Ordine internazionale e diritti umani
International Legal Order and Human Rights
Ordenamiento Jurìdico Internacional y Derechos Humanos
Ordre Juridique International et Droits de l'Homme
Diretta da Claudio Zanghì, Lina Panella, Carlo Curti Gialdino
EDITORIALE SCIENTIFICA

SILVIA ANGIOI*

HATE SPEECH E *HATE SPEECH* ONLINE. QUESTIONI DEFINITORIE E PROBLEMI DI REGOLAMENTAZIONE TRA DIRITTO INTERNAZIONALE ED INTERVENTI DEL LEGISLATORE EUROPEO

1. *Introduzione*

La locuzione *hate speech* è utilizzata frequentemente per etichettare e stigmatizzare qualsiasi forma di comunicazione volta a diffondere contenuti diffamatori, discriminanti, denigratori nonché idonei a propagare l'odio e la violenza. Per quanto si tratti di un fenomeno tutt'altro che recente, è però indubbio che abbia assunto in epoca attuale dimensioni e caratteri nuovi, soprattutto per effetto del ruolo svolto, nel settore dell'informazione e della comunicazione, da Internet e dai *social network* in quanto veicoli privilegiati per la diffusione di idee, opinioni e più in generale del pensiero sia individuale che collettivo: quotidianamente, attraverso l'uso di quei sistemi si diffondono i contenuti più vari, utili e virtuosi ma anche denigratori, diffamatori e violenti, veritieri ma anche falsi e fuorvianti.

Proprio le caratteristiche di funzionamento e le potenzialità di Internet hanno contribuito a mutare i termini nei quali il fenomeno dell'*hate speech* si manifesta: l'individuazione di strumenti di contrasto idonei a contenere e minimizzare i rischi connessi con la circolazione non controllata di contenuti definibili, in una parola, *hateful*, ha richiesto e richiede il diretto coinvolgimento degli Internet *provider* (d'ora in avanti *provider*), dato il ruolo primario dagli stessi svolto nell'ambiente digitale[1]. L'operazione, tuttavia, è apparsa ed appare ancora non poco complessa: a fronte del fatto che Internet ha certamente favorito l'affermazione, in qualche modo universalistica e democratica, del diritto alla libertà di espressione, si è posta la necessità di individuare i meccanismi idonei a garantire la tutela di

* Professore associato di Diritto internazionale, Università di Sassari.

[1] Si definiscono in generale Internet *provider* quegli operatori che offrono agli utenti la fornitura di servizi internet e che dunque si qualificano come intermediari tra la rete internet e l'utente finale.

alcuni fondamentali diritti che sono posti a rischio da sistemi di circolazione delle informazioni che, per quanto democratici e globali, sono potenzialmente scevri da forme di controllo. È indubbio, infatti, che il problema della diffusione *online* di messaggi e contenuti diffamatori, violenti e discriminatori ripropone, nella sostanza, una questione di più ampio respiro che è quella, certo non recente, come evidenziato in apertura, relativa al più generale problema dell'*hate speech*: è quanto dire del problema non solo del bilanciamento tra il diritto alla libertà di espressione ed altri diritti egualmente garantiti, ma altresì della necessità di tutelare superiori esigenze di carattere pubblico e dunque della comunità nel suo complesso. Ed è proprio il ruolo svolto dai *provider* nella gestione del sistema di circolazione delle informazioni che conferisce a tale problema tratti nuovi: è apparso infatti chiaro da tempo che la definizione di efficaci misure di contrasto all'*hate speech online* non può prescindere dal coinvolgimento di quegli attori che svolgono nella circolazione di messaggi, informazioni e più in generale di contenuti, un ruolo primario. Ciò ha comportato un mutamento fondamentale di prospettiva, nella misura in cui si è reso necessario superare il tradizionale modello, nel quale le norme che regolamentano l'esercizio della libertà di espressione sono concepite sulla base di una dialettica contrapposizione tra Stato e individuo, per definire un modello nuovo, nel quale al binomio Stato-individuo si affianca un terzo elemento rappresentato dai *provider*.

Alla luce di tali premesse, la presente indagine si propone di analizzare, in un'ottica di diritto internazionale, in che termini e con quali modalità sia stata affrontata innanzitutto la questione più generale dell'*hate speech*, per poi verificare quale impostazione sia stata invece seguita nell'adozione di specifici strumenti – in particolare nel quadro dell'Unione europea – dedicati invece al contrasto al problema dell'*hate speech online* tenuto conto delle sue peculiarità.

