



UNIVERSIDADE FERNANDO PESSOA

Faculdade de Ciências Sociais e Humanas

1º Ciclo em Criminologia

PROJETO DE GRADUAÇÃO

*Discurso de ódio online: Avaliação da prevalência de discursos de ódio
online de carácter misógino contra estudantes universitárias*

Daniela Alexandra Carvalho Rurato

Porto, 2023



UNIVERSIDADE FERNANDO PESSOA

Faculdade de Ciências Sociais e Humanas

1º Ciclo em Criminologia

PROJETO DE GRADUAÇÃO

Discurso de ódio online: Avaliação da prevalência de discursos de ódio online de caráter misógino contra estudantes universitárias

Daniela Alexandra Cavalho Ruato

Projeto de Graduação apresentado à Faculdade de Ciências Sociais e Humanas da Universidade Fernando Pessoa, como parte dos requisitos necessários para a obtenção do Grau de Licenciado do Curso de Criminologia, sob a orientação da Professora Doutora Ana Isabel Sani

Resumo

Atualmente, o discurso de ódio *online* encontra-se extremamente difundido, sendo que estudos realizados em diversos países indicam que entre 42% e 67% da população de jovens adultos observou “mensagens ou discurso *online* odioso e degradante” e 21% terão sido vítimas deste tipo de discurso. Existe, ainda, uma prevalência de 85% de violência digital contra mulheres, abrangendo, tanto mulheres que foram vítimas diretas, como mulheres que testemunharam discursos de ódio contra outras mulheres.

Assim, o presente Projeto visa, não só, compreender a problemática dos discursos de ódio *online*, particularmente de caráter misógino, procurando definir e contextualizar este fenómeno (a nível histórico, social, e do ponto de vista legal), como apresentar algumas estratégias preventivas passíveis de serem adotadas para combater este problema.

Finalmente, será apresentada uma proposta de estudo, cujo objetivo principal é o de avaliar a prevalência de discursos de ódio *online* de cariz misógino, perpetrados contra e/ou testemunhados por estudantes universitárias do sexo feminino ou que se identificam com o género feminino, e compreender o impacto deste tipo de vitimação. Espera-se que a proposta apresentada contribua para a criação de estratégias de prevenção e de combate aos discursos de ódio *online*, nomeadamente de caráter misógino e de propostas de intervenção junto das vítimas.

Palavras-chave: Discurso de ódio *online*; misoginia; mulheres; raparigas; Redes Sociais; violência digital.

Abstract

Nowadays, online hate speech is extremely widespread, now that studies carried out in many countries indicate that between 42% and 67% of the population of young adults has observed “hateful and degrading online messages or speech” and 21% have been victims of this kind of speech. There is, furthermore, a prevalence of 85% of digital violence against women, encompassing both women who have been direct victims, and women who have witnessed hate speech against other women.

Therefore, the present Project aims to not only comprehend the problem of online hate speech, particularly misogynistic hate speech, by attempting to define and contextualize this phenomenon (at an historical and social level, and from the legal standpoint), as well as to introduce some preventive strategies that are apt to be adopted in order to combat this issue.

Finally, a study proposal will be presented, whose main objective is to evaluate the prevalence of misogynistic online hate speech, perpetrated against and/or witnessed by female university students or students who identify with the female gender, and to understand the impact of this type of victimization.

It is expected that this proposal may contribute to the elaboration of preventive and combative strategies against online hate speech, mainly misogynistic, and intervention proposals for the victims.

Keywords: Online hate speech; misogyny; women; girls; Social Networks; digital violence.

Agradecimentos

A realização do presente Projeto de Graduação resulta, por um lado, de três anos de formação académica e, por outro, do apoio e cooperação de várias pessoas que, direta ou indiretamente, contribuíram para eu chegar até aqui. Assim, gostaria de expressar a minha gratidão a todos e todas.

À comunidade da Universidade Fernando Pessoa, particularmente a todos os docentes do curso de Criminologia, pelos conhecimentos transmitidos que me ajudaram a desenvolver as minhas capacidades, tanto a nível profissional, como pessoal.

À Professora Doutora Ana Isabel Sani, pela disponibilidade, pelo auxílio e esclarecimento de dúvidas, e pelos conhecimentos partilhados no âmbito do seu papel enquanto orientadora, que se revelaram essenciais à elaboração deste Projeto, e que contribuíram para o meu crescimento a nível académico e profissional.

Aos meus pais e ao meu irmão, pelo apoio e paciência incondicionais, não só durante a realização deste Projeto, mas também ao longo de todo o meu percurso académico, pelo facto de celebrarem, comigo, as (pequenas) vitórias e de permanecerem ao meu lado durante os (grandes) “infortúnios”.

À minha família, aos que estão comigo e aos que recordo em memória, pelo carinho, pela força, pelo orgulho e pelo apoio absoluto.

Aos meus amigos e amigas, pelos momentos de alegria e descontração. E aos meus colegas de curso, que tive o gosto de conhecer e travar amizade.

Índice

Introdução	1
Parte I - Enquadramento Conceptual	3
1.1 O que é discurso de ódio?	3
1.2 O que caracteriza o discurso de ódio?	3
1.2.1 Background e História	5
1.3 A Criminalização desta política	6
1.3.1 Enquadramento Legal	7
1.3.2 Discurso de Ódio versus Liberdade de Expressão	9
1.4 Discurso de Ódio Online	11
1.4.1 Papel das Redes Sociais	12
1.5 A Importância de Combater o Discurso de Ódio	14
1.5.1 Dados Estatísticos: Incidentes, Ofensas, Vítimas, Ofensores, Localização ..	15
<i>Datasets</i> em Língua Portuguesa.....	16
<i>Datasets</i> na Língua Inglesa.....	18
<i>Datasets</i> Multilíngues.....	20
1.5.2 Prevenção, Intervenção e Estratégias	22
1.6 Discursos de ódio <i>online</i> de carácter misógino	28
Parte II - Proposta de Investigação	34
2.1. Introdução	34
2.2. Objetivos	34
2.2.1. Objetivo Geral	34
2.2.2. Objetivos Específicos.....	34
2.3. Método	34
2.3.1. População Alvo e Amostra	35
2.3.2. Instrumento de Recolha de Dados	36
2.3.3. Procedimentos	36
2.4. Resultados Esperados	37
2.5. Considerações Finais	38
Referências	39
Webgrafia	42
Anexos	44

Índice de Figuras

FIGURA 1 – HIERARQUIA DE CLASSES	17
--	----

Índice de Anexos

ANEXO 1 - QUESTIONÁRIO - “DISCURSO DE ÓDIO ONLINE: AVALIAÇÃO DA PREVALÊNCIA DE DISCURSOS DE ÓDIO ONLINE DE CARÁTER MISÓGINO CONTRA ESTUDANTES UNIVERSITÁRIAS”	45
ANEXO 2 - PEDIDO DE AUTORIZAÇÃO À COMISSÃO DE ÉTICA DA UNIVERSIDADE FERNANDO PESSOA	50
ANEXO 3 - DECLARAÇÃO DE CONSENTIMENTO INFORMADO.....	51

Siglas

APAV – Associação Portuguesa de Apoio à Vítima

CE – Comissão Europeia

CE- Conselho da Europa

CPP – Código Penal Português

DM – Mensagem Direta

EAPN – Rede Europeia Anti-Pobreza

ECRI – Comissão Europeia contra o Racismo e a Intolerância

EE – *Everything Everywhere*

EIGE – Instituto Europeu para a Igualdade de Género

IPDJ – Instituto Português do Desporto e Juventude

LGBTQIA+ - Lésbicas, Gays, Bissexuais, Transgénero, *Queer*, Intersexo, Assexuais e todas as outras orientações sexuais e identidades de género em existência

ODIHR – Office for Democratic Institutions and Human Rights (Gabinete para as Instituições Democráticas e os Direitos Humanos)

OMS – Organização Mundial de Saúde

ONU – Organização das Nações Unidas

PIDCP – Pacto Internacional dos Direitos Civis e Políticos

SRS – Sites de Redes Sociais

TFUE – Tratado sobre o Funcionamento da União Europeia

TIC – Tecnologia de Informação e Comunicação

ToLD-Br – *Toxic Language Dataset for Brazilian Portuguese*

UE – União Europeia

UEFA – União das Associações Europeias de Futebol

UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura

UNFPA – Fundo de População das Nações Unidas

VCMR – Violência contra as Mulheres e as Raparigas

WMC – *Women's Media Center*

Introdução

A proliferação do discurso de ódio *online*, intimamente ligado ao (mau) uso das redes sociais, tem constituído um dos maiores desafios para a sociedade atual, provocando efeitos negativos no bem-estar, tanto da vítima, como de possíveis observadores, incluindo “depressão, isolamento, paranoia, ansiedade social, insegurança, desapontamento, solidão e falta de confiança” (Selma, 2019).

De facto, atualmente, o discurso de ódio *online* encontra-se extremamente difundido, sendo que estudos realizados em diversos países indicam que entre 42% e 67% da população de jovens adultos observou “mensagens ou discurso *online* odioso e degradante” e 21% terão sido vítimas deste tipo de discurso (Walther, 2022). Existe, ainda, uma prevalência de 85% de violência digital contra mulheres, abrangendo, tanto mulheres que foram vítimas diretas, como mulheres que testemunharam discursos de ódio contra outras mulheres.

Poderá ser argumentado que as pessoas têm o direito de se expressarem de forma livre, pelo que este fenómeno não se afigura assim tão preocupante. Porém, apesar de a liberdade de expressão estar consagrada como um dos principais Direitos Humanos, a prática deste direito, implica diversas responsabilidades e deveres – o discurso de ódio não pode ser defendido, nem justificado, principalmente, quando se verifica que este tipo de retóricas podem incentivar crimes de ódio reais, que continuam a extinguir e a destruir ativamente vidas (Keen e Georgescu, 2016).

Atualmente, o discurso de ódio é uma das demonstrações mais visíveis de intolerância. Não só se verifica um aumento no uso de afirmações de carácter odioso nos discursos políticos, como também uma aceitação e normalização deste tipo de narrativas em espaços-comuns do domínio público, designadamente na Internet. Este facto é extremamente preocupante, visto que, quando o intolerável começa a ser o “normal”, está-se perante uma ameaça aos Direitos Humanos (Keen e Georgescu, 2016).

As qualidades da Internet, enquanto ferramenta informativa, de comunicação, trabalho, entretenimento e até como mecanismo de organização da mudança social, são inegáveis, contudo, não deve ser usufruída com o propósito de disseminar ideologias de ódio (Keen e Georgescu, 2016).

A tomada de iniciativas acerca do discurso de ódio, afigura-se essencial, dado o elevado número de narrativas abusivas que se encontram na Internet, muitas das quais contendo linguagem extremista e de carácter misógino, sexista, racista, xenófobo, homofóbico, entre outras.

Não obstante, o discurso de ódio *online* é, igualmente, um problema que advém da forma como as pessoas fazem uso da Internet, o que torna este fenómeno de mais difícil compreensão, visto que, “o discurso de ódio representa as folhas de uma planta particularmente malvada cujas origens estão, profundamente, arraigadas na sociedade. Cortar as folhas não pode resolver as causas maiores do problema.” (Keen e Georgescu, 2016).

Deste modo, o presente trabalho é constituído por duas partes que se complementam: na primeira parte, faz-se um enquadramento teórico da problemática do discurso de ódio *online*, procurando definir este conceito, assim como contextualizá-lo, tanto a nível histórico e social, como do ponto de vista legal e, finalmente, apresentam-se algumas estratégias preventivas passíveis de serem adotadas para combater este problema. Por outro lado, na segunda parte, apresenta-se uma proposta de estudo, composta pela enumeração dos objetivos, métodos, participantes, instrumento e procedimentos utilizados, assim como os resultados esperados através da implementação da proposta sugerida.

Parte I - Enquadramento Conceptual

1.1 O que é discurso de ódio?

Não existe, atualmente, um consenso sobre a definição de discurso de ódio, contudo, para os propósitos do presente projeto, adotar-se-á a proposta do Conselho da Europa, que define discurso de ódio como “todas as formas de expressão que incitam, promovem, espalham ou justificam violência, ódio ou discriminação contra uma pessoa ou um grupo de pessoas, ou que os denigre, por razão de características ou estatutos pessoais, reais ou atribuídas, tais como a ‘raça’, cor, língua, religião, nacionalidade, origem nacional ou étnica, idade, incapacidade, sexo, identidade de género ou orientação sexual” (Website Conselho da Europa).

1.2 O que caracteriza o discurso de ódio?

Caracterizar o discurso de ódio, afigura-se complicado, não só devido à ambiguidade das definições existentes, que deixam bastante espaço para interpretações pessoais e subjetivas, como, também, pelo facto de que determinada afirmação pode ou não ser considerada discurso de ódio, dependendo das sensibilidades pessoais e *background* cultural de cada pessoa, podendo até a forma como o comentário está redigido, desempenhar, de igual modo, um papel importante na perceção de existência de discurso de ódio (Ross *et al.*, 2016).

Contudo, é possível afirmar que, por norma, o discurso de ódio consiste numa forma de expressão extrema de ódio e/ou incitação à violência contra terceiros, que envolve o uso de linguagem pejorativa e atribuição de “rótulos” ofensivos, abusivos e assediantes, dirigidos, tipicamente, a grupos de minorias (Cervone *et al.*, 2021), por razão da sua raça ou origem étnica, género, orientação sexual, religião, idade, incapacidade ou condição física (Oliveira, 2022).

O discurso de ódio tem, assim, o poder de incitar, promover e/ou justificar ideologias de ódio e violência, atingindo um imenso número de pessoas através da sua divulgação, feita maioritariamente através do uso das redes sociais (Cervone *et al.*, 2021).

Este tipo de discurso tem como principais objetivos, a incitação ao preconceito e a criação de um ambiente hostil, discriminatório e intolerante contra um grupo historicamente oprimido (Oliveira, 2022), favorecendo somente um lado do assunto em debate

(associação política, opiniões relativas a debates raciais, religiosos ou acerca de orientação sexual, por exemplo), enquanto menospreza o lado oposto – assim, estamos perante uma apresentação bipolar de dois extremos (totalmente a favor *versus* totalmente contra) (Erjavec e Poler, 2012).

Nesse sentido, Susan Benesch (2014), propôs um conjunto de parâmetros que auxiliam na identificação de “discursos perigosos” - *The Dangerous Speech Guidelines* -, avaliando o grau de perigosidade da narrativa de acordo com cinco fatores: a) a popularidade e influência que o interlocutor exerce sob a audiência; b) o estado emocional da audiência; c) o conteúdo do discurso, que pode ser percebido como uma incitação ao uso de violência; d) o contexto histórico e social em que o discurso ocorre; e) os meios usados para divulgar a mensagem.

