direitos humanos (Latour et al., 2017).

A campanha «Movimento Contra o Discurso de Ódio — Jovens pelos Direitos Humanos *online*», do Sector da Juventude do Conselho da Europa, criada em 2015 e implementada em vários países, incluindo em Portugal, até 2017. Os projetos «Silence Hate», da italiana COSPE, em parceria com a Federação Europeia dos Jornalistas e o

Media Diversity Institute (MDI), entre 2018 e 2019, «REACT. Respect and Equality: Acting and Communicating Together», liderado pela Associazione Arci de Itália, entre 2017 e 2019, e «Get the Trolls Out», encabeçado pelo MDI, com o apoio de diversos parceiros, desde 2023 e em curso. O movimento [#jagärhär](#) (*euestouaqui*), criado na Suécia, em 2016, e reproduzido em vários países. O projeto #MigraMyths — Desmistificando a Imigração, desenvolvido pela associação Casa do Brasil de Lisboa, desde 2020, a instalação-jogo «Enredo» ou «IN[The Hate Booth]»<sup>24</sup>, o *podcast* «Mitos e Verdades», promovido pela Beira Serra — Associação de Desenvolvimento, desde 2021, as campanhas «O Discurso de Ódio não é Argumento. #Dar a volta ao texto» da Rede Europeia Anti-Pobreza (EAPN) Portugal, iniciada em 2021, #ABCLGBTQIA+ da *Star Life* (antes *Fox life*) Portugal e a Associação ILGA Portugal — Intervenção Lésbica, Gay, Bissexual, Trans e Intersexo, criada em 2022, e #CortarOMalPelaRaiz, desenvolvida no âmbito do projeto kNOWHATE, em curso. Estes são alguns exemplos da aplicação de *contranarrativas* e *narrativas alternativas* para combater e mitigar o *discurso de ódio*, desacreditando ou desconstruindo as narrativas nas quais se baseia, particularmente, em linha, mas também fora dela.

### I. 1.6 Comunicações baseadas na esperança

*As comunicações baseadas na esperança*, tradução livre do inglês *hope-based communications*, são uma estratégia pragmática de comunicação positiva para a mudança, criada recentemente, mas «[...] informada pela história, por especialistas em comunicação, pela neurociência e pela linguística cognitiva» (Open Global Rights, 2019). Foram desenvolvidas como um caso clássico de intra-empendedorismo no seio de uma organização da sociedade civil internacional e, atualmente, são uma metodologia de fonte aberta, disponível gratuitamente e em evolução constante (Coombes, 2019; ICS Centre & JustLabs, 2019). Destinam-se a todos os agentes de mudança social, defensores dos direitos humanos, e de outros movimentos progressistas, em qualquer parte do mundo, que queiram experimentá-la e desenvolvê-la para implementar as suas causas (Hope-based comms, 2019).

---

<sup>24</sup> Vide Costa et al. (2023a, 2023b)

O termo foi articulado pela primeira vez por Thomas Coombes — estratega de comunicação global e, autointitulado, «homem da esperança» («*the hope guy*») —, numa publicação espontânea, num blogue, em 2017, enquanto «[...] um novo modelo para comunicar os direitos humanos através da esperança em vez do medo» (Coombes, 2017). Daí em diante, a metodologia evoluiu para uma lista de verificação informal — inspirada na estratégia de contar histórias «Sete Mudanças» da Greenpeace — ou num quadro de utilização facultativa dentro da Amnistia internacional (Coombes, 2019; ICS Centre & JustLabs, 2019). O interesse gerado junto de outros grupos da sociedade civil em palestras e conferências motivou, não só, o seu desenvolvimento, documentação e publicação, mas também a criação de uma empresa de consultoria estratégica, com a mesma designação (Coombes, 2019; ICS Centre & JustLabs, 2019). A entidade assume-se como «[...] uma comunidade de ativistas [comunicadores da esperança e dos direitos humanos] que querem colocar a comunicação [baseada na esperança] e a campanha [pelos direitos humanos] no centro do trabalho de mudança social», dedicando-se a apoiar a aplicação da abordagem positiva no sector das organizações da sociedade civil e a dotá-lo de um alcance tão eficaz quanto o de políticos e empresas (Coombes, 2019).

*As comunicações baseadas na esperança* partem da premissa de que a esperança é uma estratégia para a mudança e a mudança narrativa é mudança social (Hope-based cooms, 2019; Coombes, 2023). Fundam-se nos valores defendidos assim como na ação construtiva e na visão do mundo que se deseja alcançar e como lá chegar (Coombes, 2019; ICS Centre & JustLabs, 2019). Para tal, propõem ações, tais como mobilizar e colocar a esperança no centro da comunicação, discursos construtivos, isto é, «a favor de», e a ênfase em possíveis soluções, aproximando-se do conceito de *narrativa alternativa* (vide ponto I. 1.4). Caracterizam-se como uma abordagem positiva alternativa ao combate reativo às abordagens e às narrativas populistas, negativas, opressivas e autoritárias contra certas categorias ou grupos, centradas nos problemas e instigadoras do medo — onde se enquadra o *discurso de ódio* — que dominam o debate e os ecossistemas, sobretudo, político e mediático. Propõem fazê-lo através da construção de novas narrativas, autênticas e positivas, que se vivificam com textos, imagens e histórias próprias e que se deseja tornar salientes (Coombes, 2023). Pressupõem assumir o controlo da narrativa — *ser-se a narrativa* —, estabelecer a

agenda, atrair a opinião pública, reenquadrar as conversas e contribuir para a mudança de ideias, atitudes e comportamentos a longo prazo. Implicam o enquadramento claro das mensagens, a mobilização de um vocabulário que, além de factos e argumentos racionais, se baseie em valores universais partilhados — como compaixão, solidariedade e dignidade —, assim como maior consideração das emoções positivas — nomeadamente, empatia, alegria, determinação e solidariedade — e ideias, podendo ser aplicadas a qualquer estratégia e campanha (Coombes, 2019, 2023; Hope-based comms, 2019).

Assim, comunicar os direitos humanos de forma positiva através da esperança implica a realização de cinco mudanças, a saber: (1) «do medo à esperança» — mostrar que é possível criar um mundo melhor; (2) «do contra ao a favor de» — destacar o que se defende e se pretende realizar; (3) «dos problemas às soluções» — expor a alternativa, torná-la desejável e uma prioridade; (4) «da ameaça à oportunidade» — promover a pertença e a participação; e (5) «das vítimas aos heróis» — celebrar a empatia, a solidariedade e a união para construir um sentido de humanidade comum (Hope-based comms, 2019).

As campanhas da Amnistia Internacional «Brasil para todo o mundo», «A longer table» (Estados Unidos da América) e «Messages of Hope» (Nova Zelândia), e os vídeos «Unstoppable», da Planned Parenthood Federation of America, «Marriage Equality: Bring Your Family With You», da HeartWired, e «TV2 — All that we share», do canal de televisão norueguês, exemplificam a utilização de *comunicações baseadas na esperança* (Coombes, 2019; Hope-based comms, 2019; ICS Centre & JustLabs, 2019).

Em síntese, as *comunicações baseadas na esperança* são «[...] uma metodologia simples que, de forma proativa e explícita, promove os valores e as soluções que queremos ver na sociedade» (Coombes, 2019), com enfoque nos direitos humanos, que reforça e coloca em ação, para construir novas narrativas positivas e promover a mudança social.

## Capítulo II. Mapeamento de projetos, ferramentas e recursos

### II. 1. Objetivos e questões de investigação

O presente trabalho nasceu da vontade de contribuir para a compreensão do fenómeno do *discurso de ódio* na atualidade e, com base nela, para o desconstruir. Propunha-se, inicialmente, a mapear os projetos, ferramentas e recursos fundados nas *contranarrativas* e *narrativas alternativas* já implementados, com o objetivo de identificar os pontos-chave comuns e, fundando-se neles, criar um artefacto multimédia inovador, com a finalidade de combater o *discurso de ódio* e promover os direitos humanos, mobilizando a abordagem pragmática proposta pelas *comunicações baseadas na esperança* (vide I. 1.6) centradas na solução, por oposição ao medo e ao problema, e na mudança a longo prazo de ideias, atitudes e comportamentos.

Estimava-se, então, que este trabalho tivesse uma estrutura híbrida — teórica (dissertação) e aplicada (projeto) — e formularam-se três questões de investigação iniciais, especificamente:

Q1. Quais os projetos, ferramentas e recursos fundados em *contranarrativas* e *narrativas alternativas* de combate ao *discurso de ódio* e de promoção dos direitos humanos já implementados?

Q2. Quais os pontos-chave comuns que podem ser utilizados para criar um artefacto multimédia inovador de combate ao *discurso de ódio* e de promoção dos direitos humanos?

Q3. De que forma é possível sistematizá-los e combiná-los para materializar um artefacto multimédia inovador?

Todavia, vários constrangimentos, em particular, profissionais a que se somaram outros do domínio pessoal, motivaram tanto o adiamento quanto a revisão e reformulação de parte da natureza, dos objetivos globais e das perguntas de investigação. Esta reorganização acabou por ser inspirada também pela reflexão resultante da revisão de literatura, aqui materializada no Capítulo I.

Concretizando, este trabalho assume uma natureza essencialmente teórica, reflexiva e crítica sobre o tema e mantém o mapeamento e identificação acima

referidos. Contudo, passa a visar a ideação, com base em ambos os processos, de um paradigma alternativo ao combate direto e reativo ao *discurso de ódio*, fundado na solução e na ação, tendo em vista o seu desenvolvimento e a sua aplicação prática.

Deste modo, as perguntas de investigação passam a ser:

Q1. Quais os projetos, ferramentas e recursos fundados em *contranarrativas* e *narrativas alternativas* de combate ao *discurso de ódio* e de promoção dos direitos humanos já implementados que se destacam?

Q2. Quais os pontos-chave comuns que podem ser utilizados para promover um modelo alternativo ao combate direto e reativo ao *discurso de ódio*?

Q3. De que forma é possível sistematizá-los e combiná-los para lidar com o *discurso de ódio* em linha e fora dela?