Ciò comporta pertanto la necessità di affrontare, in via preliminare, le fondamentali questioni di carattere definitorio e semantico legate al concetto di *hate speech*. A dispetto, infatti, dell'utilizzo frequente di tale espressione, definire contorni e contenuti della fattispecie da un punto di vista giuridico è questione non poco complessa di cui è prova, al di là degli sforzi profusi sul piano internazionale nell'adozione di strumenti di diversa natura, l'eterogeneità che caratterizza l'approccio del legislatore interno rispetto alla tematica e alla conseguente individuazione delle misure di contrasto. Alla definizione dei contenuti dell'*hate speech* si ricollega poi la questione, sopra menzionata, del bilanciamento fra diritti fondamentali che paiono giustapposti sebbene egualmente garantiti dai principali strumenti internazionali in materia di diritti umani: su tale questione il giudice internazionale e gli organi di monitoraggio si sono più volte pronunciati, offrendo importanti indicazioni ma senza tuttavia eliminare completamente le criticità che caratterizzano l'ambito di applicazione delle norme. È solo dopo aver affrontato in che termini si pone il problema dell'*hate speech* in generale, che si potrà meglio considerare quali ulteriori questioni sollevi il fenomeno dell'*hate speech online* e in che modo si stia provvedendo a definire un sistema di regole, complesso e tutt'altro che scevro da rischi, volto a regolamentare l'esercizio della libertà di espressione nel mondo digitale.

2. *Da* genus *a* speciem*:* hate speech *e* hate speech online. *Problemi definitori e questioni semantiche*

Come si è precisato in apertura, l'*hate speech online* rappresenta la manifestazione assunta attualmente da quel fenomeno tutt'altro che recente quale è l'*hate speech*: Quest'ultima

locuzione è comunemente utilizzata per indicare comportamenti che consistono nel veicolare e diffondere – tramite l'utilizzo di strumenti diversi – "messaggi d'odio" e dunque contenuti diffamatori, discriminatori o denigratori, nei confronti di singole persone o gruppi, in particolare minoranze religiose o razziali o altri gruppi minoritari quali omosessuali, transgender, LGBQT etc. Con la medesima locuzione si fa altresì riferimento ad analoghi comportamenti che possono risultare particolarmente pericolosi perché potenzialmente prodromici rispetto alla commissione di crimini internazionali, *in primis*, i crimini di genocidio e persecuzione[2].

L'idoneità di certi messaggi, discorsi o pubblicazioni, a produrre effetti negativi sulla sfera giuridica dei destinatari – è quanto dire dell'impatto su alcuni diritti fondamentali, *in primis* il dritto alla tutela della vita privata, della reputazione e financo dell'integrità fisica e morale – e, nei casi più gravi, a generare forme di violenza e odio è strettamente legata non solo al modo in cui quegli stessi messaggi sono concepiti ma anche alle potenzialità proprie degli strumenti utilizzati per veicolare, diffondere e propagare certi contenuti. Da questo punto di vista non vi è dubbio che rispetto alle forme che il fenomeno dell'*hate speech* ha tradizionalmente assunto, la versione più attuale dello stesso, rappresentata proprio dal cosiddetto *hate speech online*, sia dotata di ben altre potenzialità: come infatti è stato efficacemente osservato già diversi anni orsono, è la stessa coscienza pubblica che oggigiorno si forma non nei luoghi di aggregazione, siano essi le pubbliche vie o piazze o parchi, ma proprio sul *web* attraverso forum di discussione e diffusione di messaggi e idee[3]. Internet, inoltre, diversamente dai tradizionali strumenti, quali l'editoria, il cinema e la televisione, si presta a veicolare in maniera immediata e amplissima i contenuti più diversi, al di fuori, teoricamente e potenzialmente, di qualsiasi forma di controllo a monte che preceda la diffusione di quei medesimi contenuti: in questo senso appare un sistema tanto efficace e rapido quanto potenzialmente dannoso. Proprio in conseguenza dell'utilizzo di Internet e dei servizi da esso offerti, l'*hate speech* ha assunto dimensioni tali da rendere sempre più pressante l'esigenza di individuare adeguati strumenti di contrasto: è evidente che rispetto a tale operazione si renderebbe necessaria quella, preliminare, che consiste nella definizione e qualificazione del fenomeno. Perseguire l'obiettivo, tuttavia, appare tutt'altro che semplice, per diverse ragioni. Ricondurre certe espressioni, idee o concetti ad una categoria, ben definita e definibile come *hate speech*, non sempre è agevole e tantomeno automatico: innanzitutto perché l'utilizzo di un certo tipo di linguaggio può essere in alcuni casi ambiguo e ciò può accadere sia perché un termine può avere significati differenti nella lingua in cui viene utilizzato, sia perché le stesse espressioni e la portata denigratoria o offensiva delle medesime possono essere percepite diversamente a seconda del contesto[4]; in secondo luogo perché in alcuni ambiti – emblematico è proprio il caso dell'*hate speech online* – è frequente l'utilizzo di termini o formule denigratorie o odiose "in codice" che, non essendo palesi sono più difficilmente individuabili e classificabili. Il problema poi può assumere una diversa dimensione quando certi termini o formule vengono utilizzati all'indirizzo non del singolo