Benesch identifica, ainda, padrões retóricos que indicam a existência de “discurso perigoso”, servindo como sinais que possibilitam a sua deteção, a saber: referência ao grupo-alvo como “pestes”, “insetos” ou “animais”; alegação de que o grupo-alvo representa uma ameaça existencial ou, até, mortal para a audiência; acusação de que o grupo-alvo será prejudicial para a integridade e pureza da audiência; identificação do grupo-alvo como um “*alien*”, reforçando que este não pertence ao resto da audiência (Benesch, 2014).

É importante referir, contudo, que, de acordo com Fortuna *et al.* (2019), os discursos de ódio “podem ocorrer em diferentes estilos linguísticos, até em formas subtis ou quando o humor é utilizado”.

Em suma, o fenómeno do discurso de ódio exige a devida atenção e intervenção, visto que:

- Promove, incita e procura justificar o ódio, a discriminação e a utilização de violência;
- É exercido contra um alvo ou grupo-alvo específico;
- Depende do recurso a um conjunto de instrumentos e argumentos retóricos (nomeadamente a evocação de teorias da conspiração) (Oliveira, 2022).

1.2.1 Background e História

O aparecimento do discurso de ódio *online* é simultâneo ao da Internet. É verdade que a Internet facilita o acesso a um maior número de informação, oferecendo a possibilidade de comunicação com praticamente qualquer pessoa no mundo, entre muitos outros benefícios, porém, também oferece um “palco” para que qualquer pessoa se transforme num orador público, facto que, infelizmente, muitos aproveitam para disseminar as suas mensagens de ódio, intolerância e discriminação (Keen e Georgescu, 2016).

De facto, tal como o escritor e filósofo italiano, Umberto Eco, menciona, “agora todos os que habitam o planeta, incluindo os loucos e os idiotas, têm direito à palavra pública. Hoje, na Internet, a sua mensagem tem a mesma autoridade que o prémio Nobel e o jornalista rigoroso” (Prieto, 2015).

Relativamente ao que poderá originar a realização de um discurso de ódio, alguns estudos identificam variáveis relativas à componente ideológica (Cervone *et al.*, 2021), como por exemplo o aumento da adesão aos partidos de extrema-direita que se tem verificado nos últimos anos, resultante da disseminação de teorias de conspiração, que alegam que os ideais liberais e componente multiculturalista do nosso país, ameaçam destruir os valores culturais e tradicionais da nação. Fenómenos como a total negação do Aquecimento Global ou os Movimentos Anti-Vacinação, são alguns exemplos desta influência – a propagação de desinformação ou informação em quantidade excessiva, origina um ambiente propício à disseminação de discursos de ódio, ao favorecer a criação de informação fraudulenta e mitos (Braga, 2018).

Por seu lado, outros estudos conseguiram identificar fatores psicológicos observados anteriormente à participação em discursos de ódio e que poderão, de certa forma, influenciá-los, estando estes fatores relacionados com: a) a personalidade autoritária e de direita do interlocutor, assim como a sua motivação para expressar os seus preconceitos e o domínio que exerce, a nível social; b) o seu estado emocional e afetivo, dirigido especificamente ao alvo ou grupo-alvo, tais como o ódio, a repugnância, o desprezo e os sentimentos de ameaça intergrupala; e c) motivações gerais, tais como *thrill-seeking*, ou “procura de emoção” e manutenção de auto-estima (Cervone *et al.*, 2021).

Alguns estudos foram ainda mais longe, distinguindo quatro tipos de ofensores *online* de acordo com a sua personalidade, a saber: *soldiers* (“soldados”), que agem de modo

organizado; e *believers* (“crentes”), *players* (“jogadores”) e *watchdogs* (“cães de vigia”), que agem de forma individual, segundo a sua própria iniciativa (Erjavec e Poler, 2012).

Finalmente, existem ainda estudos que indicam que certos eventos *offline*, tais como a realização de eleições ou protestos, podem despoletar a difusão de discursos de ódio *online* (Lupu *et al.*, 2023).

As consequências dos discursos de ódio, por outro lado, podem ser “menores” (por ex: autoimagem negativa ou isolamento e exclusão social do alvo) ou mais gravosas, como é o caso da incitação à violência ou concretização da mesma, sendo que o genocídio de 1994 do Ruanda representa um dos mais claros exemplos das atrocidades que podem advir da difusão dos discursos de ódio (Fortuna *et al.*, 2019).

Efetivamente, Schabas (2000), afirma que os discursos de ódio disseminados na Imprensa e Rádio do Ruanda - aquando do começo da guerra civil em outubro de 1990 e intensificados pela morte do então Presidente Juvénal Habyarimana a 6 de abril de 1994 -, incitaram os “executores voluntários” do genocídio do Ruanda, adicionando que devia, e podia, ter sido tomada ação mais cedo para prevenir esta tragédia, criminalizada sob a Convenção para a Prevenção e a Repressão do Crime de Genocídio. Contudo, o autor adianta que, apesar da Convenção criminalizar a incitação ao genocídio, não abrange a “propaganda do ódio” – um primeiro e importante passo na “cadeia alimentar do genocídio”, que foi posteriormente criminalizado através de outros instrumentos da Lei dos Direitos Humanos – e foca-se, principalmente, na punição (ou seja, na tomada de ação após o crime ser efetivamente concretizado) – facto deveras lamentável, visto que “O genocídio é preparado com propaganda, um bombardeamento de mentiras e ódio dirigido ao grupo-alvo, e com o objetivo de preparar os ‘executores voluntários’ para a tarefa atroz que lhes vai ser solicitada. Aqui, então, está a chave para prevenir genocídio.” (Schabas, 2000).

1.3 A Criminalização desta política

O conceito de crime de ódio é uma construção legal, que tem por base o Direito Criminal, referindo-se não ao sentimento de ódio, mas sim ao preconceito (Chakraborti, 2014). De facto, tal como Chakraborti (2014) refere, os crimes de ódio representam um grande problema face aos Direitos Humanos, indo muito além “das soluções da Justiça Criminal e a culpabilidade de ofensores individuais”.

Nesse sentido, Mason (2014) argumenta que existe pouca concordância relativamente à possibilidade do Direito Criminal ser um mecanismo adequado e eficiente para a transformação de uma sociedade injusta, numa sociedade íntegra.

Perante crimes motivados por sentimentos de ódio, estarão envolvidos sentimentos de empatia ou compaixão que desempenham um papel fundamental nos juízos éticos que se afiguram necessários para dar resposta às “chamadas pela justiça social e Direitos Humanos” (Mason, 2014).

Assim, Mason (2014) defende que, independentemente de se acreditar, ou não, que as emoções servem como “guias éticos” fiáveis, estas são partes integrantes da função normativa da Lei e, através da exploração desta componente emocional, tem-se a oportunidade de refletir acerca dos desafios que advêm do uso de sistemas de criminalização e punição para defender, moralmente, os Direitos Humanos de grupos estigmatizados e subjugados.

1.3.1 Enquadramento Legal

O discurso de ódio, apesar de não estar ainda criminalizado, é reconhecido como um “domínio da criminalidade” a nível internacional (Comissão Europeia, 2021).

Em 1997, o Conselho da Europa adotou a Recomendação nº R (97) 20 do Comité de Ministros aos Estados-Membros sobre “discurso de ódio”, que considerou que o discurso de ódio consagra incitamento ao ódio contra determinadas pessoas, ou grupos de pessoas, cujas características estão protegidas por esta mesma recomendação (desde o “ódio racial, à xenofobia, ao antissemitismo ou outras formas de ódio baseadas na intolerância, incluindo: a intolerância expressa por nacionalismo agressivo e etnocentrismo, discriminação e hostilidade contra minorias, migrantes e pessoas descendentes de migrantes”) (Conselho da Europa, 1997).

Já em 2015, a Comissão Europeia contra o Racismo e a Intolerância (ECRI) apresentou a Recomendação de Política Geral nº 15 acerca do “discurso de ódio”, definindo o discurso de ódio como a “defesa, promoção ou incitamento, sob qualquer forma, à difamação, ao ódio ou ao vilipêndio de uma pessoa ou grupo de pessoas, bem como qualquer assédio, insulto, estereótipo negativo, estigmatização ou ameaça contra essa pessoa ou grupo de pessoas e justificação de todos os tipos de expressão anteriores, por motivos de raça, cor, ascendência, origem nacional ou étnica, idade, deficiência, língua,

religião ou crença, sexo, género, identidade de género, orientação sexual e outras características pessoais ou estatuto” (ECRI, 2017).

Finalmente, em 2016, a União Europeia assinou um acordo com o Facebook, Microsoft, Twitter e Youtube – o Código de Conduta para a luta contra o discurso ilegal de ódio *online*. Mais tarde, o Instagram, Snapchat, Dailymotion, Jeuxvideo.com, TikTok, LinkedIn, Rakuten Viber e Twitch juntaram-se ao acordo, com o intuito de combater e prevenir a expansão do discurso de ódio *online* (European Commission Website).

Destaca-se, ainda, o Artigo 240º do Código Penal Português (CPP), que penaliza: 1) o envolvimento em “atividades de propaganda organizada que incitem à discriminação, ao ódio ou à violência” contra uma pessoa ou grupo(s) de pessoas que partilham uma característica protegida; e 2) a divulgação pública, através de qualquer meio, da “apologia, negação ou banalização grosseira de crimes de genocídio, guerra ou contra a paz e a humanidade”, difamando e/ou injuriando, ameaçando, incitando ao ódio ou provocando atos de violência contra uma pessoa ou grupo(s) de pessoas em razão de uma característica protegida (Artº. 240º do CPP).

Não obstante a ausência de uma definição jurídica de “discurso de ódio” no Direito da UE, a Decisão-Quadro 2008/913/JAI do Conselho da União Europeia, de 28 de novembro de 2008, estabelece que quem proceder à “incitação pública à violência ou ao ódio dirigido contra um grupo de pessoas ou um membro de um desses grupos, definido com base na raça, cor da pele, ascendência, religião ou crença religiosa ou origem nacional ou étnica”, estará a cometer um ato considerado punível enquanto infração penal (EUR-Lex – Access to European Union Law).

Tal como referido anteriormente, a principal motivação para difundir discursos de ódio é o preconceito contra pessoas que aparentam pertencer, apoiar, associar-se ou ter uma relação de filiação com determinado grupo, ou comunidade, que partilha uma característica protegida, representando atos de “mensagem ou identidade” dado que visam transmitir mensagens à vítima e ao grupo ou comunidade (designadamente, promover o isolamento das vítimas relativamente ao resto da sociedade) (ODIHR).

Assim, atendendo às especificidades do discurso de ódio – motivado pelo preconceito, transforma-se em ódio, resultando na ameaça e menosprezo da dignidade de uma pessoa, ou de um grupo de pessoas que partilham (ou aparentam partilhar) características

protegidas -, o discurso de ódio pode ser entendido como correspondente a um “domínio da criminalidade”, com base no Artigo 83º, nº1 do Tratado sobre o Funcionamento da União Europeia (EUR-Lex – Access to European Union Law - TFUE).

O objetivo basilar da criminalização do discurso de ódio será, então, o de combater esse ódio que o caracteriza como um “domínio da criminalidade”. Nesse sentido, a Comissão Europeia sugeriu o aditamento do discurso de ódio como um “domínio da criminalidade” para que seja possível propor, posteriormente, “direito derivado que abordasse especificamente a evolução e os desafios sociais conexos à medida que estes surjam e evoluam no futuro” (Comissão Europeia, 2021).

Em 1919, no seguimento do caso *Schenck versus United States*, o Juíz Oliver Wendell Holmes Jr. declarou que a perceção do risco que poderá advir de determinado discurso depende da proximidade e do grau, sendo que a questão central será a de perceber se as palavras usadas, poderão criar um perigo claro, devido à sua natureza ou às circunstâncias em que ocorre, que o Congresso tem o dever de prevenir.

Holmes acrescenta, ainda “Cada ideia é uma incitação. Ela oferece-se para ser acreditada e, se acreditada, é posta em prática, exceto se alguma outra crença a superar ou alguma falha de energia sufocar o movimento à nascença. A única diferença entre a expressão de uma opinião e um incitamento é o entusiasmo do locutor pelo resultado. A eloquência pode incendiar a razão.” (Vile, J.).

1.3.2 Discurso de Ódio versus Liberdade de Expressão

A liberdade de expressão e de opinião é um dos pilares dos Direitos Humanos e das sociedades democráticas (United Nations Website), estando protegida ao abrigo do Artigo 19º da Declaração Universal dos Direitos Humanos e do Artigo 19º do Pacto Internacional dos Direitos Cívicos e Políticos – ambos instrumentos que constituem a Carta Internacional de Direitos Humanos.

De facto, a liberdade de expressão permite a existência de “debates de interesse público vibrantes e multifacetados, dando voz a perspetivas e pontos de vista diferentes”, pelo que o respeito desta liberdade é essencial para o desenvolvimento humano sustentável e para a promoção, a nível internacional, da paz e segurança (Rabat Plan of Action, 2012).

Deste modo, o dever de preservar a liberdade de expressão, é, frequentemente, reiterado quando são tomadas medidas para regular a expressão de ódio, nomeadamente *online*, uma vez que a utilização da Internet e, em especial, das Redes Sociais acentuou a necessidade individual de procurar, receber e divulgar opiniões e informação (United Nations Website).

Infelizmente, o exercício deste Direito gera, em várias ocasiões, diversas formas de discriminação, violência e/ou hostilidade contra indivíduos, ou grupos, por razão da uma característica protegida (Rabat Plan of Action, 2012).

Nesse sentido, as Nações Unidas garantem o respeito pela liberdade de expressão como a norma, adiantando que “qualquer restrição deve ser uma exceção e procurar prevenir danos e garantir a igualdade ou a participação pública de todos(as)” (United Nations Website).

Assim, a ONU estipula, ao abrigo do Artigo 20º do Pacto Internacional dos Direitos Civis e Políticos, (PIDCP) que qualquer propaganda para a guerra e/ou advocacia de ódio nacional, racial ou religioso deve ser proibida por lei, assim como propõe o *UN Rabat Plan of Action*, que proporciona alguma orientação sobre o que distingue a liberdade de expressão da incitação à discriminação, hostilidade e violência.

O *Rabat Plan of Action* apresentou, então, um “Teste de Limiar de Seis-Partes” que procura “delimitar” a definição de restrições à liberdade de expressão e à incitação ao ódio, e a aplicação do Artigo 20º do Pacto Internacional dos Direitos Civis e Políticos, tendo em consideração:

1. Contexto: O discurso deve ser feito e divulgado dentro de um contexto social e político significativo;
2. Locutor(a): O status/posição social do(a) locutor(a) deve ser proeminente, especialmente em relação com o público-alvo;
3. Intenção: A negligência – tal como a circulação de conteúdo ofensivo - não é suficiente para que um ato seja considerado uma “incitação à discriminação, hostilidade e violência” de acordo com o Artigo 20º do PIDCP, pelo que se deverá dar particular atenção à “relação triangular entre o objeto, o sujeito do discurso e a audiência”;

4. Conteúdo e Forma: Deve ser analisado o grau em que o discurso foi direto e provocativo, assim como o estilo, a forma, a natureza dos argumentos utilizados e o equilíbrio entre estes;
5. Extensão do Ato de Fala: Relacionada com a magnitude, alcance e natureza pública do discurso e com o tamanho da audiência, atentando aos meios de divulgação usados, a frequência, quantidade e extensão dos discursos, bem como ao facto de a audiência possuir, ou não, meios para agir em concordância com o incentivo e à reduzida ou vasta acessibilidade do discurso ao público em geral;
6. Probabilidade, incluindo Iminência: A incitação é, por definição, simplesmente um crime de conduta, pelo que se deverá atentar ao grau de risco de dano identificado. Assim, é necessário que se verifique uma probabilidade razoável de que o discurso consegue incitar, diretamente, alguma ação real contra o grupo-alvo (Rabat Plan of Action, 2012).