Estas questões e objetivos de investigação guiam a análise comparativa, procurando contribuir, igualmente, para uma compreensão mais profunda das estratégias eficazes para promover discursos positivos, inclusivos e construtivos que mitiguem o *discurso de ódio*, nos espaços digitais e comuns ou concretos.

## II. 2. Estratégias de investigação

A revisão de literatura empreendida alicerça a atividade crítica e reflexiva, de identificação, exploração e sistematização dos diversos conceitos envolvidos na pesquisa, materializada nas definições concetuais inscritas no Capítulo I. A revisão de literatura sustenta as perguntas de investigação e a abordagem qualitativa de análise de conteúdo temática utilizadas neste Capítulo II.

A pesquisa foi realizada com recurso ao *SCImago Journal Ranking (SJR)* para literatura científica e ao *Google (Scholar)* para outra «literatura cinzenta», com particular enfoque em relatórios, avaliações e recomendações da União Europeia (UE), da Organização das Nações Unidas (ONU) e de outras entidades, nomeadamente Organizações Não Governamentais e da Sociedade Civil, com trabalho efetivo e reconhecido. O método de «*snowballing*» foi também utilizado.

Neste processo, os conceitos de referência deste trabalho — *discurso de ódio, narrativas, narrativas alternativas, contranarrativas, era digital, comunicações baseadas na esperança* —, assim como os seus sinónimos e termos relacionados ou derivados, por exemplo, *discurso perigoso* e *contradiscorso*, foram explorados, em diferentes combinações e conjugações, considerando igualmente o contexto, as autoras e os autores de referência, os artigos científicos, teses académicas e relatórios técnicos.

O conjunto de métodos utilizados permitiu delimitar e determinar a literatura relevante e credível utilizada na definição concetual e no estado da arte aqui realizados no Capítulo I. Além disso, permitiu igualmente identificar e destacar os projetos, as ferramentas e os recursos de referência para o tema e os objetivos deste trabalho, tanto os já mencionados na primeira parte, incluindo igualmente campanhas e iniciativas, quanto os selecionados para análise comparativa neste Capítulo II.

Daqui este estudo utiliza uma abordagem qualitativa de análise de conteúdo, para examinar de um modo comparativo e sistemático textos-chave centrados em estratégias de combate ao *discurso de ódio*, e ao radicalismo e extremismo, em linha e fora dela, e de promoção dos direitos humanos. Através da análise temática comparativa das obras selecionadas, o presente trabalho identifica pontos-chave comuns empregues no desenvolvimento de *contranarrativas* e de *narrativas alternativas*. Ao sintetizar as perceções e os conhecimentos de diversas fontes, pretende-se fornecer um quadro abrangente, baseado em provas empíricas e enquadramentos teóricos, e contribuir para o debate em curso sobre o combate ao *discurso de ódio* e a promoção de discursos construtivos, nos espaços digitais e comuns.

### II. 3. Projetos, ferramentas e recursos selecionados

Os projetos, ferramentas e recursos foram determinados do esboço teórico desta dissertação e através de seleção manual. Um dos critérios utilizados para compor a amostra foi selecionar os mais citados nas referências bibliográficas, sendo este o caso dos três primeiros aqui listados. Além disso, consideraram-se igualmente o escopo, a reputação e a especialização das entidades promotoras, a amplitude e proximidade

geográfica — em particular, Europa e Portugal —, e a relevância para especificidades temáticas — tais como, plataformas de *media* sociais.

Por conseguinte, o processo culminou na seleção de oito projetos, ferramentas e recursos, a saber:

- 1) «*Countering online hate speech*» («*Combater o discurso de ódio em linha*») — relatório publicado pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), em 2015. O estudo enumera as particularidades distintivas do *discurso de ódio* em linha e apresenta algumas das medidas adotadas para o contrariar e mitigar, destacando boas-práticas locais e globais. Oferece uma análise exaustiva dos quadros normativos internacionais, regionais e nacionais desenvolvidos para lidar com o *discurso de ódio* em linha, salientando os riscos e as limitações envolvidas. Contudo, dá particular ênfase aos mecanismos sociais e da sociedade civil, e não legislativos e dos Estados, exemplificando com diversas iniciativas executadas neste âmbito. Por fim, fornece uma tipologia geral de métodos que podem ser apropriados por vários intervenientes para desenvolver respostas, práticas e adaptadas, ao *discurso de ódio* em linha.
- 2) «*The counter-narrative handbook*» («*Manual de contranarrativas*») — concebido pelo *Institute for Strategic Dialogue (ISD)*, em 2016. Pretende ser um guia para principiantes de apoio à criação, lançamento e avaliação de campanhas de *contranarrativa* eficazes para responder proativamente a conteúdos extremistas violentos em linha. Destina-se, em particular, a agentes da sociedade civil, jovens e/ou ONG. A sua utilização pode ser completada com o conjunto de «ferramentas de contranarrativa» («*Counter-narrative Toolkit*») disponibilizado em linha, gratuitamente, pelo ISD.
- 3) «*Alternativas — Agir contra o discurso de ódio através de contranarrativas*» — versão em língua portuguesa do manual do Conselho da Europa (CoE) editado com o título original «*WE CAN! Taking Action against Hate Speech through Counter and Alternative Narratives*», em 2017. O livro integra o conjunto de

recursos e suportes educativos criados e promovidos pelo «Movimento contra o Discurso de Ódio» («*No Hate Speech*»). A primeira campanha em linha do CoE, iniciada em 2013 e concluída em 2017, e também a mais abrangente em termos internacionais, tendo estado presente em 45 países, incluindo em Portugal, onde envolveu várias entidades e associações, reunidas num Comité Nacional coordenado pelo Instituto Português do Desporto e Juventude (IPDJ). Apresenta-se como «[...] uma ferramenta essencial na promoção da literacia digital e mediática, bem como do pensamento crítico e da capacidade de denunciar e agir contra o discurso de ódio» (Latour et al., 2017, p. 5).

- 4) «*REACT — Educational toolkit for challenging hate speech and constructing counter-narrative*» («*REACT — Conjunto de ferramentas educativas para desafiar o discurso de ódio e construir uma contranarrativa*») — manual produzido, em 2019, no âmbito do projeto «*REACT. Respect and Equality: Acting and Communicating Together*» («*REACT. Respeito e Igualdade: Agir e Comunicar em Conjunto*»), coordenado pela *Associazione Arci*, de Itália. Identifica-se como um guia conciso e prático, incluindo uma série de atividades, para utilização em contextos de aprendizagem (in)formais, com o objetivo de fazer dos jovens agentes ativos, produtores e disseminadores de *contranarrativas* para desafiar o *discurso de ódio*. Além de apresentar a definição de conceitos-chave, incorpora um relatório comparativo das iniciativas, estratégias, aprendizagens e boas-práticas realizadas durante o projeto, entre 2017 e 2019, em Itália, França, Alemanha, Espanha e Reino Unido.
  
- 5) «*Be the narrative — How changing the narrative could revolutionize what it means to do human rights*» («*Seja a narrativa — Como mudar a narrativa pode revolucionar o que significa fazer direitos humanos*») — documento publicado pelo coletivo global JustLabs, em 2019. Narra a história, bem como a análise das aprendizagens e dos resultados, dos laboratórios de experimentação sobre a produção de narrativas promotoras dos direitos humanos, realizados durante um ano, e dos 12 protótipos promissores criados no final. Um processo de incubação de soluções inovadoras e concretas para as narrativas populistas que reuniu

especialistas de diversas áreas do conhecimento e 12 organizações de direitos humanos da Venezuela, Rússia, Turquia, Camboja, Hungria, México, África do Sul, Reino Unido, Filipinas, Austrália, Brasil e Índia. Identifica três respostas possíveis (cultura, cooperação e comunidades) às narrativas opressivas. Procura inspirar uma introspeção profunda sobre a transformação necessária, em termos táticos, organizacionais e operacionais, para produzir narrativas acertadas neste âmbito.

- 6) «*Courage Against Hate*» («*Coragem Contra o Ódio*») — coletânea de artigos promovida por iniciativa do *Facebook*, em 2021. Colige textos académicos e estudos de caso práticos, incluindo a menção a diretrizes e recursos existentes. Objetiva tanto desenvolver um entendimento mais abrangente sobre a natureza do fenómeno quanto incentivar o diálogo multidisciplinar e a ação transversal, a várias áreas e em toda a Europa, para combater o *discurso de ódio* e o extremismo.
- 7) «*Guia Prático | Parte 1: Prevenir e Combater a Radicalização e o Extremismo Online*», e
- 8) «*Guia Prático | Parte 2: Plano de Gestão*» — documentos editados pela Associação Portuguesa de Apoio à Vítima (APAV), em 2021, no âmbito do projeto «*Counter@ct: prevenção e combate à radicação online*», financiado pela União Europeia. Ilustram as fases e resumem a experiência prática adquirida durante o desenvolvimento e a implementação da campanha de narrativa alternativa em linha *#thisismystory*, aplicando o Modelo GAMMMA+<sup>25</sup>.

#### II. 4. Análise comparativa: pontos-chave comuns

Neste ponto, apresentam-se os resultados da análise comparativa, centrando-se nos pontos-chave comuns identificados nos oito projetos, ferramentas e recursos selecionados. As áreas de convergência das diferentes fontes são destacadas. As ideias derivadas são examinadas criticamente e sintetizadas.

---

<sup>25</sup> Sobre o Modelo GAMMMA +, *vide* Ritzmann et al. (2019).