[2] Non potranno, in questa sede, essere affrontate le questioni relative al nesso esistente tra *hate speech* e crimini internazionali che, per la loro complessità, richiederebbero ben altro spazio oltre quello consentito.

[3] Cfr. L. SHAW, *Hate Speech in Cyberspace: Bitterness without Boundaries,* in *Notre Dame Journal of Law, Ethics & Public Policy*, 2012, p. 279 ss.

[4] Cfr. ad esempio A. BROWN, A. SINCLAIR, *The Politics of Hate Speech Laws*, New York, 2020, p. 1, che richiamano i risultati condotti da un sondaggio condotto negli Stati Uniti da cui è risultato che, mentre tra gli elettori del partito democratico solo il 14% riteneva che fosse moralmente accettabile esprimere concetti o usare epiteti che potessero essere offensivi o denigratori per particolari gruppi religiosi o razziali, tra gli elettori del partito repubblicano la percentuale si attestava intorno al 24%.

individuo ma di un'intera categoria di soggetti così come una specifica valenza politica può assumere l'utilizzo di un linguaggio offensivo e diffamatorio che ha quali destinatari specifici gruppi di popolazione – minoranze, immigrati o categorie particolari – da parte di esponenti politici e autorità di governo. Se questo è il quadro di riferimento, ben si comprende perché gli strumenti – sia di *soft law* che vincolanti – dedicati al problema dell'*hate speech* in generale, nonché a quello più specifico dell'*hate speech online* provvedono a definire il fenomeno in maniera funzionale rispetto alle finalità proprie di ciascun specifico atto, senza tuttavia pervenire, nonostante le evidenti analogie riscontrabili nelle formule utilizzate, ad una definizione di *hate speech* che possa dirsi univoca e generalmente accolta. Si comprende altresì perché, strumenti che trattano tematiche specifiche, connesse con il problema della discriminazione nelle sue diverse forme e in particolare quelle relative alla criminalizzazione degli atti di natura razzista o xenofoba, pur non essendo specificamente intitolati all'*hate speech*, ad esso inevitabilmente si ricollegano, nella misura in cui i fenomeni oggetto della regolamentazione possono essere considerati come una conseguenza o una materiale manifestazione dell'*hate speech*.

3. *Segue. L'* hate speech *negli strumenti internazionali, tra* soft law, *strumenti vincolanti e tentativi di regolamentazione da parte del legislatore europeo*

Fatta la premessa di cui sopra, si intende ora verificare se sia possibile individuare perlomeno i tratti caratterizzanti del fenomeno solitamente definito *hate speech*. In quest'ottica, saranno analizzati, senza una pretesa di esaustività, alcuni significativi strumenti internazionali, sia di *soft law* che vincolanti, che, o sono all'*hate speech* specificamente dedicati, ovvero trattano tematiche, come sopra anticipato, che al medesimo fenomeno sono comunque collegate.

Un primo necessario riferimento, per quanto concerne le questioni connesse con la definizione dell'*hate speech* va fatto, innanzitutto, alla Raccomandazione n. 20 adottata nel 1997 dal Comitato dei Ministri del Consiglio d'Europa (CoE)[5]. Nel testo della Raccomandazione – dedicata in particolare al problema dell'*hate speech* diffuso attraverso i media – compare una definizione che, nei suoi contenuti essenziali, può dirsi essere stata ripresa ed ampliata in diversi strumenti adottati successivamente, sia in seno allo stesso CoE, sia nell'ambito di altre organizzazioni internazionali, nel cui contesto sono stati adottati strumenti che saranno analizzati nel prosieguo.