O *Rabat Plan of Action* nota, com preocupação, que, na realidade, os autores de atos que efetivamente atingem o limiar do Artigo 20º do PIDCP, não sofrem nenhum tipo de punição. Paralelamente, vários membros de minorias são efetivamente perseguidos através do “abuso de legislação vaga, jurisprudência e políticas domésticas” (Rabat Plan of Action, 2012).

Tal como António Guterres, Secretário-Geral das Nações Unidas, disse, “Combater o discurso de ódio não significa limitar ou proibir a liberdade de expressão. Significa impedir que o discurso de ódio escale para algo mais perigoso, particularmente incitação à discriminação, hostilidade e violência, o que é proibido sob Lei Internacional.”.

1.4 Discurso de Ódio Online

O impacto da utilização dos diversos meios de comunicação - desde a imprensa inscrita até à televisão e, mais recentemente, a Internet -, é inegável, tendo instigado uma profunda mudança a nível social, cultural, político e comportamental. De facto, o aparecimento da Internet provocou uma transformação significativa nas formas de organização, de identificação, de comunicação e de mobilização social, verificando-se a existência de interações com o intuito de construir um laço social e de interações que visam enfraquecê-lo, ou, eventualmente, quebrá-lo (Peres-Neto e Pereira, 2019).

Assim, a Internet não só disponibiliza uma pluralidade de “repositórios de conhecimento” (por ex.: Wikipedia, etc.) e aplicações (por ex.: Twitter, Instagram, Facebook, TikTok, YouTube, Reddit, etc.) acessíveis e de fácil utilização por qualquer indivíduo, como, também, constitui o fórum de comunicação da sociedade atual e uma ferramenta indispensável à garantia da liberdade de expressão e de opinião, permitindo a partilha de opinião própria individual relativamente a um determinado assunto, de caráter privado ou público, com uma audiência, mais ou menos extensa. Não obstante, a Internet pode ser utilizada como instrumento de manipulação de massas e de difamação de certos indivíduos ou comunidades, nomeadamente através da difusão de discursos de ódio (Fortuna *et al.*, 2019).

Bilewicz e Soral (2020) defendem que a sensação de anonimidade e a ampla acessibilidade oferecida pelos meios de comunicação digital, facilitaram os processos de influência social, originando uma “epidemia de discursos de ódio” que está a causar uma mudança profunda nas normas sociais relativas à expressão pública do ódio e preconceitos hostis. Brown (2018), por seu lado, identifica ainda mais algumas características dos discursos de ódio *online* que o distinguem dos discursos de ódio realizados *offline* (isto é, por meio de discursos políticos, interações pessoais, etc.), a saber: a invisibilidade fornecida pela distância física entre locutor(a) e audiência; a sensação de pertença a uma comunidade, facilitada pelo uso da Internet, que permite que pessoas com mentalidades similares interajam, sendo que, sem a mesma, não poderiam ter contacto devido, por exemplo, a restrições geográficas; o caráter instantâneo da Internet, que restringe o tempo de ponderação, entre o partilhar, ou não, uma determinada ideia ou pensamento.

As características supramencionadas podem ser exploradas no sentido de disseminar os discursos de ódio, nomeadamente recorrendo ao *trolling* (Cervone *et al.*, 2021) – o ato de postar e/ou comentar publicações *online* com o objetivo de incomodar outrem.

1.4.1 Papel das Redes Sociais

Ao longo dos últimos anos, tem-se verificado uma vasta adesão às Redes Sociais, que passaram a constituir uma parte essencial da nossa vida quotidiana, causando um grande impacto a nível social, político e económico (McCay-Peet e Quan-Haase, 2017). Aliás, as autoras adiantam que “elas (Redes Sociais) foram adotadas à volta do globo a uma

velocidade sem precedentes, revelando a natureza extraordinária do fenómeno das Redes Sociais” (p.13).

Existem inúmeras propostas de definição para o termo “Redes Sociais”, sendo que McCay-Peet e Quan-Haase (2017) citam Papacharissi (2016), que argumenta que qualquer tentativa será inevitavelmente específica a um determinado contexto, uma vez que “a nossa compreensão das redes sociais é temporal, espacial e tecnologicamente sensível – informada, mas não restrita pelas definições, práticas e materialidades de um único período ou local.” (p. 15).

Não obstante, Taprial e Kanwar (2012) resumizam o conceito, afirmando que “todas as aplicações baseadas na Internet que permitem a criação/troca de conteúdo gerado pelo utilizador e possibilitam a interação entre utilizadores pode ser classificada como “Rede Social”. De acordo com os mesmos autores, estas podem ser encontradas na forma de Sites de Redes Sociais (como Facebook, Twitter, Instagram, TikTok, Google+, Reddit, etc.), *blogs*, fóruns da Internet, sites de comunidades *online*, sites de *Q & A*, entre outros.

Para o propósito do presente Projeto, o foco centrar-se-á unicamente nos Sites de Redes Sociais (SRS), definidos como sistemas que possibilitam a criação de uma personalidade *online*, pública ou semipública, através da construção de um perfil pessoal e da interação com outros dentro do mesmo sistema, através de comentários e “*likes*” (McCay-Peet e Quan-Haase, 2017).

As Redes Sociais promovem a comunicação de variadas formas, possibilitando a interação dos seus utilizadores através dos contactos que estabelecem, que ocorrem, simultaneamente, “em massa” (por ex., através de *posts* públicos) e a nível interpessoal (por ex., através das DM – “mensagens diretas”) (Carr *et al.*, 2008).

Apesar de os Sites de Redes Sociais facilitarem a representação de uma fração da sociedade, existem estudos que indicam que nem todos os grupos sociais estão igualmente representados (McCay-Peet e Quan-Haase, 2017).

Assim, é de destacar que a adesão às Redes Sociais, assim como o seu uso e implicações sociais são “processos sociais dinâmicos que ocorrem dentro de padrões existentes de desigualdade: alguns grupos sociais são simplesmente retirados da conversação nas redes sociais.” (McCay-Peet e Quan-Haase, 2017, p.14).

Existem, ainda, diversas características das Redes Sociais que podem facilitar a disseminação de ódio *online*, podendo conduzir a uma organização social de distribuição e recompensa por mensagens de ódio (Walther, 2022).

Nesse sentido, destaca-se a facilidade com que é possível identificar um possível “alvo” - ou até um “simpatizante”, com mentalidade e opiniões semelhantes -, através do uso das Redes Sociais, assim como o facto de estas plataformas favorecerem o *feedback* entre os seus utilizadores, quer sob a forma de comentários privados, ou em publicações de terceiros ou ainda em resposta a um comentário prévio, quer sob a forma de “ferramentas de um clique”, tais como os “Likes/Gostos” (Facebook, Instagram, TikTok), os “Favoritos” (Twitter), os “+1” (Google+) e os “UpVote” (Reddit) – facto que promove a gratificação do recetor, a nível emocional, social e de *status* (Walther, 2022).

Este tipo de *feedback* tem, adicionalmente, a capacidade de estimular três processos de relevo (Walther, 2022):

- Em primeiro lugar, a receção de um *feedback* positivo (grande número de “likes/gostos” e “partilhas”) motiva o(a) locutor(a) no sentido de continuar a difundir ainda mais ódio (Walther, 2022);

- Segundo, este pode, ainda, engrandecer o ódio que o(a) locutor(a) sente pelo alvo ou grupo-alvo, uma vez que o *feedback* positivo desempenha um papel fulcral no reforço e amplificação das perceções e crenças pré-existentes (Walther, 2022);

- Finalmente, a interação mútua com outros defensores dos mesmos ideais pode facilitar o desenvolvimento de uma relação pessoal intensa, uma vez que a atração interpessoal aumenta à medida que ambas as partes descobrem novas coisas que odeiam em comum, ainda que a relação se mantenha puramente virtual (Walther, 2022).

1.5 A Importância de Combater o Discurso de Ódio

Apesar de o discurso de ódio sempre ter existido, a atual disseminação do mesmo, significativamente intensificada durante a pandemia de COVID-19, tem constituído um desafio inédito cujas consequências podem ser avassaladoras, tanto para o(s) grupo(s)-alvo – que podem ser sujeitos a atos psicológica e fisicamente abusivos, violentos e discriminatórios, e à exclusão social, económica e política –, como para a sociedade em geral – tendo em conta que expressões extremas de ódio podem originar perturbações na

paz e desenvolvimento social, gerando tensão, conflito e a violação dos Direitos Humanos.

Segundo as Nações Unidas, o “discurso de ódio é uma negação dos valores da tolerância, inclusão, diversidade e da própria essência das normas e princípios dos Direitos Humanos”, existindo uma forte ligação entre os discursos de ódio e os crimes de ódio “concretos”, pelo que a identificação e intervenção atempada com indivíduos que promovem este tipo de narrativa pode prevenir a escalada da “fala” para a ação (Waseem e Hovy, 2016). Assim, afigura-se necessário combater esta problemática, tomando atenção ao facto de que as características inerentes aos discursos de ódio *online*, representam um desafio adicional, uma vez que estas ofensas podem ocorrer a qualquer hora do dia, ou noite, e são especialmente difíceis de prevenir, monitorizar e medir.

Nesse sentido, é do entender das Nações Unidas que o combate ao discurso de ódio deverá constituir um esforço conjunto da sociedade, requerendo o apoio de todos os indivíduos e organizações (abrangendo governos, o Setor Privado, os *Media*, corporações da Internet, educadores, líderes religiosos, a juventude e a sociedade civil), que têm o dever moral de denunciar discursos de ódio e tentar combatê-los.

1.5.1 Dados Estatísticos: Incidentes, Ofensas, Vítimas, Ofensores, Localização

De um ponto de vista geral, e segundo a mais recente avaliação do progresso do Código de Conduta para a luta contra o discurso ilegal de ódio *online* – acordo assinado em 2016 entre a União Europeia e diversas empresas de Tecnologias de Informação, entre as quais o Facebook, Twitter, Instagram, Youtube, etc., tal como mencionado anteriormente –, durante o passado ano de 2022, assistiu-se a uma diminuição na eficiência das respostas dadas pelas empresas, que, na sua maioria, demorou mais de 24 horas a rever as denúncias recebidas. Assim, o número de notificações revistas no período de um dia após receção da mesma, desceu de 81% em 2021 para 64,4% em 2022, sendo que o TikTok foi a única empresa que melhorou neste aspeto (European Commission Website).

Por outro lado, a mesma avaliação mostra que a taxa de remoção de conteúdo impróprio reduziu de 71% em 2020 para 63,6% em 2022, notando-se, relativamente a este parâmetro, uma melhoria somente na empresa YouTube (que removeu 90,4% do conteúdo denunciado como impróprio). Porém, cerca de 69,9% do conteúdo a incitar à violência ou homicídio de grupos específicos, foi removido, assim como

aproximadamente 59,3% dos conteúdos que continham expressões ou imagens difamatórias, em referência a um determinado grupo social. Assim, constata-se que a remoção de manifestações de ódio consideradas “graves”, foi mais satisfatória do que o desempenho, a nível geral, neste parâmetro (European Commission Website).

Finalmente, é importante destacar que, de acordo com a mais recente monitorização, a frequência e qualidade dos *feedbacks* dados pelas diferentes empresas aos respetivos utilizadores, sofreu um desenvolvimento positivo – particularmente o TikTok, que respondeu a 74,8% das notificações recebidas (comparativamente a 28,7% em 2021), e o Instagram, cuja taxa de *feedback* foi de 72,6% (sendo que em 2021, a mesma taxa foi de 41,9%) (European Commission Website).

Após uma extensa pesquisa, verificou-se que não existem dados estatísticos relativos à prevalência do discurso de ódio *online* a nível global, nem relativamente à sua incidência em cada país, possivelmente devido à ambiguidade e falta de consenso sobre os parâmetros que ditam, se uma determinada manifestação pode constituir discurso de ódio, e que dificultam a sua identificação.

Deste modo, existem apenas diversos estudos acerca da ocorrência deste fenómeno em múltiplos idiomas. Estes estudos, por sua vez, disponibilizam *datasets*, ou “conjuntos de dados”, sendo que para os propósitos do presente Projeto, o foco irá centrar-se em *datasets* criados para as línguas: Portuguesa (Fortuna *et al.*, 2019; Leite *et al.*, 2020); Inglesa (Waseem e Hovy, 2016; Suryawanshi *et al.*, 2020); e *datasets* de carácter multilíngue, com o intuito de fornecer uma amostra dos discursos de ódio manifestados nas línguas Espanhola (Basile *et al.*, 2019) e Italiana (Fersini, Nozza e Rosso, 2018).

Será importante realçar, porém, que a grande maioria dos *datasets* atualmente disponíveis focalizam o seu estudo, quase unicamente, para a plataforma Twitter.

Datasets em Língua Portuguesa

Fortuna *et al.* (2019), criaram um *dataset* para identificar discursos de ódio *online* em Português, no qual estabelecem 81 categorias de discurso de ódio, organizadas num esquema hierárquico, que facilita a identificação dos tipos distintos de discurso de ódio e do modo como estes se inter-relacionam. Ademais, este *dataset* foca-se unicamente neste fenómeno, não considerando conteúdo ofensivo no geral, possibilitando a recolha de mais dados, sem restrições a tópicos específicos.

Para criar este *dataset*, Fortuna *et al.* (2019) utilizaram o motor de busca da plataforma Twitter para pesquisar: a) palavras-chave e *hashtags* relacionadas com discursos de ódio, tais como “fufas”, “sapatão”, “*dyke*”, “#LugarDaMulherÉNaCozinha” e “#womensPlaceIsInTheKitchen”; e b) perfis relacionados com a mesma problemática, tendo descoberto várias contas de utilizadores Portugueses cujo propósito único é a divulgação de conteúdo discriminatório e odioso contra determinadas minorias. Por fim, os autores recolheram uma amostra de 5 668 *tweets* de 1 156 perfis distintos, sendo que mais de 95% das publicações foram feitas entre janeiro e março de 2017.

Assim, um Investigador da área da Detecção Automática de Discursos de Ódio, com formação em Psicologia Social, analisou os *tweets*, tendo posteriormente classificado os mesmos de acordo com a mesma estrutura, resultando na hierarquia representada na Figura 1.