A análise de conteúdo revela 10 tópicos principais no desenvolvimento de *contranarrativas* e de *narrativas alternativas* para desafiar o *discurso de ódio* e os conteúdos extremistas violentos. Estes incluem:

1. **Análise, compreensão e segmentação da(s) audiência(s):** as suas características demográficas, dinâmicas e perspectivas, assim como as suas crenças, valores e preocupações devem determinar tanto a mensagem, ou mensagens, o(s) meio(s) e os mensageiro(s), quanto os conteúdos, estilos e tons, tendo em vista a criação de narrativas à sua medida, relevantes e ressonantes;
2. **Definição de estratégias e objetivos:** específicos, claros, mensuráveis, alcançáveis e, preferencialmente, a longo prazo, para orientar o desenvolvimento da(s) narrativa(s) e garantir o alinhamento com os resultados pretendidos;
3. **Credibilidade e autenticidade:** vozes, histórias e testemunhos na primeira pessoa, genuínas e verosímeis, incluindo de alvos, sobreviventes e, também, de antigos produtores de *discurso de ódio*, e fontes credíveis para fomentar a autoexpressão, a autorrepresentação e a identificação, criar confiança e legitimidade, e aumentar a eficácia da(s) narrativa(s);
4. **Envolvimento emocional:** promover a empatia e a ressonância emocional através da narração de histórias e do apelo a experiências humanas partilhadas para criar uma ligação mais profunda com a(s) audiência(s);
5. **Mensagens positivas e construtivas:** fundadas em valores partilhados, tais como a inclusão e o respeito, nos direitos humanos e em visões alternativas, esperançosas e transformadoras para inspirar ações positivas;
6. **Clareza e simplicidade:** comunicar de forma clara e concisa, com linguagem adaptada, correta e inclusiva, e com enfoque no essencial, para garantir a compreensão e a acessibilidade;
7. **Envolvimento e interação:** incentivar a participação da(s) audiência(s), em particular, das mais jovens, e estimular o diálogo participativo para promover o envolvimento da(s) comunidade(s) e de movimentos sociais (*grassroots movements*);

8. **Elementos visuais e multimédia:** tais como imagens, vídeos, animações, infográficos e *memes*, e formatos inovadores, por exemplo, realidade virtual, para aumentar o apelo, o envolvimento e o impacto narrativo;
9. **Consistência e persistência:** de mensagens e de esforços, em diferentes plataformas e ao longo do tempo, para reforçar as narrativas, alargar audiências, garantir um impacto abrangente e consumir mudanças duradouras a longo-prazo, e
10. **Monitorização e avaliação regulares:** das narrativas, das reações obtidas e da sua eficácia, para as readaptar e aperfeiçoar continuamente em função delas, mas também a tons, tendências, modos, conteúdos, audiências, comportamentos e tecnologias em constante evolução.

Além destes, e apesar de menos frequentes, mas comuns a alguns dos projetos, ferramentas e recursos analisados, e relevantes para o objeto desta dissertação, refiram-se ainda: a análise, compreensão e consideração do(s) contexto(s) e da(s) narrativa(s) opressiva(s); o mapeamento do trabalho já desenvolvido e a sua readaptação ou reutilização; a testagem, a experimentação e a inovação; a cooperação e as sinergias ou a construção de pontes e de uma vasta rede de parcerias; os recursos humanos e materiais, incluindo a utilização de ferramentas gratuitas ou com uma boa relação custo-benefício e a necessidade de planear o orçamento e o recurso a fontes de financiamento diversas; e a combinação imperativa de iniciativas nos espaços digitais e fora deles.

A lista apresentada não é exaustiva, mas oferece um ponto de partida para o estudo de implicações práticas para desenvolver uma compreensão mais abrangente de abordagens eficazes para lidar com o *discurso de ódio* e estimular discursos construtivos e a promoção dos direitos humanos. Por outras palavras, cria uma oportunidade para explorar e idear um modelo alternativo ao combate direto e reativo ao *discurso de ódio* em linha e fora dela, como o delineado no capítulo seguinte.

## Capítulo III. Problemática e contextualização

### III. 1. Contextualização do Projeto

O *discurso de ódio* é um fenómeno antigo, complexo e mutável, que ganhou renovada prevalência e visibilidade, complexidade e legitimação na *era digital*, da hiperinformação e da hiperbolização do real, em especial, decorrente da natureza e das especificidades da internet e, sobretudo, das plataformas de *media* sociais. Mas também resultante da centralidade que ambas ocupam na vida comum e quotidiana (Kemp, 2024a, 2024b; Roberts, 2018), e da utilização negativa, instrumentalizada, armadilhada e muito competente por produtores de ódio, humanos e automatizados (Pukallus & Arthur, 2024; Lee-Won et al., 2020). Mais ainda, consequente de técnicas de moderação, por pessoas e por algoritmos, que registam avanços e recuos, são discutíveis, condicionadas e insuficientes (Gillespie et al., 2020; Lupu et al., 2023; Roberts, 2018, 2019).

Nos últimos cinco anos, o *discurso do ódio* e a polarização por ele nutrida (Vasist et al., 2024) tiveram, e continuam a ter, um estímulo renovado e uma incidência, prevalência, visibilidade e normalização acentuadas, mormente nos espaços digitais e nas plataformas de *media* sociais, mas também nos meios de comunicação social tradicionais. Foram, e são, incentivados pela pandemia de COVID-19, pelo número significativo de conflitos armados registados em todas as latitudes, incluindo no continente europeu, bem como pela galvanização de movimentos e forças de direita radical e extrema, de *neonacionalistas* e ultradireita no mundo, no seio da União Europeia e, também, em Portugal (Almeida et al., 2023; EOOH, 2023; Faloppa et al., 2023; GAPHE, 2023; Matamoros-Fernández & Farkas, 2021; Pérez, 2023).

A tendência de agravamento do *discurso de ódio* e os danos que pode causar, tanto em linha quanto fora dela, colocaram-no, e mantêm-no, no centro das agendas académicas, legais, políticas e mediáticas (Almeida et al., 2023; Carvalho et al., 2024; Montero et al., 2022). Concomitantemente, motivaram, e ocasionam, a produção, cada vez mais vasta e diversa, científico-académica e também noutras esferas e por múltiplos atores (Montero et al., 2022; Paz et al., 2020). Do mesmo modo, impulsionaram, e incentivam, a implementação significativa de ações e estratégias diversas para o

combater e promover os direitos humanos, também nos espaços digitais. Nestas incluindo-se igualmente as recentes estratégias de automatização, com recurso a inteligência artificial, de modelos de linguagem de deteção de *discurso de ódio*, também em português europeu (Ramos et al., 2024), e de *discurso de esperança* (*hope speech* [Chakravarthi, 2022]), assim como de produção de *contradiscursos* (Mathew et al., 2019) e de contranarrativas (Chung et al., 2019), nas plataformas de *media* sociais. Todavia, uma definição ou descrição definitiva de *discurso de ódio* e das suas funções e/ou efeitos continuam por alcançar e o debate permanece, sobretudo, centrado na dicotomia *discurso de ódio/liberdade de expressão* e na lógica *lei-e-ordem* dominante, segundo a qual o *discurso de ódio* deve ser combatido, regulado, proibido e sancionado (Udupa et al., 2021; Udupa & Pohjonen, 2019).

As *contranarrativas* e as *narrativas alternativas* têm sido, e continuam a ser, consideradas estratégias promissoras para desafiar e desmantelar o *discurso de ódio* e as narrativas negativas, problemáticas e opressivas nas quais se baseia, oferecendo uma resposta positiva, inclusiva e construtiva (kNOWHATE, 2024; Latour et al., 2017; REACT, 2019; Silva, 2023; Silva et al., 2024). Contudo, carecem ainda de mais investigação, designadamente, em relação ao seu impacto e à sua eficácia (Carthy, 2021; Carthy et al., 2020; Haide, 2022; Silverman et al., 2016). Ainda que salvaguardada a devida distância e, por isso, a uma escala própria, o mesmo pode aplicar-se às *comunicações baseadas na esperança*, pela sua génese e pelo seu tempo de existência.

Em Portugal, o *discurso de ódio* tem vindo igualmente a escalar (Almeida et al., 2023; Casquillo-Martins et al., 2022; Silva et al. 2021), mas a produção científica e os dados e estatísticas oficiais permanecem escassos (Almeida et al., 2023; Silva et al., 2021). Portugal tem participado em importantes projetos, em parceria e consórcios, e campanhas internacionais, promovidas por instituições governamentais, OSC e ONG contra o *discurso de ódio* (Silva et al., 2021). Todavia, as *contranarrativas* e as *narrativas alternativas* são pouco abundantes e carecem de exploração — bem como, algo dispersas —, e um projeto ou uma iniciativa colaborativa consistente está ainda por implementar (C. Silva, 2023; M. T. Silva, 2019).

Pese embora todos os esforços envidados e progressos alcançados para o identificar, prevenir, mitigar e, sobretudo, combater, o *discurso de ódio* continua a

reproduzir-se, em particular, em linha e nas plataformas de *media* sociais, reafirmando-se como um desafio significativo e uma ameaça crescente aos direitos humanos, à paz e coesão sociais, e à democracia. A disseminação de mensagens de intolerância, preconceito e violência nos espaços digitais alimenta divisões e antagonismos, virtuais e reais, minando a confiança nas instituições e exacerbando conflitos sociais.

Perante estes cenários, a adoção de abordagens inovadoras e proativas de prevenção e neutralização do *discurso de ódio* afigura-se como uma prioridade necessária, senão mesmo imperativa. Ao desafiar as narrativas de divisão e desumanização, estas estratégias poderão constituir-se como um antídoto contra o ódio, capacitando indivíduos e comunidades para a promoção da estabilidade social, da harmonia intercomunitária e da cultura de civismo, no fundo, para o restabelecimento da humanidade comum, para a construção de um presente e de um futuro mais humano e justo para todas as pessoas.

### III. 2. Processo de Ideação

A formação e relacionamento de ideias e conceitos acompanhou todo o trabalho desenvolvido no âmbito desta dissertação, tenho conhecido várias fases. No entanto, o aprofundamento potenciado pela reflexão crítica, pela identificação e análise de projetos, ferramentas e recursos, tanto os referidos quanto os selecionados e comparados, aqui materializadas nos dois primeiros capítulos, acabaram por ser determinantes no processo que aqui se delineia. Isto é: a exploração e ideação de um paradigma alternativo ao combate direto e reativo ao *discurso de ódio*, em linha e fora dela, fundado na solução e na ação, tendo em vista o seu desenvolvimento e a sua aplicação prática.