La Raccomandazione definisce l'*hate speech* come «all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin»[6]. Il primo elemento, che, in maniera evidente, appare strutturalmente connesso all'*hate speech*, è quello del rapporto tra discriminazione ed *hate speech*: il pregiudizio ideologico ancorato ad elementi di tipo razziale, etnico, religioso o nazionale rappresenta infatti il meccanismo attivatore dell'*hate speech* ed è attraverso quest'ultimo che quel pregiudizio ha modo di manifestarsi e propagarsi. Ciò è peraltro confermato dal fatto che la

[5] Cfr. Consiglio d'Europa (CoE), *Recommendation No. R(97)20* del Comitato dei Ministri del 30 ottobre 1997, *"Hate Speech"*, file:///C:/Users/utente/Downloads/Rec(97)20%20(3).pdf.
[6] CoE, *Recommendation No. R(97)20*, cit., scope.

stessa Raccomandazione precisa che il contrasto all'*hate speech* è da intendersi come parte di una più ampia strategia di lotta alla discriminazione e all'intolleranza[7]. Ulteriore tratto caratterizzante può individuarsi nella condotta nella quale l'*hate speech* si sostanzia: la Raccomandazione individua tale condotta nell'incitamento all'odio razziale, all'intolleranza e alla discriminazione. L'incitamento è dunque indicato quale elemento costitutivo dell'*hate speech* e in questo senso è associato ad altre attività similari quali *promotion*, *spread* e *advocacy*: queste ultime, dunque, acquistano la medesima valenza per quanto non possano considerarsi esattamente equivalenti, né possano dirsi, i termini utilizzati, dei perfetti sinonimi.

Accanto alla definizione, alcune indicazioni contenute nella Raccomandazione servono ulteriormente a qualificare l'*hate speech*, innanzitutto perché chiariscono il rapporto tra *hate speech* e libertà di espressione. In questo senso si comprende infatti il richiamo, rivolto agli Stati, perché adottino un quadro normativo idoneo non solo a perseguire *l'hate speech* – tramite il ricorso a misure diverse, di natura penale, civile e amministrativa – ma altresì a garantire un adeguato bilanciamento tra l'esigenza di contrasto all'*hate speech* e la salvaguardia del diritto alla libertà di espressione[8]. La Raccomandazione evidenzia infatti che l'articolo 10 della Convenzione europea dei diritti umani (CEDU) – norma fondamentale posta a presidio della libertà di espressione – non tutela tale diritto nei casi in cui lo stesso venga esercitato con modalità idonee a determinare «the destruction of the rights and freedoms laid down in the Convention or at their limitation to a greater extent than provided therein». Stabilire quali manifestazioni del pensiero e delle opinioni siano idonee a produrre tale effetto, assume una rilevanza centrale perché serve non solo a definire i limiti entro i quali la libertà di espressione può essere garantita, ma altresì ad individuare i casi in cui si produce un fenomeno qualificabile come *hate speech*.

La Raccomandazione poi contiene delle indicazioni utili ad individuare quello che dovrebbe essere il corretto approccio dello Stato ai casi di *hate speech*: si afferma infatti un principio – che verrà ribadito in termini ancora più chiari nei diversi strumenti adottati successivamente, sui quali ci si soffermerà a breve – secondo il quale, poiché «the imposition of criminal sanctions generally constitutes a serious interference» con la libertà di espressione, il legislatore dovrebbe riconoscere all'autorità giudiziaria un sufficiente margine di discrezionalità nel trattare i casi di *hate speech*; tale margine dovrebbe consentire al giudice di tenere in adeguata considerazione il diritto alla libertà di espressione della persona accusata, ferme restando, inoltre, le garanzie che devono essere fornite, laddove venga comminata una sanzione penale, quanto al pieno rispetto del principio di proporzionalità[9].

Nel solco tracciato dalla Raccomandazione n. 20/1997 del CoE, si collocano, secondo una logica di continuità, altri atti adottati più di recente in materia di *hate speech*, sia nel quadro del CoE, sia nel più ampio contesto delle Nazioni Unite. Sotto il primo profilo vengono in

---

[7] Tale impostazione trova conferma anche in altri atti adottati, per esempio, nell'ambito delle Nazioni Unite, in particolare dall'Assemblea generale (cfr. UNGA Res. 60/143 del 16 dicembre 2005; UNGA Res. 62/142 del 18 dicembre 2007; UNGA Res. 64/147 del 18 dicembre 2009; UNGA Res. 66/143 del 29 marzo 2012) e dal Consiglio dei diritti umani (A/HRC/RES/16/18 del 12 aprile 2011 e A/HRC/RES/19/25 del 10 aprile 2012, intitolate *Combating intolerance, negative stereotyping and stigmatization of, and discrimination, incitement to violence, and violence against persons based on religion or belief*; ris. 18/15 del 14 ottobre 2011 - spec. par. 8 - intitolata *The incompatibility between racism and democracy*). Nel testo di tali documenti l'*hate speech* assume infatti rilievo rispetto ai fenomeni ai quali quegli atti sono primariamente dedicati – razzismo, xenofobia, discriminazione, rigurgiti neonazisti e negazionismo – in quanto rappresenta una fra le diverse forme attraverso cui quei medesimi fenomeni si manifestano.