Class	ND	Parent nodes	Freq	Class	ND	Parent nodes	Freq
Hate speech	0	-	1228	Ageing	1	Hate speech	4
Sexism	1	Hate speech	672	Angolans	3	Africans	4
Women	2	Sexism	544	Nordestines	3	Rural people, Brazilians	4
Homophobia	1	Hate speech	322	Chinese	3	Asians	3
Homossexuais	2	Homophobia	288	Homeless	2	Other/Lifestyle	3
Lesbians	3	Homossexuais, Woman	248	Arabic	2	Origin	2
Body	1	Hate speech	164	Bisexuals	2	Homophobia	2
Fat people	2	Body	160	Blond women	2	Women, Body	2
Fat women	3	Women, Fat people	153	East europeans	2	Origin	2
Ugly people	2	Body	131	Jews	2	Religion	2
Ugly women	3	Women, Ugly people	130	Journalists	2	Other/Lifestyle	2
Racism	1	Hate speech	94	Old people	2	Ageing	2
Ideology	1	Hate speech	92	Thin people	2	Body	2
Migrants	1	Hate speech	82	Thin women	3	Women, Thin people	2
Men	2	Sexism	70	Vegetarians	2	Other/Lifestyle	2
Refugees	2	Migrants	70	White people	2	Racism	2
Feminists	2	Ideology, Sexism	65	Young people	2	Ageing	2
Gays	3	Homossexuais	56	Agnostic	2	Ideology	1
Black people	2	Racism	52	Argentines	3	Latins	1
Religion	1	Hate speech	30	Autists	2	Health	1
Left wing ideology	2	Ideology	26	Brazilian women	3	Women, South Americans	1
Origin	1	Hate speech	26	Egyptians	3	Arabic	1
Trans women	3	Women, Transsexuals	26	Football players women	2	Women, Other/Lifestyle	1
Other/Lifestyle	1	Hate speech	20	Gamers	2	Other/Lifestyle	1
Islamists	2	Religion	17	Homeless women	3	Women, Homeless	1
Immigrants	2	Migrants	15	Indigenous	2	Racism	1
Transsexuals	2	Sexism	14	Iranians	3	Arabic	1
Muslims	2	Religion	11	Japaneses	3	Asians	1
Black Women	3	Women, Black people	8	Men Feminists	3	Feminists, Men	1
Criminals	2	Other/Lifestyle	8	Mexicans	3	Latins	1
Latins	2	Racism, Origin	7	Muslim women	3	Muslims, Women	1
Health	1	Hate speech	6	Old women	3	Women, Old people	1
Rural people	2	Origin	6	Polyamorous	2	Other/Lifestyle	1
Travestis	3	Women	6	Poor people	2	Other/Lifestyle	1
Aborting women	3	Women	5	Russians	3	East europeans	1
Asians	2	Racism, Origin	5	Sertanejos	3	Rural people, Brazilians	1
Brazilians	3	South Americans	5	Street artists	2	Other/Lifestyle	1
Disabled people	2	Health	5	Ucranians	3	East europeans	1
South Americans	2	Origin	5	Venezuelans	3	Latins	1
Africans	2	Origin	4				

Figura 1 – Hierarquia de Classes

Finalmente, os autores obtiveram um *dataset* multi-categoria que registou que 22% dos *tweets* analisados continham discurso de ódio. É de destacar, ainda, que este *dataset* identifica categorias menos estudadas, tais como o “discurso de ódio contra pessoas

pobres”, o “discurso de ódio contra pessoas ‘gordas’ ” ou o “discurso de ódio contra homens”.

Mais recentemente, Leite *et al.* (2020) propuseram um novo *dataset* em Português do Brasil – ToLD-Br (*Toxic Language Dataset for Brazilian Portuguese*) -, em larga escala, com *tweets* classificados como “tóxico” ou “não-tóxico” ou de acordo com diferentes tipos de toxicidade. Os autores definem “comentários tóxicos” como “comentários que contêm insultos e linguagem obscena também é considerada, para além do discurso de ódio”.

Assim, recolheram mais de 10 milhões de *tweets*, publicados entre julho e agosto de 2019, através de duas estratégias: a) procura de *hashtags* e “palavras-chave” com maior probabilidade de serem usadas em contexto tóxico (por exemplo, “gay”, “mulherzinha” e “nordestino”); e b) recolha de *tweets* que mencionam figuras influentes e, portanto, mais propensas a ser alvo de comentários tóxicos. Desta colheita, foram selecionados aleatoriamente 21 000 *tweets*.

Por fim, foram registados: 9 255 *tweets* “tóxicos”; 138 *tweets* “racistas”; 6 652 *tweets* “obscenos”; 463 *tweets* “misóginos”; 151 *tweets* “xenófobos”; 4 385 *tweets* “insultuosos”; e 344 *tweets* “LGBTQ+fóbicos”.

Datasets na Língua Inglesa

Existem diversos *datasets* disponíveis na Língua Inglesa, sendo que o *dataset* proposto por Waseem e Hovy (2016) é um dos mais comumente utilizados.

Waseem e Hovy (2016) criaram um *dataset* com *tweets* anotados como “Sexista”, “Racista” ou “Nenhum”. Os autores realizaram uma pesquisa manual, durante um período de dois meses, por termos ou calões comuns e relativos a minorias religiosas, étnicas, sexuais e de género - tendo encontrado, adicionalmente, um pequeno número de utilizadores prolíficos na divulgação deste tipo de conteúdo -, da qual resultou um total de 136 052 *tweets*. Assim, os autores selecionaram 16 914 *tweets* que foram posteriormente classificados pelos mesmos e revistos por uma anotadora exterior, no sentido de mitigar vieses por parte dos anotadores.

Para realizar esta classificação, os anotadores consideraram um *tweet* como ofensivo se este:

- Usa um calão sexista ou racista;
- Ataca uma minoria;
- Procura silenciar uma minoria;
- Critica uma minoria (sem apresentar um argumento bem fundamentado);
- Promove, mas não usa diretamente, discurso de ódio ou um crime violento;
- Critica uma minoria e utiliza a falácia do espantalho (uma falácia usada de modo malicioso, que consiste na distorção intencional de um argumento adversário);
- Deturpa descaradamente a verdade ou procura distorcer a opinião sobre uma minoria, através de alegações sem fundamento;
- Mostra apoio a *hashtags* problemáticas, tais como #BanIslam (“banam o Islão”), #Whoriental (jogo de palavras entre “p*ta” e “oriental”) e #WhiteGenocide (“genocídio branco”);
- Estereotipa negativamente uma minoria;
- Defende xenofobia ou racismo;
- Contém um *username* ofensivo, de acordo com os critérios supramencionados, o *tweet* é ambíguo, e o *tweet* refere-se a um tópico mencionado anteriormente.

Deste modo, os mesmos identificaram: 3 383 *tweets* de teor sexista, 1 972 *tweets* com conteúdo racista; e 11 559 *tweets* sem retóricas sexistas nem racistas. Este estudo notou, ainda, que os perpetradores deste tipo de discursos são, na sua grande maioria, pessoas que se identificam como “homens”.

Suryawanshi *et al.* (2020), por seu lado, criaram um *dataset* multimodal para identificar *memes* com conteúdo ofensivo. De acordo com os autores, um *meme* é uma forma de *media* que partilha uma ideia ou emoção, através da Internet, “um elemento de uma cultura ou sistema de comportamento transmitido de um indivíduo para outro por imitação ou outros comportamentos não-genéticos”, com um impacto crescente nos processos de comunicação através das Redes Sociais. Os *memes* podem aparecer sob diversos formatos, nomeadamente imagens, vídeos e *tweets*, sendo que o mais popular corresponde aos *memes* na forma de imagens com texto inserido nestas.

Os autores destacam a importância de estudar o fenómeno da propagação de conteúdo ofensivo sob a forma de *memes*, uma vez que estes são responsáveis por disseminar muito ódio na sociedade, definindo *meme* ofensivo como “um meio que propaga uma ideia ou emoção que pretende danificar a identidade social de um indivíduo-alvo, comunidade, ou diminuir o seu prestígio”, considerado abusivo de modo implícito dado utilizar, por norma, uma frase não-ofensiva em combinação com uma imagem provocatória, ou vice-versa – facto que frequentemente ofusca o verdadeiro significado do *meme* ofensivo, dificultando a sua deteção.

Assim, os autores utilizaram um *dataset* relativo a *memes*, recolhidos de Sites de Redes Sociais como o Twitter, Facebook, Instagram e Reddit, sobre a Eleição Presidencial dos E.U.A. de 2016 para criar o *dataset* MultiOFF.

Após o processo de anotação, foram identificados 305 *memes* “ofensivos” e 438 *memes* “não-ofensivos”.

É importante notar, contudo, que apesar da natureza multimodal deste estudo ser bastante inovadora, a restrição a uma temática específica, limita a sua utilidade e pode introduzir vieses na investigação. Paralelamente, Suryawanshi *et al.* (2020) demonstram que esta abordagem de combinação de modalidades pode ser aplicada noutros tipos de conteúdo multimédia, como vídeos e gravações de áudio.

Datasets Multilíngues

Basile *et al.* (2019) criaram o *dataset* HatEval, para deteção de discurso de ódio contra imigrantes e mulheres no Twitter, nas Línguas Inglesa e Espanhola, focando-se nestes dois alvos devido à sua particular vulnerabilidade.

Neste sentido, os autores realizaram uma recolha de *tweets*, maioritariamente entre julho e setembro de 2018. Através da monitorização dos perfis de potenciais vítimas de discurso de ódio, *download* do histórico de “*haters*” pré-identificados e filtração de “*streams*” no Twitter, pesquisando por diversas “palavras-chave”, os autores recolheram um total de 19 600 *tweets*, dos quais 13 000 estão em Inglês e 6 600 em Espanhol, e 9 091 *tweets* são sobre imigrantes, enquanto 10 509 se referem a mulheres.

Assim, identificaram: no *dataset* de treino e desenvolvimento (com 15 000 *tweets*), 3 209 *tweets* misóginos em Espanhol e 1 991 *tweets* contra imigrantes no mesmo idioma, e, em

Inglês, 5 000 *tweets* contra cada alvo; no *dataset* de teste (com 4 600 *tweets*), 660 *tweets* em Espanhol com discurso de ódio (metade contra mulheres, metade contra imigrantes), e 940 *tweets* sem discurso de ódio, enquanto em Inglês encontraram 1 260 *tweets* que constituíam discurso de ódio (com distribuição idêntica, relativamente aos alvos) e 1 740 que não continham retórica “odiosa”.

Fersini, Nozza e Rosso (2018), realizaram um *dataset*, no âmbito da tarefa de Identificação Automática de Misoginia (*Automatic Misogyny Identification*), que identifica e classifica *tweets* com conteúdo misógino nas Línguas Inglesa e Italiana.

O foco deste estudo é, unicamente, o ódio contra mulheres, visto que a retórica misógina tem vindo a aumentar nos últimos anos, especialmente *online*. Desta forma, os autores definem misoginia como “o ódio ou preconceito contra mulheres”, que pode manifestar-se de variadas maneiras (por ex.: desde a exclusão social e discriminação, até ameaças de violência e objetificação sexual), pelo que propuseram distinguir conteúdo misógino de conteúdo não-misógino, caracterizar comportamentos misóginos e classificar o alvo do *tweet*.

Nesse sentido, identificaram 5 tipos distintos de misoginia, a saber:

- Estereótipo e Objetificação: Ideia comum, fixa e demasiado simplificada de uma mulher (“o lugar da mulher é na cozinha”); descrição da atratividade física de uma mulher e/ou comparação a padrões restritos;
- Dominação: Afirmar a superioridade dos homens em relação às mulheres para realçar a desigualdade de género;
- “Descarrilamento”: Justificar abuso contra mulheres, rejeitando a responsabilidade masculina; tentativa de interromper a conversa com o objetivo de redirecionar as conversas das mulheres para algo menos desconfortável para os homens;
- Assédio Sexual e Ameaças de Violência: Descrever ações como avanços sexuais, pedidos de favores sexuais, assédio de natureza sexual; intenção de exercer poder sob as mulheres fisicamente, através de ameaças de violência;
- Descrédito: Falar “por cima” de mulheres com nenhuma intenção maior.

Os autores recolheram *tweets* misóginos, tanto em Inglês como em Italiano, através do “*streaming download*”, usando palavras-chave definidas manualmente (por ex.: “*bi**h*”, “*w**re*”, “*c*nt*” em Inglês, e “*pu****a*”, “*tro**a*”, “*f**a di legno*” em Italiano), a monitorização de perfis de potenciais vítimas e o *download* do histórico de utilizadores identificados como misóginos (por ex.: utilizadores que explicitam crenças misóginas nos perfis), tendo composto dois *datasets*, cada um com 5 000 *tweets*, correspondentes a ambos os idiomas.

De seguida, os *tweets* foram classificados por seis anotadores especialistas, que identificaram:

- em Inglês, 2 755 *tweets* “não misóginos” e 2 245 “misóginos”, dos quais 1 155 *tweets* constituíam uma situação de “descrédito”, 396 de “assédio sexual e ameaça de violência”, 103 de “descarrilamento”, 319 de “estereótipo e objetificação” e 272 de “dominação”;

- já em Italiano, encontraram 2 660 *tweets* “não-misóginos” e 2 340 “misóginos”, sendo que 738 *tweets* correspondiam a situações de “descrédito”, 601 de “assédio sexual e ameaça de violência”, 26 de “descarrilamento”, 843 de “estereótipo e objetificação” e 132 de “dominação”.

1.5.2 Prevenção, Intervenção e Estratégias

A Organização das Nações Unidas acredita que o combate ao discurso de ódio implica, em primeira instância, a monitorização e análise do mesmo, no sentido de compreender as suas dinâmicas, e consequente imposição de limites a este tipo de retóricas, visto que podem representar um “sinal de aviso” para a ocorrência efetiva de atos violentos (incluindo crimes de atrocidade, nos quais se inserem o genocídio, os crimes de guerra e os crimes contra a humanidade). Do mesmo modo, no interesse do fim da impunidade, os autores de discursos de ódio deveriam ser responsabilizados.

Neste sentido, e em resposta à alarmante e perturbadora onda de xenofobia, racismo e intolerância, antissemitismo, islamofobia, perseguição de Cristãos e misoginia violenta a nível global, o Secretário-Geral das Nações Unidas, António Guterres, lançou, a 18 de junho de 2019, a Estratégia e Plano de Ação das Nações Unidas contra o Discurso de Ódio (United Nations Strategy and Plan of Action, 2019).

Assim, este Plano de Ação visa: a) reforçar os esforços da ONU para investigar as causas e motivações que originam discursos de ódio, e b) permitir respostas efetivas da ONU sob o impacto do discurso de ódio nas sociedades.

Não obstante a tomada de ação contra este fenómeno, a ONU reitera que todas as medidas tomadas deverão seguir as normas e padrões dos Direitos Humanos Internacionais, particularmente no que concerne à liberdade de expressão e opinião. A ONU reforça, ainda, que o combate ao discurso de ódio é um dever comum a todos os cidadãos – incluindo governos, sociedades, o setor privado, e todos os indivíduos -, pelo que cada pessoa deve agir em conformidade quando confrontada com este fenómeno (United Nations Strategy and Plan of Action, 2019).