Num primeiro momento, influenciado pelo objetivo inicial deste trabalho — a criação de um novo artefacto multimédia —, esboçaram-se como possibilidades: a conceção de uma narrativa digital interativa gamificada ou de um jogo interativo; ou a transposição de um movimento criado no *mundo físico* para o digital, por exemplo, adaptando o projeto brasileiro «Mais Amor Por Favor»; ou, ainda, a invenção de um filme documentário em realidade virtual, inspirado no «*Clouds Over Sidra*» (Arora &

Pousman, 2015), enquanto «derradeira máquina da empatia» (*ultimate empathy machine*), uma expressão popularizada por Chris Milk (2015).

Depois, e sem prejuízo da importância que, se antecipa, um objeto transformador poderá ter nesta esfera e em outras conexas, afigurou-se mais apropriada a inspiração de renovação daqueles que se afirmam como o modelo, métodos e recursos dominantes. Dito de outra forma, ao atender ao estado da arte, à centralidade da dicotomia *discurso de ódio/liberdade de expressão*, da lógica *lei-e-ordem* e de *combate* preponderantes, como também, e por oposição, ao examinar as estratégias e as práticas de vanguarda, em especial, os princípios inerentes às *comunicações baseadas na esperança* — alguns dos quais comuns aos das *narrativas alternativas* —, sobressaiu a proposta de alteração de paradigma da adaptação e combinação de vários elementos pré-existentes. Isto significa, idear a operacionalização, em grande escala, das *comunicações baseadas na esperança*, das cinco mudanças nela implicadas — *esperança, a favor de, solução, oportunidade e humanidade* — e das respostas possíveis identificadas — cultura, cooperação e comunidades (Gomez & Coombes, 2019; Hope-based cooms, 2019).

Posto isto, sugere-se descentrar o foco, tanto do debate como da atuação, do *problema* — *discurso de ódio* — e recentrá-lo na *solução* — *discursos positivos, inclusivos e construtivos a favor da diversidade, da igualdade e do respeito* —, colocando os direitos humanos em ação, e estimulando propósito, resiliência, atividade, progresso e, com isso, mudança e paz sociais. Realizá-lo através de uma intervenção preventiva, proativa, integral, interseccional e híbrida — em linha e fora dela —, da base para o topo, do local para o global, dos espaços comuns para os digitais, num esforço coletivo, consciente e contínuo. Além disso, fazê-lo com um incentivo particular à produção e à disseminação de novas narrativas autênticas, positivas, construtivas e inclusivas, imbuídas de valores, emoções e ideias, por cidadãos-utilizadores de comunidades locais.

Para tal, aproveitar e engrandecer os conhecimentos especializados, capacidades, investimento e permanência particulares das pessoas a nível local, como obreiras da *solução*, agentes de mudança social e fazedoras da paz (Pukallus & Arthur, 2024). Reconhecer as comunidades locais, já organizadas em coletivos ou que possam sê-lo, privilegiando aquelas que são visadas, especialmente, pelo *discurso de ódio*,

sobretudo afrodescendentes, migrantes, comunidades muçulmanas, judaicas, ciganas e LGBTI+, bem como, em particular, mulheres (Almeida et al., 2023; Carvalho et al., 2024; Casquilho-Martins et al., 2022; Cohen-Almagor, 2011; Costa, 2021; FRA, 2023; Miranda et al., 2022; Silva et al., 2021; Sousa de Almeida, 2022). Nelas e com elas desenhar e implementar espaços e processos de incubação, de partilha de conhecimentos, de inovação e de cocriação, capacitando-as para *serem a narrativa* (Coombes, 2019, 2023; Gomez & Coombes, 2019). Habilitá-las igualmente para conhecer e reconhecer o *discurso de ódio*, numa prática combinada com a educação para a cidadania e a literacia digitais para, do mesmo modo, apoiar o desenvolvimento de competências cívico-comunicativas (Pukallus & Arthur, 2024). Integrar as organizações locais em consórcios multissetoriais e pluridisciplinares existentes ou por criar, envolvendo especialistas de outras áreas de conhecimento e campos do saber — tais como: comunicação, *media* digitais, tecnologia e dados, marketing, design, arte de contar histórias (*storytelling*), psicologia social, neurociência, artes, comédia e empreendedorismo —, numa relação de benefício mútuo.

Em síntese, criar e dinamizar uma plataforma permanente, composta por todas as partes referidas — comunidades locais, capacitadores e especialistas —, de ideação, capacitação e ação positiva para a mudança através da comunicação de narrativas coletivas de esperança, unidade e solidariedade, inspiradoras e mobilizadoras, incorporando-a em estratégias, planos e programas nacionais existentes e financiando-a através de fundos públicos, privados e comunitários.

Na mesma medida, propõe-se a exploração da possibilidade de mapear, identificar e analisar, nos domínios académicos e além deles, todas as práticas, nas suas diversas formas, fundadas nas *comunicações baseadas na esperança* e, pela sua proximidade, nas *narrativas alternativas*, dos respetivos elementos fundadores, mais-valias e impactos. Assim, tendo em vista a criação de uma coletânea de fonte aberta, incluindo boas-práticas e artefactos pré-existentes, em formatos, que possam ser readaptados e reutilizados, com agilidade e expressão, por uma grande variedade de agentes e em diversos contextos, como seja no âmbito da plataforma aqui esboçada.

Por fim, a abordagem aqui sugerida é interdisciplinar e combinatória, não excludente, até porque a complexidade e grandiosidade do *discurso de ódio* convocam

múltiplos métodos e instrumentos. Por conseguinte, este paradigma alternativo deve ocorrer, em simultâneo, com outros elementos contemporâneos essenciais e a congregação de esforços para, designadamente: estabelecer uma definição ou descrição unívoca de *discurso de ódio*; mapear o fenómeno do *discurso de ódio*; democratizar a literacia para os *media* e para a igualdade; ampliar a colaboração entre os intermediários da Internet e as plataformas de *media* digitais, os governos e as sociedades, e promover mais autorregulação e maior transparência; criar novas formas de moderação de espaços e conteúdos digitais; retificar os algoritmos para deixarem de ampliar, ainda que inadvertidamente, *discurso de ódio* e passarem a promover, de propósito, direitos humanos; e encorajar ou automatizar *contradiscursos* alternativos e positivos, em linha e fora dela.

## Capítulo IV. Considerações finais

A presente dissertação contribui para a compreensão do fenómeno do *discurso de ódio* e para a sua desconstrução na atualidade, apresentando o estado da arte dos principais conceitos e examinando as estratégias implementadas para o combater e promover os direitos humanos. Mais ainda, revela 10 pontos-chave comuns de oito projetos, ferramentas e recursos, fundados em *contranarrativas* e *narrativas alternativas*. Uma lista inicial, não exaustiva, que oferece um ponto de partida ou uma oportunidade para explorar e idear tanto novas narrativas quanto uma abordagem alternativa ao preponderante combate direto e reativo ao *discurso de ódio*, em linha e fora dela. Além disso, da adaptação e combinação de vários elementos pré-existentes, propõe a ideação e operacionalização, em grande escala, da metodologia das *comunicações baseadas na esperança* e dos seus princípios fundadores, visando o seu desenvolvimento e a sua aplicação prática, em particular, nas e com as comunidades locais, dos espaços comuns para os digitais.

Não obstante os contributos realçados, a presente dissertação compreende limitações. Aqui se anotam, tal como se delineiam recomendações tendo em vista a estruturação e desenvolvimento de futuros trabalhos, incluindo aqueles que possam inspirar-se no já realizado nesta dissertação.

A abordagem global adotada pode ser entendida como demasiado abrangente, ainda que, assumidamente, o seja para oferecer uma perspetiva geral do tema e dos conceitos de referência para ser transposta para o particular numa nova investigação. Posto isto, afigura-se relevante que possa ser desenvolvida num novo estudo com um recorte geográfico mais específico, em Portugal, e/ou centrado nas particularidades do contexto sociocultural português, determinando-o para depois o correlacionar e enquadrar na Europa, no *ocidente* e no mundo.

Apesar da amostra utilizada ter permitido a génese de uma lista inicial de pontos-chave comuns de projetos, ferramentas e recursos implementados, ela pode ser considerada escassa, mais especificamente, quando observada a quantidade significativa de obras produzidas, incluindo as que ainda se encontram por documentar enquanto *contranarrativa* ou *narrativa alternativa*, em muito pela ausência de

tipificação nesses termos. Posto isto, salienta-se a importância de, em futuros trabalhos, alargar a amostra de projetos, ferramentas e recursos executados, nomeadamente em Portugal, utilizando outros processos que agilizem o seu mapeamento, recolha e categorização. Do mesmo modo, o método e os critérios utilizados para compor, analisar e sistematizar a amostra podem ser considerados diminutos ou carecendo de afinação. Assim, sugere-se igualmente a aplicação em futuras análises de um modelo de pesquisa quantitativa e qualitativa automatizado e mais preciso, aumentando a sua validade.

Por fim, e como antes referido, as *comunicações baseadas na esperança* estão conceitualmente subdesenvolvidas, por completar e precisar. Portanto, afigura-se pertinente que novas investigações desenvolvam quadros teóricos específicos que as informem, por exemplo, analisando-as à luz da *teoria da esperança* (Snyder et al., 2002) e do *discurso da esperança* (Chakravarthi, 2022), completando-as com entrevistas exploratórias, nomeadamente ao seu *fundador* e à sua comunidade de comunicadores, e/ou com a realização de um estudo de caso que delas derive.

Posto isto, e ainda que dela não se possa extrair uma conclusão geral, esta dissertação dá um justo contributo para o avanço do estudo do *discurso de ódio*, das *contranarrativas*, das *narrativas alternativas* e, de forma inovadora, das *comunicações baseadas na esperança*. Campos que são ainda, relativamente, subexplorados em Portugal. Deste modo, encoraja os estudos futuros a aprofundarem o conhecimento acerca das *comunicações baseadas na esperança* e, pela sua proximidade, a respeito das *narrativas alternativas*, com a intenção de dar uma nova vitalidade à discussão teórica e à produção científica nestas áreas específicas, também, em Portugal.