Para combater os discursos de ódio, a ONU compromete-se a implementar ações, quer a nível global, quer nacional, assim como promover a cooperação interna entre entidades relevantes das Nações Unidas. A Estratégia e Plano de Ação das Nações Unidas disponibiliza, ainda, as seguintes recomendações:

- **Monitorização e análise de discursos de ódio:** As entidades da ONU devem ser capazes de reconhecer, monitorizar, colher dados e analisar as tendências deste fenómeno;
- **Investigar causas “raiz”, motivadores e atores de discurso de ódio:** No sentido de combater com maior eficácia e/ou amenizar o impacto dos discursos de ódio, a ONU deve procurar compreender as causas e motivações que originam os discursos de ódio;
- **Apoiar as vítimas de discursos de ódio:** As entidades da ONU devem ser solidárias para com as vítimas de discursos de ódio – promovendo os seus direitos e respondendo às suas necessidades, nomeadamente através da advocacia de remédios, acesso à justiça e acompanhamento psicológico -, e garantir a implementação de medidas centradas nos Direitos Humanos e que visam prevenir discursos de ódio em retaliação e escalada de violência;
- **Convocação de atores relevantes:** Quando relevante para o contexto, a ONU deve apoiar a convocação de atores relevantes, reformular os problemas de modo que as soluções sejam mais exequíveis, facilitar mediação e especialidade independente e criar alianças;

- **Envolvimento com formas de *media* novas e tradicionais:** A ONU deve criar e fortalecer parcerias com diversas formas de *media* para combater discursos de ódio e promover narrativas positivas de tolerância, não-discriminação, diversidade e liberdade de expressão e opinião;
- **Usar a tecnologia:** A ONU deve manter-se atualizada relativamente às inovações tecnológicas, promover estudos acerca da relação entre o uso da Internet e das Redes Sociais para partilhar discursos de ódio e os fatores que motivam as pessoas a recorrer à violência, e transmitir a setores privados, como companhias de Redes Sociais, recomendações sobre como promover os princípios da ONU e como combater os discursos de ódio;
- **Usar a educação como ferramenta para combater discursos de ódio:** As entidades da ONU devem favorecer a educação, formal e informal, para implementar o Objetivo de Desenvolvimento Sustentável 4, assim como promover os valores e valências da Educação para a Cidadania Global e a Alfabetização Informacional e Mediática;
- **Fomentar sociedades pacíficas, inclusivas e justas para combater as causas “raiz” e motivadores de discursos de ódio:** A ONU deve educar o público acerca dos Direitos Humanos, tolerância, não-discriminação, igualdade de género e compreensão sobre diferentes culturas e religiões, tanto *online* como *offline*, e favorecer o diálogo intercultural, inter-religioso e inter-crenças e a compreensão mútua;
- **Envolvimento em advocacia:** A ONU deve servir-se da advocacia, tanto em privado, como em público, para realçar tendências alarmantes de discursos de ódio e para expressar simpatia e apoio para os indivíduos ou grupos-alvo;
- **Desenvolvimento de protocolos para comunicações externas:** As comunicações devem ser utilizadas de modo a combater e mitigar o impacto dos discursos de ódio, evitando potenciais escalas de violência, sem restringir o direito à liberdade de expressão e opinião;
- **Fomentar parcerias:** A ONU deve procurar criar ou fortalecer parcerias com empresas interessadas e de relevo, tais como na indústria da tecnologia, uma vez que o combate aos discursos de ódio é, igualmente, responsabilidade dos governos, organizações regionais e multilaterais, companhias privadas, os meios de comunicação, organizações religiosas e outros atores da sociedade civil;

- **Desenvolver as competências do staff da ONU:** As competências laborais e de liderança de membros do *staff* de entidades da ONU de relevo, devem ser desenvolvidas, designadamente através de programas existentes;

- **Apoiar Estados-Membros:** Face a uma solicitação, a ONU deve disponibilizar apoio aos Estados-Membros no sentido de desenvolver capacidades e políticas para combater os discursos de ódio. Assim, a ONU convocou uma conferência internacional sobre Educação para a Prevenção com foco no combate aos discursos de ódio, evento que iria envolver os Ministros da Educação (esta conferência foi realizada a 30 de setembro e 1 de outubro de 2021, em Paris, França) (United Nations Strategy and Plan of Action, 2019; UNESCO, 2021).

Mais recentemente, no âmbito de uma Campanha realizada anualmente, de 25 de novembro a 10 de dezembro, pelo Fundo da População das Nações Unidas, denominada “16 Dias de Ativismo contra a Violência de Género”, foi criada a iniciativa “Bodyright”, com o intuito de combater a violência digital, destacando que os logótipos das empresas e a propriedade intelectual, têm maior valor e são protegidas, *online*, com maior eficiência do que imagens de corpos humanos, que são regularmente publicadas, e possivelmente utilizadas de forma maliciosa, na Internet sem o devido consentimento. Nesse sentido, a Campanha “Bodyright” disponibiliza o símbolo “” - passível de ser adicionado a qualquer imagem, quer diretamente nas “histórias” do Instagram através de *stickers*, quer fazendo o *download* na página da Campanha (<https://www.unfpa.org/bodyright>) -, visando responsabilizar os formuladores de políticas, as companhias e os indivíduos, promovendo simultaneamente a valorização e respeito pelos direitos *online* das mulheres, raparigas, minorias étnicas e raciais, comunidade LGBTQIA+ e de outros grupos marginalizados.

Ainda no âmbito da Campanha “Bodyright”, a United Nations Population Fund (UNFPA) criou o *website* “*The Virtual is Real*” (“O Virtual é Real”, acessível através de <https://www.unfpa.org/thevirtualisreal>), que apresenta histórias de vítimas e sobreviventes de violência digital de todo o mundo, assim como o trabalho realizado pela UNFPA para combater estes fenómenos, destacando o facto de que, segundo uma pesquisa (questionário) realizada pela Unidade de Inteligência Economista, no sentido de identificar a prevalência de violência digital contra mulheres, 85% das mulheres que

utilizam a Internet reportam ter testemunhado algum tipo de violência *online* contra outras mulheres, sendo que quase 40% foi alvo desta violação dos Direitos Humanos.

Para além disso, de acordo com o mesmo estudo, 65% das mulheres questionadas foram vítimas de discurso(s) de ódio, 67% de difamação e desinformação, 66% de “ciber-assédio”, 57% de abuso através do uso de vídeos e imagens e 52% de ameaças violentas (onlineviolencewomen.eiu.com).

Já o Setor da Juventude do Conselho da Europa, desenvolveu o Projeto “*Young People Combating Hate Speech Online*” (Jovens pelo Combate ao Discurso de Ódio *Online*), que decorreu entre 2012 e 2014, com o intuito de combater discriminação e racismo através da tomada de ação contra discursos de ódio, fortalecendo as competências dos jovens e organizações de juventude no sentido de as capacitar a identificar e combater este fenómeno. Este Projeto beneficiou de um envolvimento a nível Europeu e nacional, tendo proporcionado múltiplas atividades, nomeadamente: cursos formativos, desenvolvimento de materiais educativos, seminários e conferências nacionais (Titley, Keen e Földi, 2014).

Um dos elementos centrais deste Projeto, foi a Campanha “Movimento Contra o Discurso de Ódio – Jovens Pelos Direitos Humanos *Online*”, lançada em março de 2013 e que decorreu até ao final de 2017 (www.odionao.com.pt), sendo que a primeira fase da campanha destinou-se à sensibilização dos jovens em relação a este fenómeno, mudança de atitudes e mobilização dos mesmos, enquanto a segunda fase se focou nas estratégias preventivas e respostas a nível educativo (Keen e Georgescu, 2016).

Deste modo, este Movimento não visa, simplesmente, o recurso a meios legais para combater os discursos de ódio, nem a eliminação do ódio sempre que este surge – pelo contrário, a Campanha promove o respeito pela liberdade de expressão e pelos Direitos Humanos no mundo *online*, pelo que propõe o desenvolvimento de respostas alternativas, como a educação, prevenção, sensibilização, o fortalecimento do autocontrolo dos(as) utilizadores(as) e o apoio às vítimas de discursos de ódio, com o objetivo de tornar a Internet um espaço seguro para todos os seus utilizadores.

Vale a pena destacar, ainda, que o Manual “Alternativas”, criado no âmbito do “Movimento Contra o Discurso de Ódio – Jovens Pelos Direitos Humanos *Online*”, opta por combater o discurso de ódio através da utilização de “contra narrativas”, com o objetivo de prevenir a opressão e a “violência cultural” – termo que procura explicar como

qualquer elemento de uma cultura ou ideia (por ex.: a Língua, as histórias, as canções, etc.) pode ser utilizado no sentido de justificar violência, direta ou estrutural.

Esta Campanha foi desenvolvida em 45 países, beneficiando de vários parceiros e uma comunidade internacional de ativistas *online* sendo que, em Portugal, o Projeto foi coordenado pelo Instituto Português do Desporto e Juventude, com o apoio de várias entidades e organizações juvenis ou que lutam pelo combate à discriminação e apoio a vítimas, como a APAV (Website APAV para jovens; Website IPDJ).

Finda a Campanha Europeia, o IPDJ afirma que o movimento pelos Direitos Humanos *online* vai prosseguir. De facto, de acordo com a Declaração final da Conferência de Avaliação realizada em abril de 2018, em Estrasburgo, “o IPDJ continua e continuará ativo, vigilante e empenhado em manter vivo o movimento pelos Direitos Humanos *online*, preservando a sua juventude e relevância, convidando todas as entidades parceiras e associações de jovens a participar” (Instituto Português do desporto e da Juventude, 2017).

Por sua vez, a Recomendação Geral nº 15 sobre “Discurso de Ódio”, da Comissão Europeia contra o Racismo e a Intolerância (ECRI, 2016), sugere a adoção dos seguintes comportamentos, no sentido de combater os discursos de ódio:

- Consciencializar o público acerca da importância de respeitar a diversidade e das consequências dos discursos de ódio, demonstrando a falsidade e inaceitabilidade das premissas-base deste tipo de retórica;
- Apoiar as vítimas de discursos de ódio, tanto a nível individual, como coletivo;
- Apoiar a autorregulação de instituições públicas e privadas como meio de combate contra os discursos de ódio;
- Clarificar a responsabilidade civil e administrativa acerca do uso de discursos de ódio, mantendo o devido respeito pela liberdade de expressão e opinião de cada indivíduo (ECRI, 2016).

Para além da Comissão Europeia contra o Racismo e a Intolerância (ECRI, 2016) e da Campanha do Movimento contra o Discurso de Ódio, o Conselho da Europa dispõe ainda do Departamento de Antidiscriminação, dedicado à política, monitorização e

desenvolvimento de capacidades para combater a discriminação, os discursos de ódio e os crimes de ódio, e fomentar sociedades inclusivas.

A UNESCO, por seu lado, combate os discursos de ódio através da educação e alfabetização digital do público, com o intuito de promover um padrão internacional de liberdade de expressão e opinião e investigar as causas dos discursos de ódio. Esta organização apoia, igualmente, diversos países a desenvolver respostas eficientes a este fenómeno, nomeadamente através do fomento das capacidades construtivas de operadores regulamentares e judiciais e da advocacia à transparência das companhias de Internet (UNESCO Website).

Em Portugal, está a decorrer atualmente o Projeto de Investigação “*SocialHate – I Love to Hate!*: Cartografia do ódio em rede”, coordenado pela Escola Superior de Comunicação Social e financiado pelo Instituto Politécnico de Lisboa, que visa “cartografar e caracterizar os discursos de ódio racistas que proliferam nas Redes Sociais Digitais, em Portugal”, no sentido de incrementar um “Barómetro dos Discursos de Ódio” referente ao Facebook, Twitter e YouTube.

Durante o mês de abril de 2021 e 2022, no âmbito da Iniciativa da Semana da Interculturalidade, a Rede Europeia Anti-Pobreza (EAPN) promoveu a Campanha “O Discurso de Ódio não é argumento - #DarAVoltaAoTexto #EAPN”, com o objetivo de consciencializar o público acerca da problemática dos discursos de ódio, através da partilha de frases populares do senso comum, relacionadas com preconceitos generalizados, e sua desconstrução (por ex.: “Se vêm para cá têm é que respeitar – e ser respeitados”, “Toda a gente sabe que o lugar da mulher é – onde ela quiser”, “Vai mas é para a tua terra – aqui não há lugar para o racismo.”, entre outros). Esta Campanha beneficiou do apoio de diversas figuras públicas, nomeadamente de vários atores, músicos, humoristas e treinadores de futebol e de 70 autarquias a nível nacional, incluindo a Região Autónoma da Madeira (Website EAPN).

Para terminar, a título de curiosidade, as Nações Unidas disponibilizam, através do seu *website*, um pequeno questionário para testar o conhecimento individual relativamente a discursos de ódio e a fenómenos relacionados com esta temática.

1.6 Discursos de ódio *online* de carácter misógino

De acordo com o Instituto Europeu para a Igualdade de Género (EIGE, 2017), diversos

fatores supramencionados – tais como o alcance, em constante evolução, da Internet, a progressiva adesão às Redes Sociais e a propagação de uma grande quantidade de informação, tanto de caráter fidedigno, como dúbio (ou, em certas situações, simplesmente erróneo), a um ritmo acelerado -, associados à atual “pandemia” de violência contra as mulheres e as raparigas (VCMR), criaram um novo problema, a nível mundial, com consequências possivelmente impactantes nas esferas económica e social: a “VCMR cibernética” (EIGE, 2017).

Neste sentido, a Organização Mundial de Saúde estima que, uma em cada dez mulheres já foi vítima de pelo menos uma forma de ciber-violência desde os seus 15 anos (EIGE, 2017) e, segundo um estudo realizado pela Amnistia Internacional em 2017, 41% das mulheres questionadas que haviam sido vítimas de abuso ou assédio *online*, reportaram que sentiram que a sua segurança física foi ameaçada, pelo menos numa ocasião e que cerca de 24% partilharam receio também pela segurança de familiares (Amnistia Internacional, 2017). Para além disso, tal como mencionado anteriormente, a prevalência global de violência digital contra mulheres é de cerca de 85%.

Não obstante a atual ausência de uma conceitualização concreta do fenómeno de discursos de ódio *online* de caráter misógino, para o propósito do presente estudo, propõe-se a seguinte definição: “todas as formas de expressão que incitam, promovem, espalham ou justificam violência, ódio ou discriminação” (Website do Conselho da Europa) contra uma mulher ou um grupo de mulheres, ou que as denigre, por razão do seu sexo.

De facto, “a investigação disponível sugere que as mulheres são desproporcionadamente mais afetadas por certas formas de violência cibernética do que os homens” (EIGE, 2017). Nesse sentido, um estudo realizado pelo Pew Research Center, nos EUA, averiguou que apesar de os homens serem alvo de assédio *online* “moderado” (por ex.: “chamar nomes”) com maior frequência, as mulheres – especialmente as mulheres jovens, entre os 18 e os 24 anos – tendem a sofrer, de modo desproporcional, de formas mais graves de ciber-assédio (por ex.: assédio sexual *online* ou ciber-*stalking*) (EIGE, 2017).