Ao salientar uma abordagem alternativa fundada na solução e na ação, esta dissertação lança ainda as bases para uma potencial nova linha de investigação. A qual poderá ser conduzida para promover uma compreensão mais precisa sobre a produção e comunicação de *narrativas alternativas* positivas, construtivas e inclusivas, baseadas na *esperança*, entrelaçando a empatia, a solidariedade e a afinidade coletiva, para desnutrir e desativar o *discurso de ódio*, ocupando os seus espaços, em linha e fora dela.

## Referências bibliográficas

- Alkiviadou, N. (2023). The Internet, Internet Intermediaries and Hate Speech: Freedom of Expression in Decline?. *SCRIPTed*, 20, 243. <https://justitia-int.org/wp-content/uploads/2023/02/Article-The-Internet-Internet-Intermediaries-and.pdf>
- Amaral, I. (2019). A imagem na era da cultura digital e das narrativas alternativas. In P. Lopes & B. Reis (Eds.), *Comunicação Digital: media, práticas e consumos* (pp. 99–113). <https://doi.org/10.26619/978-989-8191-87-8.5>
- 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>
- Anderson, L., & Barnes, M. (2022). Hate speech. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition). <https://plato.stanford.edu/entries/hate-speech/>
- APAV (2021). *Guia Prático | Parte 2: Plano de Gestão*. Counter@ct: prevenção e combate à radicação online. União Europeia. [https://apav.pt/apav\\_v3/index.php/pt/2824-guias-praticos-para-a-prevencao-e-combate-a-radicalizacao-online](https://apav.pt/apav_v3/index.php/pt/2824-guias-praticos-para-a-prevencao-e-combate-a-radicalizacao-online)
- APAV (2021). *Guia Prático | Parte 1: Prevenir e Combater a Radicalização e o Extremismo Online*. Counter@ct: prevenção e combate à radicação online. União Europeia. [https://apav.pt/apav\\_v3/index.php/pt/2824-guias-praticos-para-a-prevencao-e-combate-a-radicalizacao-online](https://apav.pt/apav_v3/index.php/pt/2824-guias-praticos-para-a-prevencao-e-combate-a-radicalizacao-online)
- Article 19 (2015). ‘Hate speech’ explained. A toolkit. Londres: Article 2019. <https://www.article19.org/data/files/medialibrary/38231/%27Hate-Speech%27-Explained---A-Toolkit-%282015-Edition%29.pdf>
- Babo, M. A. P. S. (2017). Considerações sobre a máquina narrativa. In A. Peixinho & B. Araújo (Eds.), *Narrativa e Media: géneros, figuras e contextos* (pp. 71–101). [https://doi.org/10.14195/978-989-26-1324-6\\_3](https://doi.org/10.14195/978-989-26-1324-6_3)
- Bahador, B. (2021) Countering hate speech online. In H. Tumber & S. Waisbord (Eds.) *The Routledge Companion to Media Misinformation and Populism*. Routledge. <https://doi.org/10.4324/9781003004431>
- Bakhtiary, M. J., & Behzadi, M. (2023). Digital Storytelling: Unleashing the Power of Narrative in the Digital Age. *Research Gate*.
- Bal, M., & Van Boheemen, C. (2009). *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press. [Google Books](https://books.google.com/books?id=9781442690361).
- Bamberg, M. G., & Wipff, Z. (2020). Reconsidering counter-narratives. <https://doi.org/10.4324/9780429279713-7>
- Bamberg, M. (2007a). Stories: Big or small—Why do we care? In M. Bamberg (Ed.), *Narrative—State of the art* (pp. 165–174). Amsterdam, The Netherlands: John Benjamins. [https://www.researchgate.net/publication/233492031\\_Stories\\_Big\\_or\\_small\\_Why\\_do\\_we\\_care](https://www.researchgate.net/publication/233492031_Stories_Big_or_small_Why_do_we_care)

- Bamberg, M. (2007b). Considering counter-narratives. Clark University.
- Bamberg, M., & Andrews, M. (Eds.). (2004). *Considering counter-narratives: Narrating, resisting, making sense* (Vol. 4). John Benjamins Publishing.
- Barthes, R. (1966) 1977. Introduction to the Structural Analysis of Narratives. *Image Music Text*. New York: Hill and Wang. 79–124.
- Batista, J. (2022, April 8). *Deplatforming o discurso de ódio?!*. Medium. iNOVAMediaLab. <https://medium.com/inovamedialab/deplatforming-o-discurso-de-ódio-86791f8ef077>
- Benesch, S., Dillon, K. P., Ruths, D., Saleem & H. M., Wright, L. (2016). *Counterspeech on Twitter: A Field Study*. Dangerous Speech Project. <http://doi.org/10.15868/socialsector.34066>
- Benesch, S. (2014a). Defining and diminishing hate speech. *State of the world's minorities and indigenous peoples*, 18–25. <https://minorityrights.org/app/uploads/2024/01/mrg-state-of-the-worlds-minorities-2014.pdf>
- Benesch, S. (2014b). Countering dangerous speech: New ideas for genocide prevention. Available at SSRN 3686876. <https://www.ushmm.org/m/pdfs/20140212-benesch-countering-dangerous-speech.pdf>
- Benesch, S. (2012). Dangerous speech: A proposal to prevent group violence. *Voices That Poison: Dangerous Speech Project*.
- Briggs, R., & Feve, S. (2013). Review of programs to counter narratives of violent extremism.
- Brockmeier, J., & Harré, R. (2003). Narrativa: problemas e promessas de um paradigma alternativo. *Psicologia: reflexão e crítica*, 16, 525–535. <https://doi.org/10.1590/S0102-79722003000300011>
- Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18, 297–326. <https://doi.org/10.1177/1468796817709846>
- Brown, A. (2017). What is hate speech? Part 2: Family resemblances. *Law and Philosophy*, 36(5), 561–613. <https://doi.org/10.1007/s10982-017-9300-x>
- Buyse, A. (2014). Words of violence: "Fear speech," or how violent conflict escalation relates to the freedom of expression. *Hum. Rts. Q.*, 36, 779.
- Canavilhas, J., Baccin, A., & Satuf, I. (2017). Era pós-PC: a nova tessitura da narrativa jornalística na web. In A. Peixinho & B. Araújo (Eds.), *Narrativa e Media: géneros, figuras e contextos* (pp. 317–344). [https://doi.org/10.14195/978-989-26-1324-6\\_12](https://doi.org/10.14195/978-989-26-1324-6_12)
- Carthy, S. L., & Sarma, K. M. (2023). Countering terrorist narratives: Assessing the efficacy and mechanisms of change in counter-narrative strategies. *Terrorism and Political Violence*, 35(3), 569–593. <https://doi.org/10.1080/09546553.2021.1962308>

- Carthy, S. (2021). Lessons Learned from Alternative Narrative Campaigns. RAN. [https://home-affairs.ec.europa.eu/system/files/2022-03/ran\\_lessons\\_learned\\_from\\_alternative\\_narrative\\_campaigns\\_032022\\_en\\_1.pdf](https://home-affairs.ec.europa.eu/system/files/2022-03/ran_lessons_learned_from_alternative_narrative_campaigns_032022_en_1.pdf)
- Carthy, S. L., Doody, C. B., Cox, K., O'Hora, D., & Sarma, K. M. (2020). Counter-narratives for the prevention of violent radicalisation: A systematic review of targeted interventions. *Campbell Systematic Reviews*, 16(3), e1106. <https://doi.org/10.1002/cl2.1106>
- Carvalho, P., Caled, D., Silva, C., Batista, F., & Ribeiro, R. (2024). 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>
- Casquilho-Martins, I., Belchior-Rocha, H., & Alves, D. R. (2022). Racial and Ethnic Discrimination in Portugal in Times of Pandemic Crisis. *Social Sciences*, 11(5), 184. <https://doi.org/10.3390/socsci11050184>
- Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., & López, H. M. H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior*, 58, 101608. <https://doi.org/10.1016/j.avb.2021.101608>
- Castells, M. (2007). Communication, power and counter-power in the network society. *International journal of communication*, 1(1), 29.
- Castells, M. (2001). *The Internet galaxy: Reflections on the Internet, business, and society*. New York: Oxford University Press.
- CERN (2023). *A Short Story of the Web*. CERN. <https://home.cern/science/computing/birth-web/short-history-web>
- Cervone, C., Augoustinos, M., & Maass, A. (2021). The Language of Derogation and Hate: Functions, Consequences, and Reappropriation. *Journal of Language and Social Psychology*, 40(1), 80–101. <https://doi.org/10.1177/0261927X20967394>
- Chakravarthi, B. R. (2022). Hope speech detection in YouTube comments. *Social Network Analysis and Mining*, 12(1), 75.
- Chen, A. (2014, November 23). The laborers who keep Dick pics and beheadings out of your Facebook feed. *Wired*. <https://www.wired.com/2014/10/content-moderation/>
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017, February). Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, Portland, Oregon. 1217–123. New York: Association for Computing Machinery. <https://doi.org/10.1145/2998181.2998213>
- Chung, Y. L., Kuzmenko, E., Tekiroglu, S. S., & Guerini, M. (2019). CONAN — Counter Narratives through Nichesourcing: a Multilingual Dataset of Responses to Fight Online Hate Speech. arXiv preprint arXiv:1910.03270