Assim, com o intuito de procurar compreender o fenómeno da VCMR cibernética e sua prevalência, o EIGE realizou uma investigação, que consistiu na revisão de literatura existente acerca das diferentes formas de violência cibernética contra mulheres e raparigas e na compreensão da “disponibilidade de dados administrativos e de inquérito

sobre este fenómeno”, tendo concluído que a abordagem neutra à ciber-violência, relativamente ao género do alvo, apresenta algumas limitações, uma vez que existem estudos que apontam que a violência exercida e os danos consequentes são experienciados de maneiras distintas pelos diferentes géneros.

Do mesmo modo, o Instituto Europeu para a Igualdade de Género, defende que, a VCMR cibernética deve ser entendida como uma continuação da violência exercida no “mundo real”.

A violência cibernética contra mulheres e raparigas pode assumir diversas formas, tais como: partilha não-consensual de conteúdo pornográfico (“*revenge porn*”), ciber-*stalking*, objetificação sexual, “*slut-shaming*”, envio de conteúdo pornográfico não solicitado pela vítima, *grooming* e predação sexual, *doxing* (a recolha e partilha não autorizada de informação pessoal de outrem), tráfico facilitado por via eletrónica, e o assédio cibernético – onde se incluem o assédio e insultos com base no género da vítima, as ameaças de morte ou de violência física e/ou sexual e os discursos de ódio de carácter misógino (WMC Website).

A VCMR cibernética pode, então, abranger violência psicológica, sexual e, ainda, económica (por ex.: quando informações pessoais da vítima são divulgadas sem o seu consentimento, a sua situação laboral atual e futura pode sofrer consequências), afigurando-se igualmente necessário avaliar a probabilidade de a violência *online* ser concretizada fisicamente.

Não obstante, é imperativo procurar compreender melhor o impacto da VCMR cibernética nas suas vítimas, atentando aos seus testemunhos e experiências pessoais.

A emergência desta problemática foi reconhecida pela Dra. Dubravka Šimonović - Diplomata e antiga Relatora Especial das Nações Unidas sobre Violência contra as Mulheres -, no seu Relatório de junho de 2018, explicando que “Apesar dos benefícios e do potencial empoderador da Internet e das TIC, mulheres e raparigas por todo o mundo, têm expressado cada vez mais as suas preocupações relativamente ao conteúdo danoso, sexista, misógino e violento e ao comportamento *online*”, pelo que se afigura “importante reconhecer que a Internet está a ser usada num ambiente mais amplo de discriminação generalizada e sistémica e de violência, baseada no género, contra mulheres e raparigas” – facto que restringe as mulheres e raparigas, impedindo-as de usufruir plenamente dos

seus Direitos Humanos e, por conseguinte, de alcançar a igualdade de género (Šimonović, 2018).

Ging e Siapera (2018) destacam, ainda, que a Internet não só agrega e favorece a proliferação de diferentes formas de misoginia, como também fomenta a criação de novos fenómenos misóginos dependentes dos recursos tecnológicos disponibilizados pelos novos *media* (por ex.: *revenge porn* ou *doxxing*), das políticas e dos algoritmos utilizados por certas plataformas, da cultura de espaços laborais que produzem estas tecnologias, e dos indivíduos e comunidades que os utilizam.

As autoras defendem, ainda, que a relação entre o mundo *online* e *offline* não é, necessariamente, linear, argumentando que existem, por exemplo, determinadas plataformas *online* que permitem que homens (que se autoproclamam de “*beta males*”) utilizem a misoginia e o racismo como modo de “proteger” essas mesmas plataformas - que, no seu entender, são espaços reservados, exceccionalmente, a homens brancos -, apesar de, na realidade, poderem possuir baixo capital económico, social e/ou cultural.

Segundo Ging e Siapera (2018) é, igualmente, necessário adotar uma abordagem interseccional aquando da elaboração de pesquisas ou respostas preventivas à misoginia *online* em contextos culturais, legais e geográficos distintos, adiantando que estudos realizados em diferentes países asiáticos (China, Índia e Coreia do Sul) reportam preocupações significativamente diferentes – facto que indica uma demanda urgente de um maior número de estudos acerca desta temática, para além do contexto Anglófono. As autoras consideram, ainda, que esta abordagem interseccional, seria importante no sentido de procurar compreender o modo como a misoginia *online* por vezes se encontra associada a outras formas de abuso discriminatório *online*, tais como a homofobia, o racismo, o classicismo e o capacitismo.

Ging e Siapera (2018) reconhecem, também, o alcance e a diversidade do ódio misóginico *online*, argumentando que este pode existir em contextos públicos, laborais, sociais (“nas ruas”) e domésticos, assim como sob a forma de ataques contra homens “aliados”.

Assim, as autoras inspiraram-se no trabalho de Hess e Waller (2014) – que comparam o fenómeno atual de “humilhação *online*”, às práticas medievais de humilhação pública (sugerem que estes rituais medievais deixavam, com frequência, marcas permanentes da humilhação sofrida sob a forma de cicatrizes, à semelhança da “humilhação *online*”, que

provoca uma “marca digital de humilhação” igualmente duradoura), apelidando o conceito de “pelourinho digital”, em referência ao pelourinho medieval, um instrumento de madeira que imobilizava os ofensores no pescoço e pulsos, expondo-os ao abuso do público sem possibilidade de se defenderem – e propõem que determinadas figuras se têm tornado alvos preferenciais deste fenómeno, nomeadamente as feministas, as “guerreiras de justiça social” (“*social justice warriors*”), as “promíscuas” (“*slut*”), as mulheres que trabalham em empresas tecnológicas ou de videojogos, ou qualquer mulher que se oponha ou questione publicamente as dinâmicas de poder dos géneros.

Segundo o estudo realizado pela Amnistia Internacional em 2017, já mencionado, 76% das mulheres que foram vítimas de abuso ou assédio *online*, mudaram a forma como utilizam as plataformas de Redes Sociais, das quais 32% deixaram de publicar conteúdo opinativo relativamente a determinados tópicos.

Inclusive, 61% das vítimas reportaram baixa auto-estima e auto-confiança após o(s) incidente(s), 55% partilharam episódios de ansiedade, *stress* e até ataques de pânico, 56% experienciaram dificuldades em manter a concentração e 63% em adormecer (Amnistia Internacional, 2017) – factos que confirmam o elevado impacto da vitimação por abuso e/ou assédio misógeno *online* e suas consequências.

Ademais, o Relatório de Misoginia *Online* (2022) - realizado pelas empresas Saatchi e Saatchi e *Everything Everywhere* (EE) no âmbito da Campanha *Hope United 2022* no sentido de combater o ódio misógeno *online* direcionado a atletas da UEFA Women’s EURO 2022 -, registou que 43% das mulheres questionadas referem que a misoginia interfere com a capacidade de se expressarem livremente.

Neste sentido, Ging e Siapera (2018) fazem alusão ao trabalho de Jane (2014), que estudou alguns casos específicos de misoginia *online*, reunindo as perceções de mulheres que foram vítimas deste fenómeno. Deste trabalho surgiram novos conceitos, tais como: “vandalismo económico” e “*gendered cyberhate*” - sendo que o primeiro se refere ao “custo” deste tipo de vitimação (por ex.: oportunidades de trabalho perdidas, produtividade diminuída e abandono da Internet e Redes Sociais), e o segundo ao ódio, motivado pelo género, disseminado na Internet -, concluindo que “as mulheres que são alvo de abuso acabam muitas vezes duplamente penalizadas, não só a sofrer o impacto afetivo do ódio, como também os próprios danos materiais causados, contra os quais (...) mais proteções institucionais devem ser implementadas”.

De um modo generalizado, a utilização da Internet não só é uma comodidade que constitui, cada vez mais, um Direito Humano, como também se afigura, com frequência, uma necessidade para o bem-estar económico dos seus utilizadores. Assim, o “mundo” digital deve constituir um espaço seguro e empoderador para todas as pessoas, incluindo para as mulheres e raparigas (EIGE, 2017) – que têm o direito implícito de desfrutar da Internet sem receio de serem silenciadas, ameaçadas, injuriadas, abusadas ou exploradas sexualmente -, sendo que o dever de combater o ódio misógino *online*, de proteger as vítimas e de lutar pela igualdade de género *online* é responsabilidade de toda a sociedade (Ging e Siapera, 2018).

Parte II - Proposta de Investigação

2.1. Introdução

Conforme referido anteriormente, no decorrer do Capítulo II procura-se expor e delinear uma proposta de estudo empírico, iniciando-se pela apresentação dos objetivos gerais e específicos, da metodologia, da amostra, dos instrumentos e dos procedimentos recomendados, concluindo-se com a identificação dos resultados esperados.

2.2. Objetivos

Explicitar os objetivos é indicar o que se pretende com a realização de um projeto de investigação (Ander-Egg e Aguilar, 1999).

Ainda segundo Guerra (2002), os objetivos gerais descrevem as grandes orientações do projeto e são coerentes com as finalidades da investigação, enquanto que os objetivos específicos, descrevem, de uma forma mais pormenorizada, as especificações que se têm de dar para concretizar o objetivo geral (Ander-Egg e Aguilar, 1999).

2.2.1. Objetivo Geral

O objetivo geral do estudo empírico proposto é a avaliação da prevalência de discursos de ódio *online* de carácter misógino, testemunhados por, ou perpetrados contra estudantes universitárias.

2.2.2. Objetivos Específicos

Com o intuito de cumprir o objetivo geral do estudo empírico proposto, foram estabelecidos objetivos específicos que se afiguram cruciais à determinação explícita da finalidade de cada uma das etapas da investigação (Tamayo, 2004), a saber:

- 1) Compreender se as jovens universitárias têm presente a noção de discurso de ódio misógino.
- 2) Informar as jovens acerca das diferentes formas de discursos de ódio misóginos.
- 3) Identificar a prevalência das diferentes formas de discursos de ódio misóginos, de acordo com a perceção das jovens.
- 4) Avaliar o impacto deste tipo de vitimação, direta ou indireta, nas jovens universitárias.

2.3. Método

Tendo em conta que o objetivo principal do estudo proposto, é avaliar a prevalência de discursos de ódio *online* de carácter misógino contra estudantes universitárias e o impacto deste tipo de vitimação, a investigação deverá seguir uma metodologia quantitativa, procurando estudar este fenómeno através da recolha e análise de dados, com o auxílio

de métodos matemáticos, tais como a estatística, permitindo a generalização da “verdade” encontrada na amostra à população geral (Sukamolson, 2007).

A investigação será do tipo quantitativo e de caráter exploratório – visto existir pouca literatura acerca do fenómeno em estudo, pretendendo-se que a presente pesquisa forneça mais informação e fomente a criação de novas hipóteses -, e descritivo – uma vez que se pretende, de igual modo, compreender o fenómeno de vitimação (DeCarlo, 2018).

O presente estudo envolve a utilização de um inquérito por questionário auto-administrado, via a plataforma Google Forms, seguindo um desenho observacional, transversal, exploratório, descritivo e retro/prospetivo.

Antes da realização efetiva do estudo proposto, é, contudo, necessário obter o consentimento informado por parte das pessoas inquiridas, explicar os objetivos da investigação e recolher os dados sociodemográficos, que servirão para caracterizar a amostra.

2.3.1. População Alvo e Amostra

DeCarlo (2018) define a população alvo, como o conjunto de pessoas de particular interesse ao estudo em questão, notando, contudo, que a utilização da totalidade da população de interesse numa determinada investigação é bastante rara, devido à sua complexidade e caráter de longa-duração. Deste modo, recorre-se à seleção de uma amostra, com base nas observações que se pretende analisar (DeCarlo, 2018).

Atentando aos objetivos propostos, a população alvo corresponderá a estudantes universitárias do sexo feminino ou que se identificam com o género feminino, que utilizam a Internet e/ou Redes Sociais, com idades, identidades de género, cursos e graus curriculares diferenciados.

Assim, dever-se-á selecionar uma amostra não-probabilística e intencional (DeCarlo, 2018), de um número mínimo de 200 estudantes universitárias do sexo feminino que utilizam a Internet e Redes Sociais. As variáveis sociodemográficas supramencionadas pretendem dotar a investigação proposta com um caráter interseccional.

No sentido de recolher esta amostra, deve considerar-se determinados critérios de inclusão – nomeadamente, pessoas maiores de idade; pessoas do sexo feminino e cisgénero, ou pessoas que se identificam com o género feminino; pessoas que frequentam o Ensino Superior numa Instituição Portuguesa, privada ou pública, na altura da aplicação do questionário, incluindo estudantes a obter grau de Licenciatura, Mestrado ou Doutoramento; pessoas que utilizam a Internet e/ou Redes Sociais – e de exclusão – tais como, pessoas menores de idade; pessoas do sexo masculino ou intersexo e/ou que se identificam com o género masculino ou não-binário; pessoas que não frequentam, na altura de aplicação do questionário, o Ensino Superior numa Instituição Portuguesa; pessoas que não utilizam a Internet e/ou Redes Sociais.

2.3.2. Instrumento de Recolha de Dados

Tal como mencionado anteriormente, o instrumento de recolha de dados aplicado no âmbito do estudo proposto deverá ser o inquérito por questionário auto-administrado e transversal (DeCarlo, 2018), por intermédio da plataforma Google Forms (Anexo 1), constituído por perguntas formuladas pela autora e inspiradas em questionários realizados por outros autores.

O questionário será composto por 3 grupos distintos de questões, com um total de 19 questões:

- no primeiro grupo, serão colocadas questões relativas às características sociodemográficas da pessoa respondente (tais como a idade, o sexo, a identidade de género, e o curso e grau curricular);
- o segundo grupo, por sua vez, será constituído por questões referentes à utilização da Internet e/ou Redes Sociais, ao conhecimento da participante sobre o conceito de discurso de ódio *online* misógino e a deste tipo de vitimação, que pode ocorrer de forma, direta ou indireta, e, possivelmente, até já ter sido experienciado pela participante;
- finalmente, o terceiro grupo pretenderá avaliar a perceção da participante acerca da sua experiência de vitimação - em primeiro lugar, a pessoa deverá responder de acordo com o que sentiu durante e/ou após a situação em causa, através de uma questão de escolha múltipla e, de seguida, deverá classificar um conjunto de afirmações relativas ao modo como esta situação impactou a sua vida, utilizando uma Escala de Likert de quatro pontos (1 – discordo completamente; 2 – discordo; 3 – concordo; 4 – concordo completamente).

2.3.3. Procedimentos

Antes da realização efetiva do estudo proposto, será, contudo, necessário elaborar um pedido à Comissão de Ética da Universidade Fernando Pessoa (Anexo 2) e, após a obtenção da devida autorização, afigura-se igualmente fundamental obter a aprovação das Universidades que se pretende estudar no sentido de partilhar o *link* para o questionário a ser preenchido pelas estudantes.

De seguida, será, ainda, crucial obter o consentimento informado (Anexo 3) por parte das pessoas inquiridas, assim como explicar os objetivos da investigação e recolher os dados sociodemográficos, que servirão para caracterizar a amostra.

Após os procedimentos enunciados, pretende-se, no mínimo, conseguir cerca de 200 respostas obtidas de modo aleatório.

O instrumento de recolha de dados será disponibilizado na plataforma Google Forms, de forma a garantir o anonimato das participantes. A duração média estimada para o preenchimento do questionário deverá ser de 5 minutos.

Finalmente, utilizar-se-á o *software Statistical Package for the Social Sciences* (SPSS) para se proceder à análise e tratamento dos dados obtidos anteriormente.

Estes dados deverão ser, em primeira instância, analisados de forma descritiva, de modo a se poder organizar os resultados obtidos numa tabela de frequências e, por fim, de forma correlacional no sentido de compreender o impacto deste tipo de vitimação. Outros testes estatísticos terão de ser equacionados *a posteriori*.

2.4. Resultados Esperados

De destacar, em primeiro lugar, que, tanto quanto foi possível perceber e/ou averiguar, relativamente ao estudo proposto, não se detetou a existência de investigações similares em relação a esta temática, pelo que a projeção/inferência de possíveis resultados constitui um desafio adicional.

Deste modo, espera-se que a presente proposta constitua uma contribuição inicial para futuras investigações acerca da problemática dos discursos de ódio *online* de carácter misógino.

Não obstante, no sentido de obter os resultados pretendidos, será necessária a colaboração de, no mínimo, 200 estudantes universitárias do sexo feminino ou que se identificam com o género feminino, e com características sociodemográficas diversas, nomeadamente idades, identidades de género, cursos e graus curriculares diferentes.

No que concerne o objetivo geral do estudo proposto - a avaliação da prevalência de discursos de ódio *online* de carácter misógino, testemunhados por ou perpetrados contra estudantes universitárias -, espera-se que, à semelhança do estudo realizado pela Unidade de Inteligência Economista em 2020, o número de estudantes que tenham sido vítimas, tanto diretas como indiretas, deste fenómeno seja relativamente elevado (cerca de metade das participantes).

Relativamente aos objetivos específicos, por seu lado, é expectável que as participantes tenham presente a noção de discurso de ódio *online* misógino, ainda que possivelmente de modo mais superficial.

Paralelamente, as formas de discursos de ódio *online* de carácter misógino mais prevalentes deverão ser, em concordância com o Relatório de Misoginia Online (Saatchi e Saatchi e Everything Everywhere, 2022), os “insultos e utilização de calões com base no género” e o “descrédito e desvalorização das opiniões de mulheres”.

Finalmente, presume-se que o impacto deste tipo de vitimação, direta ou indireta, seja variável de participante para participante. Porém, considerando o inquérito realizado pela Amnistia Internacional em 2017, é possível inferir que grande parte das participantes que foram vítimas de discursos de ódio *online* misóginos, poderão reportar diminuição de auto-estima e/ou autoconfiança, ataques de pânico e/ou ansiedade e/ou *stress* persistente, perturbações de sono, e que a sua segurança física foi comprometida.

Em suma, espera-se que, o presente estudo venha evidenciar a importância de combater os discursos de ódio *online* de carácter misógino e a necessidade de apoiar as vítimas deste fenómeno.

2.5. Considerações Finais

A realização do presente projeto de graduação permitiu concluir que os discursos de ódio *online*, nomeadamente de caráter misógino, apesar de serem uma temática pouco conhecida, investigada e trabalhada, constituem, cada vez mais, uma problemática de particular interesse e relevância, devido não só ao seu caráter controverso, como, também, ao seu poder de fomentar e “justificar” a concretização de crimes efetivamente violentos no mundo “real”.

Deste modo, a concretização deste projeto permitiu o aperfeiçoamento e a aplicação de um conjunto de conhecimentos adquiridos no decorrer dos três anos de Licenciatura, assim como de adquirir novos conhecimentos acerca da temática dos discursos de ódio *online*, particularmente de índole misógina.

Através do presente projeto foi possível perceber que os discursos de ódio *online*, em particular contra mulheres e raparigas, constituem um fenómeno social em ascensão e de grande impacto para as suas vítimas, podendo até agir como um fator preditor de problemas sociais mais gravosos, tais como a violência física e/ou sexual e o femicídio.

Não obstante, esta problemática não é, de um modo geral, entendida como suficientemente relevante – facto que condiciona a compreensão plena deste fenómeno e, consequentemente, a criação de planos de ação e/ou prevenção unânimes e eficientes.

Paralelamente, aquando do combate a este tipo de retóricas, torna-se necessário compreender a distinção entre uma opinião meramente negativa ou depreciativa e um discurso de ódio, no sentido de promover o bem-estar e a segurança *online* de todas as pessoas que utilizam a Internet, sem desprezar o seu Direito à Liberdade de Expressão e Opinião – um dos Direitos Fundamentais que deve ser assegurado a todos os cidadãos.

Finalmente, a proposta de estudo apresentada, visa avaliar a prevalência de discursos de ódio *online* de caráter misógino, perpetrados contra e/ou testemunhados por estudantes universitárias do sexo feminino ou que se identificam com o género feminino.

De igual modo, a investigação proposta procura compreender este fenómeno de vitimação, avaliando o seu impacto nas estudantes. Assim, após a análise dos dados resultantes do questionário, poderão ser criadas, não só, estratégias de prevenção e de combate aos discursos de ódio *online*, particularmente de caráter misógino, como a apresentação de propostas de intervenção junto das vítimas.

Referências

- Ander-Egg, E. e Aguilar, M. J. (1999). Como Elaborar um Projecto: Guia para Desenhar Projectos Sociais e Culturais. (15.^a Ed.). Lisboa: Centro Português de Investigação em História e Trabalho Social.
- Basile, V., Bosco, C., Fersini, E., Nozza, D., Patti, V., Rangel, F., Rosso, P. e Sanguinetti, M. (2019). SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter. *In: Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*, pp. 54-63. Association for Computational Linguistics.
- Benesch, S. (2014). Countering Dangerous Speech: New Ideas for Genocide Prevention. [Em linha] Disponível em: <https://www.ushmm.org/m/pdfs/20140212-benesch-countering-dangerous-speech.pdf> [Consultado em 21.03.2023].
- Bilewicz, M., e Soral, W. (2020). Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology*, 41(S1), pp. 3–33.
- Braga, R. (2018). A Indústria das Fake News e o Discurso de Ódio. *In: Pereira, R. (Org.). Direitos Políticos, Liberdade de Expressão e Discurso de Ódio*. Volume I. Belo Horizonte: Instituto para o Desenvolvimento Democrático, 2018. pp. 203-220.
- Brown, A. (2018). What is so Special about Online (as compared to offline) Hate Speech?. *Ethnicities*, 18(3), pp. 297-326, DOI: <https://doi.org/10.1177/1468796817709846>.
- Carr, C., Choi, S., DeAndrea, D., Van Der Heide, B., Kim, J., Tong, S. e Walther, J. (2008). Interaction of Interpersonal, Peer and Media Influence Sources Online: A Research Agenda for Technology Convergence. [Em linha] Disponível em: <http://www.etchouse.com/mcma503/readings.old/carr-2008.pdf> [Consultado em 23.01.2023].
- Cervone, C., Augoustinos, M. e Maass, A. (2021). The language of derogation and hate: Functions, Consequences, and Reappropriation. *Journal of Language and Social Psychology*, 40(1), pp. 80–101. [Em linha] Disponível em: <https://doi.org/10.1177/0261927X20967394> [Consultado em 22.01.2023].
- Chakraborti, N. (2014). Framing the Boundaries of Hate Crime. *In: N. Hall, A. Corb, P. Giannasi, & J. G. D. Grieve (Eds.), The Routledge International Handbook on Hate Crime* (pp. 13-23). Routledge. [Em linha] Disponível em: https://cabodostrabalhos.ces.uc.pt/n23/recensoes/03_GinaMoura.pdf [Consultado em 21.01.2023].
- Comissão Europeia (2021). Comunicação da Comissão ao Parlamento Europeu e Conselho. Bruxelas, 9.12.2021. COM(2021) 777 final. [Em linha] Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021DC0777&from=IT> [Consultado em 23.01.2023].
- DeCarlo, M. (2018). Scientific Inquiry in Social Work. pp. 164-166. Radford University. Publisher: Open Social Work Education. ISBN 13: 9781975033729

ECRI. (2016) Recomendação de Política Geral nº15 sobre a luta contra o discurso de ódio, p. 3. [Em linha] Disponível em: <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01> [Consultado em 23.01.2023].

Erjavec, K. e Poler, M. (2012) You Don't Understand, this is a New War: Analysis of Hate Speech in News Web Sites' comments. *Mass Communication & Society*, 15(6), pp. 899–920. [Em linha] Disponível em: <https://doi.org/10.1080/15205436.2011.619679> [Consultado em 21.03.2023].

European Institute for Gender Equality (2017). Cyber Violence Against Women and Girls. [Em linha] Disponível em: <https://eige.europa.eu/publications/cyber-violence-against-women-and-girls>. [Consultado em 02.02.2023].

Fersini, E., Nozza, D. e Rosso, P. (2018). “Overview of the EVALITA 2018 Task on Automatic Misogyny Identification (AMI)” in Caselli, T., Novielli, N., Patti, V. e Rosso, P. (dir.), *EVALITA Evaluation of NLP and Speech Tools for Italian*, Torino, Accademia University Press, pp. 59-66.

Fortuna, P., da Silva, J., Soler-Company, J., Wanner, L. e Nunes, S. (2019). A Hierarchically-Labeled Portuguese Hate Speech Dataset. In: *Proceedings of the Third Workshop on Abusive Language Online*, pp. 94-104.

Ging, D. e Siapera, E. (2018). Special Issue on Online Misogyny. In: *Feminist Media Studies*, 18(4), pp. 515-524. <https://doi.org/10.1080/14680777.2018.1447345>.

Guerra, I. (2002). *Fundamentos e Processos de uma Sociologia da Acção: O Planeamento em Ciências Sociais*. Lisboa: Principia.

Hess, K. e Waller, L. (2014). ‘The Digital Pillory: Media Shaming of ‘Ordinary’ People for Minor Crimes’, *Continuum Journal of Media & Cultural Studies* 28(1), pp. 101-111, DOI: <https://doi.org/10.1080/10304312.2013.854868>

Instituto Português do Desporto e da Juventude. (2017). Alternativas: Agir contra o Discurso de Ódio através de Contra narrativas. [Em linha] Disponível em: [16d2315e-a2a3-06e4-5de8-3fd0ffae7e58 \(ipdj.gov.pt\)](https://www.ipdj.gov.pt/16d2315e-a2a3-06e4-5de8-3fd0ffae7e58) [Consultado em 02.02.2023].

Keen, E. e Georgescu, M. (2016). “Referências: Manual para o Combate do Discurso de Ódio Online Através da Educação para os Direitos Humanos”. Fundação Calouste Gulbenkian para a tradução em Língua Portuguesa. [Em linha] Disponível em: https://www.internetsegura.pt/sites/default/files/2019-12/ReferenciasPT2016_BOOK.pdf [Consultado em 03.02.2023].

Leite, J., Silva, D., Bontcheva, K. e Scarton, C. (2020). Toxic Language Detection in Social Media for Brazilian Portuguese: New Dataset and Multilingual Analysis. In: *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pp. 914–924, Suzhou, China. Association for Computational Linguistics.

Lupu, Y., Sear, R., Velásquez, N., Leahy, R., Restrepo, N., Goldberg, B. e Johnson, N. (2023). Offline Events and Online Hate. [Em linha] Disponível em: <https://doi.org/10.1371/journal.pone.0278511> [Consultado em 21.03.2023].

Mason, G. (2014). The Symbolic Purpose of Hate Crime Law: Ideal Victims and Emotion. *Theoretical Criminology*, 18(1), pp. 75-92.

McCay-Peet, L. e Quan-Haase, A. (2017). “What is Social Media and What Questions Can Social Media Research Help us Answer?” *In: The SAGE Handbook of Social Media Research Methods*, pp. 13-26. [Em linha] Disponível em: <https://books.google.pt/books?id=9oewDQAAQBAJ&lpg=PA13&ots=eOOSl3s-AU&dq=WHAT%20IS%20SOCIAL%20MEDIA&lr&hl=pt-PT&pg=PP1#v=onepage&q=WHAT%20IS%20SOCIAL%20MEDIA&f=false> [Consultado em 23.01.2023].

Oliveira, F. (2022). *Ativismo Contra os Discursos de Ódio nas Redes Sociais: Estudo Exploratório com Jovens Adultos*. Edição do Autor. Dissertação de mestrado em Psicologia. Universidade Católica Portuguesa. [Em linha] Disponível em: <https://repositorio.ucp.pt/bitstream/10400.14/38409/1/203041461.pdf> [Consultado em 23.02.2023].

Peres-Neto, L. & Pereira, G. (2019). Ética, Liberdade de Expressão e Discurso de Ódio de Género em Sites de Redes Sociais. *In: Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação* (22), pp. 1-25.

Prieto, D. (2015). Umberto Eco: 'No estoy seguro de que internet haya mejorado el periodismo'. *El Mundo*. [Em linha] Disponível em: <https://www.elmundo.es/cultura/2015/03/26/551385fc22601dfd398b456b.html> [Consultado em 5.04.2023].

Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N. e Wojatzki, M. (2016). Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis. *In: Proceedings of NLP4CMC III: 3rd Workshop on Natural Language Processing for Computer-Mediated Communication (Bochum)*, Bochumer Linguistische Arbeitsberichte, 17, pp. 6-9. [Em linha] Disponível em: <https://doi.org/10.48550/arXiv.1701.08118> [Consultado em 22.01.2023].

Saatchi & Saatchi e Everything Everywhere. (2022). *Online Misogyny Report. Women's Euros 2022*. London.

Schabas, W. (2000). Hate Speech in Rwanda: The Road to Genocide. *In: McGill Law Journal*, 46(1), pp. 141-171.

Selma (2019). *Hacking Online Hate: Building an Evidence base for Educators*. European Schoolnet. [Em linha] Disponível em: <https://hackinghate.eu/assets/documents/hacking-online-hate-research-report-1.pdf>. [Consultado em 23.01.2023].

Šimonović, D. (2018). *Report of the Special Rapporteur on Violence Against Women, its Causes and Consequences on Online Violence against Women and Girls from a Human Rights perspective*. Human Rights Council. United Nations;

Sukamolson, S. (2007). *Fundamentals of Quantitative Research*. Language Institute Chulalongkorn University.

Suryawanshi, S., Chakravarthi, B., Arcan, M. e Buitelaar, P. (2020). Multimodal Meme Dataset (MultiOFF) for Identifying Offensive Content in Image and Text. *In: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pp. 32-41. European Language Resources Association (ELRA).