- Citron, D. K., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *BUL Rev.*, 91, 1435. [https://digitalcommons.law.umaryland.edu/fac\\_pubs/1056/](https://digitalcommons.law.umaryland.edu/fac_pubs/1056/)
- Cohen-Almagor, R. (2011). Fighting hate and bigotry on the Internet. *Policy & Internet*, 3(3), 1–26. <https://doi.org/10.2202/1944-2866.1059>
- Coombes, T. (2023, November 10). *Resettlement Narratives: Changing the Conversation on Humanity in Motion. Hope-based.* <https://www.hope-based.com/post/resettlement-narratives-changing-the-conversation-on-humanity-in-motion>
- Coombes, T. (2019, November 11). *Hope-based comms: the story so far.* Hope-based. <https://www.hope-based.com/post/hope-based-comms-the-story-so-far>
- Coombes, T. (2019, November 11). *How to change narratives with hope.* Medium. [https://medium.com/@the\\_hope\\_guy/how-to-change-narratives-with-hope-52f8a15a3b02](https://medium.com/@the_hope_guy/how-to-change-narratives-with-hope-52f8a15a3b02)
- Coombes, T. (2017, December 10). *Hope, not fear: A new model for communication human rights.* Medium. [https://medium.com/@the\\_hope\\_guy/hope-not-fear-a-new-model-for-communicating-human-rights-d98c0d6bf57b](https://medium.com/@the_hope_guy/hope-not-fear-a-new-model-for-communicating-human-rights-d98c0d6bf57b)
- Costa, A. P. (2021). Discurso de ódio e imigração em Portugal. *Lisboa: Casa do Brasil de Lisboa.* <https://casadobrasildelisboa.pt/wp-content/uploads/2022/01/relatorio-migramyths-202111-1.pdf>
- Costa, S., Tavares, M., Bidarra, J., Mendes da Silva, B.: The Enredo game-installation: a proposal to counter hate speech online. In Martins, N. & Brandão, D. (Eds.), *DIGICOM2022* (SSDI, Vol. 27, pp. 307–320). Springer, Cham (2023a). [https://doi.org/10.1007/978-3-031-20364-0\\_27](https://doi.org/10.1007/978-3-031-20364-0_27)
- Costa, S., Tavares, M., Bidarra, J., da Silva, B.M.: IN [The Hate Booth]: a gamified installation to counteract hate speech. In A. L. Brooks (Ed.) *ArtsIT 2022* (LNICS, SITE, Vol. 479, pp. 161–173). Springer, Cham (2023b). [https://doi.org/10.1007/978-3-031-28993-4\\_12](https://doi.org/10.1007/978-3-031-28993-4_12)
- Couldry, N. (2008). Mediatization or mediation? Alternative understandings of the emergent space of digital storytelling. *New Media & Society*, 10(3), 373–391. <https://doi.org/10.1177/1461444808089414>
- Daniels, J. (2008). Race, civil rights, and hate speech in the digital era. [https://academicworks.cuny.edu/gc\\_pubs/193/](https://academicworks.cuny.edu/gc_pubs/193/)
- Davids, N. (2018). On the (in) tolerance of hate speech: does it have legitimacy in a democracy?. *Ethics and Education*, 13(3), 296–308. <https://doi.org/10.1080/17449642.2018.1477036>
- Deuze, M. (2006). Participation, remediation, bricolage: Considering principal components of a digital culture. *The information society*, 22(2), 63–75. <https://doi.org/10.1080/01972240600567170>
- Di Fátima, B., Munoriyarwa, A., Gilliland, A., Msughter, A. E., Vizcaíno-Verdú, A., Gökaliler, E., & de Sousa, V. (2023). Hate Speech on Social Media: A Global

Approach. LabComBooks&EdiPUCE, Covilhã, Portugal.  
<https://doi.org/10.25768/654-916-9>

Di Fátima, B., & Miranda, S. (2022). Discurso de ódio, fake news e redes sociais: uma breve introdução. *Razón y Palabra*, 26(113), 12–16.  
<https://revistarazonypalabra.org/index.php/ryp/article/view/1932>

Di Girolamo, L. (2021). *The emergence of social networks and the proliferation of hate speech online: features and offline consequences*. [Tese de Mestrado, Università degli Studi di Modena e Reggio Emilia]. Sistema Bibliotecario di Ateneo MoReThesis. <https://morethesis.unimore.it/theses/available/etd-03292021-192814/>

ECRI (2016). *ECRI General Policy Recommendation n.15 on combating hate speech*. Estrasburgo: Conselho da Europa. <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.15>

EOOH (2023). Online Hate Speech in 2023. European Observatory of Online Hate. <https://eoooh.eu/articles/online-hate-2023>

Ernst, J., Schmitt, J. B., Rieger, D., Beier, A. K., Vorderer, P., Bente, G., & Roth, H. J. (2017). Hate beneath the counter speech? A qualitative content analysis of user comments on YouTube related to counter speech videos. *Journal for Deradicalization*, (10), 1–49.  
<https://journals.sfu.ca/jd/index.php/jd/article/view/91>

Facebook (2021). Courage Against Hate. Media Diversity Institute. <https://www.media-diversity.org/resource-courage-against-hate/>

Faloppa, F., Gambacorta, A., Odekerken, R., & van der Noordaa, R. (2023). Study on preventing and combating hate speech in times of crisis. Conselho da Europa. <https://rm.coe.int/-study-on-preventing-and-combating-hate-speech-in-times-of-crisis/1680ad393b>

Farrington, N., Hall, L., Kilvington, D., Price, J., & Saeed, A. (2015). *Sport, racism and social media*. Routledge. [Google Books](#).

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*.

Foxman, A. H., & Wolf, C. (2013). *Viral hate: Containing its spread on the Internet*. [Google Books](#).

FRA (2023). Online content moderation — Current challenges in detecting hate speech. Report. European Union Agency for Fundamental Rights. <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>

GAPHE (2023, June 26), *Novo relatório do GPAHE identifica grupos de ódio e de extrema-direita radicais em Portugal; inclui o Chega!*. Global Extremism. <https://globalextrémism.org/post/novo-relatorio-do-projeto-global-contra-o->

[odio-e-o-extremismo-identifica-grupos-de-odio-e-de-extrema-direita-radicais-em-portugal-inclui-o-chega/](#)

- Gelber, K. (2012). Reconceptualizing counterspeech in hate-speech policy (with a focus on Australia). The content and context of hate speech: Rethinking regulation and responses. In M. E. Herz & P. Molnar (Eds.). New York, NY, U.S.A.: Cambridge University Press. 198–216. <https://doi.org/10.1017/CBO9781139042871.016>
- Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A. & Myers West, S. (2020). Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review*, 9(4). <https://doi.org/10.14763/2020.4.1512>
- Glazzard, A. (2017). Losing the plot: Narrative, counter-narrative and violent extremism. <https://icct.nl/sites/default/files/import/publication/ICCT-Glazzard-Losing-the-Plot-May-2017.pdf>
- Gomez, K. & Coombes, T. (2019). *Be the Narrative: How Changing the Narrative Could Revolutionize What It Means to Do Human Rights*. Fund for Global Human Rights.
- Google (2024). Relatório de Transparência da Google. [https://transparencyreport.google.com/?hl=pt\\_PT](https://transparencyreport.google.com/?hl=pt_PT)
- Gulbrandsen, I. T., & Just, S. N. (2013). Collaboratively constructed contradictory accounts: online organizational narratives. *Media, Culture & Society*, 35(5), 565–585. <https://doi.org/10.1177/0163443713485492>
- Haider, H. (2022). Countering Online Misinformation, Hate Speech or Extremist Narratives in the Global South. K4D Helpdesk Report 1113. Brighton, UK: Institute of Development Studies. <https://doi.org/10.19088/K4D.2022.070>
- Hare, I., & Weinstein, J. (Eds.). (2009). *Extreme speech and democracy*. Oxford University Press, USA. [Google Books](#).
- Herz, M., & Molnár, P. (Eds.). (2012). *The content and context of hate speech: rethinking regulation and responses*. Cambridge University Press. [Google Books](#).
- Hope-based comms (2019). *Hope-based comms: a strategy for change*. Hope-based. <https://www.hope-based.com/about>
- ICS Centre & JustLabs (2019). Civil society innovation and populism report in a digital era. International Report. ICS Centre. <https://icscentre.org/wp-content/uploads/2019/10/Innovation-Report-2019-civil-society-innovation-and-populism-in-a-digital-era.pdf>
- Kearney, R. (2012). Narrativa. *Educação & Realidade*, 37, 409–438. <https://www.scielo.br/j/edreal/a/NHRYZHnHWHrR9vDbHxfshTN/abstract/?lang=pt#>
- Kemp, S. (2024a, 31 de janeiro). *Digital 2024: Global Overview Report*. DataReportal. <https://datareportal.com/reports/digital-2024-global-overview-report>
- Kemp, S. (2024b, 23 de fevereiro). *Digital 2024: Portugal*. DataReportal. <https://datareportal.com/reports/digital-2024-portugal>

- Kilvington, D. (2021). The virtual stages of hate: Using Goffman's work to conceptualise the motivations for online hate. *Media, Culture & Society*, 43(2), 256–272. <https://doi.org/10.1177/0163443720972318>
- Kirk, R., & Schill, D. (2021). Sophisticated Hate Stratagems: Unpacking the Era of Distrust. *American Behavioral Scientist*. <https://doi.org/10.1177/00027642211005002>
- kNOwHATE (2024). O que é o contradiscurso?. <https://knowhate.eu/wp-content/uploads/2024/07/FLYER-PT.pdf>
- Koenitz, H., Barbara, J., Holloway-Attaway, L., Nack, F., Eladhari, M. P., & Bakk, A. (2023). INDCOR White Paper 0: Interactive Digital Narratives (IDNs) — A Solution to the Challenge of Representing Complex Issues. <https://doi.org/10.48550/arXiv.2306.17498>
- Lamartine, C., & Silva, M. T. D. (2022). Cyberspace as denunciation: harassment and discrimination linked to coloniality in the project Brasileiras Não Se Calam. *Comunicação e sociedade*, (41), 209–229.
- Latour, A. d., Perger, N., Salaj, R., Tocchi, C., & Otero, P. V. (2017). ALTERNATIVAS: Agir contra o discurso de ódio através de contranarrativas. *Edição revista*.
- Lee-Won, R. J., White, T. N., Song, H., Lee, J. Y., & Smith, M. R. (2020). Source magnification of cyberhate: Affective and cognitive effects of multiple-source hate messages on target group members. *Media Psychology*, 23(5), 603–624. <https://doi.org/10.1080/15213269.2019.1612760>
- Leets, L. (2002). Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of social issues*, 58(2), 341–361. <https://doi.org/10.1111/1540-4560.00264>
- Lepoutre, M. (2019). Hate speech in public discourse. [A Pessimistic Defense of Counterspeech] *Social Theory and Practice*, 43(4), 851–883. <https://ora.ox.ac.uk/objects/uuid:cca6c73d-6283-45b8-92ec-5e2c87a33822>
- Lupu, Y., Sear, R., Velásquez, N., Leahy, R., Restrepo, N. J., Goldberg, B., & Johnson, N. F. (2023). Offline events and online hate. *PLoS one*, 18(1), e0278511. <https://doi.org/10.1371/journal.pone.0278511>
- Maarouf, A., Pröllochs, N., & Feuerriegel, S. (2024). The virality of hate speech on social media. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), 1–22. <https://arxiv.org/pdf/2210.13770>
- Madio, Leonardo and Quinn, Martin, Content moderation and advertising in social media platforms (March 7, 2023). SSRN: <https://ssrn.com/abstract=3551103> ou <https://dx.doi.org/10.2139/ssrn.3551103>
- Manchola-Castillo, C. (2022). Contrahistorias: una propuesta a partir de la bioética narrativa para contrarrestar el discurso de odio. *Revista Colombiana de Bioética*, 17(1). <https://doi.org/10.18270/rcb.v17i1.3948>
- Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & new media*, 22(2), 205–224. <https://doi.org/10.1177/1527476420982230>

- Matamoros-Fernández, A. (2018). Inciting anger through Facebook reactions in Belgium: The use of emoji and related vernacular expressions in racist discourse. *First Monday*, 23(9). <https://doi.org/10.5210/fm.v23i9.9405>
- Matamoros-Fernández, A. (2017). Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, 20(6), 930–946. <https://doi.org/10.1080/1369118X.2017.1293130>
- Mathew, B., Saha, P., Tharad, H., Rajgaria, S., Singhanian, P., Maity, S. K., Goyal, P. & Mukherjee, A. (2019). Thou shalt not hate: Countering online hate speech. In *Proceedings of the international AAAI conference on web and social media* (Vol. 13, pp. 369–380). <https://ojs.aaai.org/index.php/ICWSM/article/view/3237>
- Matsuda M. J., Lawrence C. R. III, Delgado R., Williams Crenshaw K. (1993). *Words that wound*. Westview.
- Meta (2024). Community Standards Enforcement Report. <https://transparency.meta.com/reports/community-standards-enforcement/>
- Milk, C. (2015). *How virtual reality can create the ultimate empathy machine*. TED Talk. [https://www.ted.com/talks/chris\\_milk\\_how\\_virtual\\_reality\\_can\\_create\\_the\\_ultimate\\_empathy\\_machine](https://www.ted.com/talks/chris_milk_how_virtual_reality_can_create_the_ultimate_empathy_machine)
- Miranda, S., Malini, F., Di Fatima, B., & Cruz, J. (2022, April). I love to hate!: the racist hate speech in social media. In *European Conference on Social Media* (Vol. 9, No. 1, pp. 137–145). <https://doi.org/10.34190/ecsm.9.1.311>
- Mirrlees, T. (2021), "GAFAM and Hate Content Moderation: Deplatforming and Deleting the Alt-right", In M. Deflem & D. M. D. Silva (Eds.) *Media and Law: Between Free Speech and Censorship (Sociology of Crime, Law and Deviance*, Vol. 26, pp. 81–97). Emerald Publishing Limited, Leeds. <https://doi.org/10.1108/S1521-613620210000026006>
- Monteiro, A. C. & Carvalhais, M. (2023). An Enactive Approach to the Construction of Meaning in Interactive Digital Narratives. In *Digital Narratives – Theories, Criticism(s), Achievements* (pp. 61–78). <https://doi.org/10.34619/prbo-xb07>
- Montero, A. I., Laforgue-Bullido, N., & Abril-Hervás, D. (2022). Hate speech: a systematic review of scientific production and educational considerations. *Revista Fuentes*, 24(2), 222–233. <https://doi.org/10.12795/revistafuentes.2022.20240>
- Motta, L. G. (2017). Análise pragmática da narrativa: teoria da narrativa como teoria da ação comunicativa. In A. Peixinho & B. Araújo (Eds.), *Narrativa e Media: géneros, figuras e contextos* (pp. 43–69). [https://doi.org/10.14195/978-989-26-1324-6\\_2](https://doi.org/10.14195/978-989-26-1324-6_2)
- Murray, J. H. (2003). *Hamlet no Holodeck. O futuro da narrativa no ciberespaço*. São Paulo: Unesp.
- Nielsen, L. B. (2002). Subtle, pervasive, harmful: Racist and sexist remarks in public as hate speech. *Journal of Social Issues*, 58(2), 265–280. <https://doi.org/10.1111/1540-4560.00260>
- Niemann, M., Riehle, D. M., Brunk, J., & Becker, J. (2019, February). What is abusive language?. In *Multidisciplinary International Symposium on Disinformation in*

*Open Online Media* (pp. 59–73). Springer, Cham. [https://doi.org/10.1007/978-3-030-39627-5\\_6](https://doi.org/10.1007/978-3-030-39627-5_6)

- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016, April). Abusive language detection in online user content. In *Proceedings of the 25th international conference on world wide web* (pp. 145–153).
- Omand, D. “Countering international terrorism: The use of strategy”. *Survival* 47, No. 4 (2005), pp. 107–16.
- Open Global Rights (2019, February 19). *A Guide to Hope-based Communications*. <https://www.openglobalrights.org/hope-guide/>
- O’Regan, C. (2018). Hate speech online: an (intractable) contemporary challenge?. *Current Legal Problems*, 71(1), 403–429. <https://doi.org/10.1093/clp/cuy012>
- Page, R., & Thomas, B. (Eds.). (2011). *New narratives: Stories and storytelling in the digital age*. U of Nebraska Press.
- Parekh, B. (2012). Is There a Case for Banning Hate Speech? In M. Herz & P. Molnar (Eds.), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (pp. 37–56). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139042871.006>
- Parekh, B. (2006). Hate speech. *Public policy research*, 12(4), 213–223. <https://doi.org/10.1111/j.1070-3535.2005.00405.x>
- Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate Speech: A Systematized Review. *SAGE Open*. <https://doi.org/10.1177/2158244020973022>
- Peixinho, A. T., & Araújo, B. (2017). *Narrativa e media: géneros, figuras e contextos*. Imprensa da Universidade de Coimbra/Coimbra University Press. <https://doi.org/10.14195/978-989-26-1324-6>
- Pérez, A. L. (2021). The “hate speech” policies of major platforms during the COVID-19 pandemic. [https://unesdoc.unesco.org/ark:/48223/pf0000377720\\_por.locale=en](https://unesdoc.unesco.org/ark:/48223/pf0000377720_por.locale=en)
- Perkis, A., & Taveter, K. (2022). Applications of complex narratives. *New Review of Hypermedia and Multimedia*, 28(3-4), 97–111. <https://doi.org/10.1080/13614568.2022.2150323>
- Perry, B., & Olsson, P. (2009). Cyberhate: the globalization of hate. *Information & Communications Technology Law*, 18(2), 185–199. <https://doi.org/10.1080/13600830902814984>
- Pizzi, F. (2022). Online hate speech: Education as prevention. *Journal of Inclusive Methodology and Technology in Learning and Teaching*, 2(1). <https://inclusiveteaching.it/index.php/inclusiveteaching/article/view/26>
- Pohjonen, M., & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. *International Journal of Communication*, 11, 19. <https://ijoc.org/index.php/ijoc/article/view/5843>

- Poole, E., Giraud, E. H., & de Quincey, E. (2021). Tactical interventions in online hate speech: The case of #stopIslam. *New Media & Society*, 23(6), 1415–1442. <https://doi.org/10.1177/1461444820903319>
- Pukallus S, Arthur C. Combating Hate Speech on Social Media: Applying Targeted Regulation, Developing Civil-Communicative Skills and Utilising Local Evidence-Based Anti-Hate Speech Interventions. *Journalism and Media*. 2024; 5(2), 467–484. <https://doi.org/10.3390/journalmedia5020031>
- Ramos, G., Batista, F., Ribeiro, R., Fialho, P., Moro, S., Fonseca, A., Guerra, R., Carvalho, P., Marques, C. & Silva, C. (2024). Leveraging Transfer Learning for Hate Speech Detection in Portuguese Social Media Posts. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3430848>
- RAN (2015, October). Counter narratives and alternative narratives [Ran Issue Paper]. [https://home-affairs.ec.europa.eu/system/files/en?file=2020-09/issue\\_paper\\_cn\\_oct2015\\_en.pdf](https://home-affairs.ec.europa.eu/system/files/en?file=2020-09/issue_paper_cn_oct2015_en.pdf)
- REACT (2019). Educational toolkit for challenging hate speech and constructing counter-narrative. arci. <http://www.reactnohate.eu>
- Rettberg, S., & Rettberg, J. W. (2010) Narrative and Digital Media.
- Reynders, D. (2022). Countering illegal hate speech online 7th evaluation of the Code of Conduct. Fact Sheet. <https://commission.europa.eu/system/files/2022-12/Factsheet%20-%207th%20monitoring%20round%20of%20the%20Code%20of%20Conduct.pdf>
- Reynolds, L., & Tuck, H. (2016). The counter-narrative monitoring & evaluation handbook. *Institute for Strategic Dialogue*. <https://www.isdglobal.org/isd-publications/the-counter-narrative-monitoring-evaluation-handbook/>
- Richardson-Self, L. (2018). Woman-Hating: On Misogyny, Sexism, and Hate Speech. *Hypatia, Online First*, 1–17. <https://doi.org/10.1111/hypa.12398>
- Ritzmann, A., Wouterse, L., & Verdegaal, M. (2019). Effective Narratives: Updating the GAMMMA+ model. *Radicalisation Awareness Network*, 19. [https://home-affairs.ec.europa.eu/system/files/2019-12/ran\\_cn\\_academy\\_creating\\_implementing\\_effective\\_campaigns\\_brussels\\_14-15112019\\_en.pdf](https://home-affairs.ec.europa.eu/system/files/2019-12/ran_cn_academy_creating_implementing_effective_campaigns_brussels_14-15112019_en.pdf)
- Roberts, S. T. (2019). *Behind the screen*. Yale University Press. [Google Books](#).
- Roberts, S. T. (2018). Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday*. <https://doi.org/10.5210/fm.v23i3.8283>
- Ryan, M. L. (2017). Narrative. *A companion to critical and cultural theory*, 517–530. <https://doi.org/10.1002/9781118472262.ch33>
- Ryan, M. L. (2007). Toward a definition of narrative. *The Cambridge companion to narrative*, 22.
- Ryan, M. L. (2006). *Avatars of story* (Vol. 17). U of Minnesota Press.