Tamayo, M. (2004). *El Proceso de la Investigación Científica – Incluye evaluación y administración de proyectos de investigación*. Cuarta Edición. Editorial Limusa, Grupo Noriega Editores.

Taprial, V. e Kanwar, P. (2012). *Understanding Social Media*. Varinder Taprial & Priya Kanwar & Ventus Publishing.

Walther, J. (2022). Social Media and Online Hate. *Current Opinion in Psychology*, 45:101298, DOI: <https://doi.org/10.1016/j.copsyc.2021.12.010>.

Waseem, Z. e Hovy, D. (2016). Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. *In: North American Chapter of the Association of Computer Linguistics – Human Language Technologies (NAACL-HLT) 2016*, pp. 88-93.

Webgrafia

Amnesty International (2017). [Em linha] Disponível em: [Amnesty reveals alarming impact of online abuse against women - Amnesty International](#). [Consultado em 25.04.2023].

APAV para Jovens” – Movimento contra o Discurso de Ódio. [Em linha] Disponível em: <https://www.apavparajovens.pt/pt/go/movimento-contra-o-discurso-do-odio> [Consultado em 25.01.2023].

EUR-LEX – Access to European Union Law. [Em linha] Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=LEGISSUM:l33178&from=PT> – Decisão-Quadro 2008/913/JAI. [Consultado em 26.01.2023].

EUR-LEX – Access to European Union Law. [Em linha] Disponível em: [Tratado sobre o Funcionamento da União Europeia \(europa.eu\)](#). [Consultado em 26.01.2023].

OSCE Office for Democratic Institutions and Human Rights (ODIHR). 2009. HATE CRIME LAWS A Practical Guide. [Em linha] Disponível em: [36426.pdf \(osce.org\)](#). [Consultado em 26.01.2023].

Rabat Plan of Action (2012) [Em linha] Disponível em: [The Rabat Plan of Action | OHCHR](#) [Consultado em 25.01.2023].

UNESCO Events Website. [Em linha] Disponível em: <https://events.unesco.org/event?id=2689680899> — Addressing hate speech through education: Global Education Ministers Conference. [Consultado em 27.01.2023].

United Nations Website. [Em linha] Disponível em: <https://www.un.org/en/hate-speech/understanding-hate-speech/hate-speech-versus-freedom-of-speech> [Consultado em 26.01.2023].

Vile, J. R., *Incitement to Imminent Lawless Action*, First Amendment Encyclopedia. [Em linha] Disponível em: <https://www.mtsu.edu/first-amendment/article/970/incitement-to-imminent-lawless-action>. [Consultado em 26.01.2023].

Website Amnistia Internacional. (2017). Amnesty reveals alarming impact of online abuse against women. [Em linha] Disponível em: <https://www.amnesty.org/en/latest/press-release/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/> [Consultado em 28.01.2023].

Website Conselho da Europa. Hate Speech. [Em linha] Disponível em: <https://www.coe.int/en/web/freedom-expression/hate-speech> - [Consultado em 26.01.2023].

Website EAPN. “O Discurso de Ódio não é argumento. #DarAVoltaAoTexto#EAPN”. [Em linha] Disponível em: <https://www.eapn.pt/campanhas-nacionais/nacionais-campanhas-nacionais/o-discurso-de-odio-nao-e-argumento-dar-a-volta-ao-texto-eapn/> [Consultado em 28.01.2023].

Website Instituto Português de Desporto e Juventude (IPDJ). [Em linha] Disponível em: <https://ipdj.gov.pt/-/movimento-contra-o-discurso-de-%C3%93dio> — Movimento contra o Discurso de Ódio. [Consultado em 25.01.2023].

WMC Website. [Em linha] Disponível em: [Women’s Media Center \(womensmediacenter.com\)](https://www.womensmediacenter.com) [Consultado em 29.01.2023].

Anexos

Anexo 1 - Questionário - “Discurso de ódio online: avaliação da prevalência de discursos de ódio online de caráter misógino contra estudantes universitárias”

No âmbito de um trabalho de investigação, com vista à obtenção do grau de Licenciatura em Criminologia na Faculdade de Ciências Humanas e Sociais (FCHS) da Universidade Fernando Pessoa (UFP), está em curso o projeto “Discurso de ódio *online*: avaliação da prevalência de discursos de ódio *online* de caráter misógino contra estudantes universitárias”.

Pretende-se com o presente inquérito determinar a prevalência de discursos de ódio *online*, de caráter misógino, testemunhados por, ou perpetrados contra estudantes universitárias, assim como compreender o impacto deste tipo de vitimação, direta ou indireta.

Este estudo foi aprovado pela Comissão de Ética da UFP, de modo a garantir a proteção dos direitos, segurança e bem-estar de todas as participantes incluídas e garantir prova pública dessa proteção.

Deste modo, solicito a colaboração com o preenchimento do inquérito que se segue, com uma duração de cerca de 5 minutos. **Não existem respostas certas ou erradas**, apenas a sua opinião conta.

A participação é voluntária, e a recolha de dados é **anónima e confidencial**.

As informações que acabou de ler destinam-se a esclarecê-lo acerca da natureza, alcance, consequências e risco do estudo, de modo a permitir que, depois de esclarecida, se encontre capaz de decidir participar ou não.

Caso não tenha qualquer dúvida, escolha se pretende participar ou não. Se não quiser, não sofrerá qualquer tipo de penalização.

É inteiramente livre de aceitar ou recusar participar neste estudo. Pode retirar o seu consentimento em qualquer altura, através da notificação ao investigador, ou interrompendo o preenchimento do questionário, sem qualquer consequência, sem precisar de explicar as razões, sem qualquer penalização ou perda de benefícios e sem comprometer a sua relação com o investigador que lhe propõe a participação neste estudo.

O consentimento entretanto retirado não abrange os dados recolhidos e tratados até essa data.

A qualquer momento poderá colocar novas questões ao investigador responsável pelo estudo.

Caso pretenda participar, responda “Sim” à questão que se segue. Ao responder “Sim” a esta questão, está a declarar o seu consentimento informado e a permitir a utilização de dados para este estudo.

Grata pela colaboração. Daniela Rurato – 40327@ufp.edu.pt

Concorda em participar neste inquérito?

<input type="checkbox"/>	Sim
<input type="checkbox"/>	Não

PARTE I – DADOS SOCIODEMOGRÁFICOS

1. Qual é a sua idade?

- 18-23
- 24-29
- 30-35
- 36-41
- 42-47
- 48-53
- 54-59
- 60-65
- 66 ou mais

2. Qual é o seu sexo? (Nota: considera-se que o sexo se refere à anatomia com que a pessoa nasceu)

- Feminino
- Intersexo
- Prefiro não dizer

3. Qual é a sua identidade de género?

- Feminina
- Prefiro não dizer
- Outro. Qual? _____

4. Encontra-se, de momento, a frequentar o Ensino Superior numa Instituição Portuguesa?

- Sim
- Não

4.1. Se sim, qual?

4.2. Qual é o seu Curso? _____

4.3. Encontra-se a obter o grau de:

- Licenciatura
- Mestrado
- Doutoramento

PARTE II – INTERNET, REDES SOCIAIS E DISCURSO DE ÓDIO MISÓGINO

1. Utiliza a Internet?

- Sim
- Não

1.1. Utiliza Redes Sociais?

- Sim
 Não

1.2. Se sim, quais?

- Facebook
 Twitter
 Instagram
 TikTok
 Reddit
 Outro. Qual? _____

2. Conhece o conceito de discurso de ódio *online* misógino?

- Sim
 Não

2.1. O discurso de ódio *online* misógino corresponde a “todas as formas de expressão que incitam, promovem, espalham ou justificam violência, ódio ou discriminação” contra uma mulher ou um grupo de mulheres, ou que as denigre, por razão do seu sexo.

Tendo em conta a definição fornecida, que tipo de comportamentos considera constituir discurso de ódio *online* misógino? (Fersini *et al.*, 2018; Saatchi & Saatchi e Everything Everywhere, 2022).

Assinale todos os exemplos com que concorda de acordo com esta escala:

- discordo completamente (1);
- discordo (2);
- concordo (3);
- concordo completamente (4).

	1	2	3	4
Insultos e utilização de calões com base no género (ex: “cabra”, “cadela”).				
<i>Body-shaming</i> e <i>slut-shaming</i> .				
Estereotipização e objetificação de mulheres.				
Afirmção da superioridade masculina em relação às mulheres, com o intuito de fomentar a desigualdade de género.				
Justificar abuso contra mulheres.				
Assédio sexual e ameaças de violência física e/ou sexual.				
Descrédito e desvalorização das opiniões de mulheres.				
Promoção dos comportamentos supramencionados				

3. Alguma vez testemunhou discursos de ódio *online* de carácter misógino contra outra(s) mulher(es)?

- Sim, no último ano.
 Sim, há mais de um ano.
 Não.

3.1. Alguma vez foi vítima de discursos de ódio *online* de caráter misógino?

- Sim, no último ano.
- Sim, há mais de um ano.
- Não.

3.2. Se sim, em que plataforma(s) testemunhou e/ou foi vítima deste tipo de retóricas?

- Website ou fórum na Internet.
- Rede Social – Facebook
- Rede Social – Twitter.
- Rede Social – Instagram.
- Rede Social – TikTok.
- Rede Social – Reddit.
- Rede Social – Outro. Qual? _____

3.3. De que forma de discurso de ódio *online* misógino, foi vítima/testemunhou? (Fersini *et al.*, 2018; Saatchi & Saatchi e Everything Everywhere, 2022).)

- Insultos e utilização de calões com base no género (ex: “cabra”, “cadela”).
- Body-shaming* e *slut-shaming*.
- Estereotipização e objetificação de mulheres.
- Afirmção da superioridade masculina em relação às mulheres, com o intuito de fomentar a desigualdade de género.
- Justificar abuso contra mulheres.
- Assédio sexual e ameaças de violência física e/ou sexual.
- Descrédito e desvalorização das opiniões de mulheres.
- Promoção dos comportamentos supramencionados.

3.4. Se já foi vítima/testemunhou discursos de ódio *online* de caráter misógino contra si, ou contra outrem, que ação, ou ações, tomou quando confrontada com esse tipo de discurso? (Pode assinalar mais do que uma situação, caso se aplique)

- Ignorei.
- Reportei a situação à plataforma/ website.
- Reportei a situação às autoridades policiais.
- Reportei a situação à Internet Segura.
- Conteí a um amigo/ familiar / conhecido.
- Não sei/ Não me lembro.
- Outro. Qual? _____

PARTE III – IMPACTO

Se testemunhou e/ou foi vítima alguma vez, vítima de discursos de ódio *online* de caráter misógeno, por favor, responda às questões seguintes:

1. Como é que esta situação a fez sentir? (Saatchi & Saatchi e Everything Everywhere, 2022)

- Triste/ desanimada.
- Desconfortável.
- Ansiosa.
- Raiva.
- Indiferente.
- Desprezada/ menosprezada.
- Ameaçada.
- Sozinha/ isolada.
- Pouco segura.
- Pouco confiante.
- Outro. Qual? _____
- Prefiro não dizer.

2. Considera que esta situação teve impacto na sua vida?

- Sim
- Não

2.1. Se sim, classifique as seguintes afirmações de acordo com esta escala (Amnesty International, 2017; Saatchi & Saatchi e Everything Everywhere, 2022):

- discordo completamente (1);
- discordo (2);
- concordo (3);
- concordo completamente (4).

	1	2	3	4
Senti que não posso expressar-me livremente				
Deixei / restringi a publicação de conteúdo opinativo nas minhas Redes Sociais				
Senti que a minha segurança física e/ou a segurança de familiares / amigos foi comprometida				
A minha situação laboral (presente e/ou futura) foi comprometida				
A minha relação com amigos e/ou familiares foi comprometida				
Tive dificuldade em concentrar-me e/ou manter o foco				
Tive problemas com o sono (insónia, hipersónia, pesadelos, etc.)				
Senti dificuldade em tomar decisões				
Experienciei ataques de pânico, ansiedade e/ ou <i>stress</i> persistente				
Senti-me apreensiva quando pensava na Internet ou Redes Sociais				
Senti-me apreensiva quando recebia notificações de Redes Sociais ou e-mails				
Senti-me com baixa auto-estima				
Experienciei mudanças repentinas de humor				
Senti-me indefesa e/ou sem capacidade para responder ao abuso <i>online</i>				
Senti-me isolada e/ou necessidade de me isolar				

Anexo 2 - Pedido de autorização à Comissão de Ética da Universidade Fernando Pessoa

Exma. Senhora Presidente,
Da Comissão de Ética
da Universidade Fernando Pessoa

Eu, Daniela Alexandra Carvalho Rurato, aluna do 3º ano da licenciatura em Criminologia na Universidade Fernando Pessoa, venho por este meio submeter o meu projeto de investigação de forma que este seja avaliado pela Comissão de Ética da Universidade Fernando Pessoa.

A proposta de investigação apresentada no projeto intitulado “Discurso de ódio *online*: avaliação da prevalência de discursos de ódio *online* de carácter misógino contra estudantes universitárias”, visa realizar um estudo para avaliar a prevalência de discursos de ódio *online* misóginos testemunhados por ou perpetrados contra estudantes universitárias, assim como para compreender este fenómeno de vitimação.

Neste âmbito será utilizado um questionário elaborado com o objetivo de determinar a prevalência deste fenómeno e avaliar o seu impacto, da perspetiva das vítimas diretas ou indiretas. Este questionário será preenchido por estudantes do sexo feminino que frequentam o Ensino Superior em Portugal.

Sem mais nenhum assunto de momento, aguardo a autorização.

Porto, 24 de abril de 2023.

A requerente

(Daniela Alexandra Carvalho Rurato)

Anexo 3 - Declaração de Consentimento Informado

DECLARAÇÃO DE CONSENTIMENTO INFORMADO

Designação do Estudo: “Discurso de ódio *online*: avaliação da prevalência de discursos de ódio *online* de carácter misógino contra estudantes universitárias”

Eu, abaixo-assinado, (nome completo do participante no estudo) -----

-----, compreendi a explicação que me foi fornecida acerca da participação na investigação que se tenciona realizar, bem como do estudo em que serei incluído. Foi-me dada oportunidade de fazer as perguntas que julguei necessárias, e de todas obtive resposta satisfatória.

Tomei conhecimento de que a informação ou explicação que me foi prestada versou os objetivos e os métodos. Além disso, foi-me afirmado que tenho o direito de recusar a todo o tempo a minha participação no estudo, sem que isso possa ter como efeito qualquer prejuízo pessoal.

Foi-me ainda assegurado que os registos em suporte papel e/ou digital (sonoro e de imagem) serão confidenciais e utilizados única e exclusivamente para o estudo em causa, sendo guardados em local seguro durante a pesquisa e destruídos após a sua conclusão.

Por isso, consinto em participar no estudo em causa.

Data: ____/____/ 2023

Assinatura do participante no projecto: _____

O Investigador responsável:

Nome:

Assinatura:

(Daniela Alexandra Carvalho Rurato)

Comissão de Ética da Universidade Fernando Pessoa