- Ryan, M. L. (2005). *Defining Media from the Perspective of Narratology*. Århus : Handelshøjskolen i Århus, Institut for Sprog og Erhvervskommunikation, 14 s. (VÆRK working paper series)
- Ryan, M. L. (2002). Beyond myth and metaphor: Narrative in digital media. *Poetics Today*, 23(4), 581–609.
- Santana, F. C. (2017). Narrativas em mudança: do folhetim aos textos transmedia. In A. Peixinho & B. Araújo (Eds.), *Narrativa e Media: géneros, figuras e contextos* (pp. 377–407). [https://doi.org/10.14195/978-989-26-1324-6\\_14](https://doi.org/10.14195/978-989-26-1324-6_14)
- Schäfer, S., Rebasso, I., Boyer, M. M., & Planitzer, A. M. (2023). Can We Counteract Hate? Effects of Online Hate Speech and Counter Speech on the Perception of Social Groups. *Communication Research*, 51(5), 553–579. <https://doi.org/10.1177/00936502231201091>
- Schieb, C., & Preuss, M. (2016, June). Governing hate speech by means of counterspeech on Facebook. In *66th ica annual conference, at fukuoka, japan* (pp. 1–23).
- Schlegel, L. (2021). Storytelling against extremism: How fiction could increase the persuasive impact of counter- and alternative narratives in P/CVE. <https://journals.sfu.ca/jd/index.php/jd/article/view/467>
- Schmid, A. P. (2014). Al-Qaeda’s “single narrative” and attempts to develop counter-narratives: The state of knowledge. *The Hague: ICCT*, 26(2), 208–225.
- Schmid, U. K., Kümpel, A. S., & Rieger, D. (2024). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *New media & society*, 26(5), 2614–2632. <https://doi.org/10.1177/146144482210911>
- Sellers, A. (2016). Defining hate speech. *Berkman Klein Center Research Publication*, (2016-20), 16–48. <https://doi.org/10.2139/ssrn.2882244>
- SELMA (2019). *Hacking Online Hate: Building an Evidence Base for Educators*. <https://hackinghate.eu/assets/documents/hacking-online-hate-research-report-1.pdf>
- Siegel, A. A. (2020). Online hate speech. *Social media and democracy: The state of the field, prospects for reform*, 56–88. [Google Books](#).
- Silva, C., Guerra, R., & Carvalho, P. (2024). Desmascarar o discurso de ódio: O conhecimento é a nossa arma. ISBN: 978-989-781-821-9. <https://knowhate.eu/wp-content/uploads/2024/07/BOOKLET-PT.pdf>
- Silva, C., Aguado, J. M., & Dren Gërguri, L. (2023). INDCOR White Paper 5: Addressing Societal Issues in Interactive Digital Narratives. <https://doi.org/10.48550/arXiv.2306.09831>
- Silva, C. (2023). Fighting Against Hate Speech: A Case for Harnessing Interactive Digital Counter-Narratives. In L. Holloway-Attaway & J. T. Murray (Eds.): *ICIDS 2023, LNCS 14383*, pp. 1–16. Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-47655-6\\_10](https://doi.org/10.1007/978-3-031-47655-6_10)
- Silva, M. T., Gonçalves, J., Coelho, P., Brites, M. J. (2021). *Jornalismo, Discurso de Ódio e Participação das Audiências*. Almedina. 2021

- Silva, M. T. (2019). Literacia dos média e discurso de ódio. In M. J. Brites, I. Amaral, & M. T. D. Silva (Eds.), *Literacias cívicas e críticas: Refletir e praticar* (pp. 31–42). Centro de Estudos Comunicação e Sociedade (CECS) – Universidade do Minho. [http://www.lasics.uminho.pt/ojs/index.php/cecs\\_ebooks/issue/view/258/showTOC](http://www.lasics.uminho.pt/ojs/index.php/cecs_ebooks/issue/view/258/showTOC)
- Silverman, T., Stewart, C., Amanullah, Z. and Birdwell, J. (2016). The Impact of Counter-Narratives: Insights from a year-long cross-platform pilot study of counter-narrative curation, targeting, evaluation and impact. [https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives\\_ONLINE\\_1.pdf](https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf)
- Sinnreich, A. (2018). Moderation, Community, and Democracy. Democracy cannot survive algorithmic content moderation. In T. Gillespie, & P. Aufderheide (Eds.), *Internet Policy Review*, 9(4). <https://doi.org/10.14763/2020.4.1512>
- Sinnreich, A. (2018). Four crises in algorithmic governance. In J. C. Joerden, R. Schmücker, & E. Ortland (Eds.), *Annual Review of Law and Ethics* (Vol. 26, pp. 190–199). Duncker & Humblot.
- Snyder, C. R., Rand, K. L., & Sigmon, D. R. (2002). Hope theory: A member of the positive psychology family. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 257–276). Oxford University Press.
- Sousa de Almeida, P. (2022). *Racismo e Xenofobia em Portugal: a normalização dos discursos de ódio no espaço público da internet*. FCT. <https://racismoexenofobia.cria.org.pt/>
- Tekiroglu, S. S., Bonaldi, H., Fanton, M., & Guerini, M. (2022). Using Pre-Trained Language Models for Producing Counter Narratives Against Hate Speech: a Comparative Study. arXiv preprint arXiv:2204.01440
- TikTok (2024). Community Guidelines Enforcement Report. <https://www.tiktok.com/transparency/en-gb/community-guidelines-enforcement-2024-1/>
- Tontodimamma, A., Nissi, E., Sarra, A., & Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126, 157–179. <https://doi.org/10.1007/s11192-020-03737-6>
- Torres, R., Rettberg, J. W., Rettberg, S., & Tabbi, J. (2023). Narrativas Digitais — Teorias, Crítica(s), Resultados: Introdução. *Revista De Comunicação E Linguagens*, (58), 7–11/12. <https://rcl.fcsh.unl.pt/index.php/rcl/article/view/300>
- Tsesis, Alexander. Hate in Cyberspace: Regulating Hate Speech On the Internet. *San Diego Law Review*. Vol. 38 (2001). pp. 817–874. <https://digital.sandiego.edu/sdlr/vol38/iss3/4>
- Tuck, H., & Silverman, T. (2016). The counter-narrative handbook. *Institute for Strategic Dialogue*, 1. <https://www.isdglobal.org/isd-publications/the-counter-narrative-handbook/>

- Turner, G. (2018). The media and democracy in the digital era: is this what we had in mind? *Media International Australia*, 168(1), 3–14. <https://doi.org/10.1177/1329878X18782987>
- Twitter (2023). Removal Requests Report and Rules Enforcement. Transparency Reports. <https://transparency.x.com/en/reports>
- Udupa, S. (2021). Digital Technology and Extreme Speech: Approaches to Counter Online Hate. Commissioned Research Paper for the United Nations Peacekeeping Technology Strategy [https://epub.ub.uni-muenchen.de/77473/1/Digital technology and extreme speech Udupa.pdf](https://epub.ub.uni-muenchen.de/77473/1/Digital%20technology%20and%20extreme%20speech%20Udupa.pdf)
- Udupa, S., Gagliardone, I., & Hervik, P. (2021). Digital Hate: The Global Conjunction of Extreme Speech. Indiana University Press.
- Udupa, S., & Pohjonen, M. (2019). Extreme speech | Extreme speech and global digital cultures — introduction. *International Journal of Communication*, 19. <https://ijoc.org/index.php/ijoc/article/view/9102>
- Ummah, A.H. (2021). Digital Media and Counter Narrative of Radicalism. <http://dx.doi.org/10.21580/teo.2020.31.2.6762>
- Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2021). Deplatformization and the governance of the platform ecosystem. *New Media & Society*. <https://doi.org/10.1177/14614448211045662>
- Van Eerten, J. J., Doosje, B., Konijn, E., de Graaf, B. A., & de Goede, M. (2017). Developing a social media response to radicalization: The role of counter-narratives in prevention of radicalization and de-radicalization. [https://www.researchgate.net/publication/327667910 Developing a social media response to radicalization The role of counter-narratives in prevention of radicalization and de-radicalization](https://www.researchgate.net/publication/327667910_Developing_a_social_media_response_to_radicalization_The_role_of_counter-narratives_in_prevention_of_radicalization_and_de-radicalization)
- Vasist, P. N., Chatterjee, D., & Krishnan, S. (2024). The polarizing impact of political disinformation and hate speech: a cross-country configural narrative. *Information Systems Frontiers*, 26(2), 663–688. <https://doi.org/10.1007/s10796-023-10390-w>
- Waldron, J. (2012). The harm in hate speech. In *The Harm in Hate Speech*. Harvard University Press. [Google Books](#).
- Warner, W., & Hirschberg, J. (2012, June). Detecting hate speech on the world wide web. In *Proceedings of the second workshop on language in social media* (pp. 19–26).
- Weber, I., Laban, A., Masullo, G. M., Gonçalves, J., Torres da Silva, M., & Hofhuis, J. (2020). International perspectives on what's considered hateful or profane online. *JOURNAL OF MEDIA ETHICS*, 3. Center for Media Engagement. <https://mediaengagement.org/research/perspectives-on-online-profanity/>
- Whillock R. K., Slayden D. (Eds.). (1995). *Hate speech*. Sage.