[8] CoE, *Recommendation No. R(97)20*, cit., principle 2.

[9] *Ibid.*, principle 5.

esame la Raccomandazione n. 16/2022[10], nonché la Raccomandazione n. 15 sul contrasto all'*hate speech*, adottata nel 2016 dalla Commissione europea contro il razzismo e l'intolleranza (ECRI)[11]; sotto il secondo profilo, rilevano invece la *General Recommendation* n. 35 del Comitato sull'eliminazione della discriminazione razziale (CERD) sul *racist hate speech*[12] ed il più recente *United Nations Strategy and Plan of Action on Hate Speech* adottato nel 2019[13].

Dall'esame degli strumenti menzionati può rilevarsi come anche le definizioni di *hate speech* in essi contenute evidenzino chiaramente – in linea con la definizione di *hate speech* contenuta nella Raccomandazione CoE n. 20/1997 – l'esistenza di un nesso strutturale tra *hate speech* e discriminazione; alla base dell'*hate speech* si individua infatti il pregiudizio culturale e/o ideologico su base razziale, religiosa, etnica, o sessuale. Ciò si ricava senz'altro dalla lettura delle definizioni – che appaiono sotto questo profilo del tutto simili – contenute sia nella Raccomandazione CoE n. 16/2022, sia nella Raccomandazione n. 15 dell'ECRI[14]; altrettanto può dirsi per l'*UN Strategy and Plan of Action* che definisce l'*hate speech* come «any kind of speech, writing, behaviour that attacks or uses pejorative or discriminatory language with reference to a person or group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factors».

L'altro tratto che accomuna gli ultimi documenti menzionati e la Raccomandazione CoE n. 20/1997 è l'enfasi posta sulla necessità di salvaguardare l'esercizio del diritto a manifestare pensieri ed opinioni: a tale esigenza si ricollega poi quella di adottare, nel contrasto a tale fenomeno, un approccio che faccia perno su una gamma di misure di diversa natura e severità e dunque tenga conto sia della gravità della condotta sia dell'impatto che la stessa è suscettibile di produrre non solo sulla sfera giuridica di altri soggetti ma anche sul contesto sociale.

Quanto poi all'individuazione delle condotte, mentre negli strumenti adottati in senso al CoE la definizione di *hate speech* utilizzata mantiene fermo il collegamento tra l'*hate speech* e l'*incitement* – nonché l'*advocacy* e la *promotion*, allo stesso assimilate – altrettanto non può dirsi per quanto concerne gli strumenti adottati nell'ambito delle Nazioni Unite, cioè l'UN Strategy

---

[10] CoE, *Recommendation CM/Rec (2022)16* del Comitato dei Ministri del 20 maggio 2022, *"Hate Speech"*.

[11] Commissione europea contro il razzismo e l'intolleranza (ECRI), *General Policy Recommendation No. 15* dell'8 dicembre 2015, *Combating Hate Speech*.

[12] Cfr. CERD, *General Recommendation No. 35* del Comitato delle Nazioni Unite per l'eliminazione della discriminazione razziale (CERD) del 26 settembre 2013, *Combating Racist Hate speech*, CERD/C/GC/35. Va peraltro evidenziato che si è scelto di esaminare la *general Recommendation* n. 35, sia perché è lo strumento adottato più di recente sul tema, sia perché ben più ampio nei contenuti rispetto ad alcune *General Recommendation* adottate in precedenza dedicate all'appliazione dell'art. 4 della ICERD: cfr. *General Recommendation*, n. 7 del 23 agosto 1985, *Implementation of Article 4. Legislation to Eradicate Racial Discrimination*, A/40/18 e n. 15 del 23 marzo 1993, *Article 4* A/48/18.

[13] United Nations Strategy and Plan of Action on Hate speech, https://www.un.org/en/genocideprevention/documents/advising-and mobilizing/Action_plan_on_hate_speech_EN.pdf.

[14] Nel caso della Raccomandazione n. 16/2022, l'*hate speech* è definito come «all types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as "race", colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation» (punto 1). Ai sensi invece della Raccomandazione n. 15 dell'ECRI, l'*hate speech* è da intendersi come «the advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons (…) on the ground of "race", colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status».

and Plan of Action e la *General Recommendation* n. 35 sul *racist hate speech* del CERD. Per quanto concerne il primo, si può notare come da un lato la definizione di *hate speech* utilizzata non contenga alcun riferimento all'incitamento; dall'altro, come malgrado tale omissione, proprio tale concetto venga poi espressamente richiamato per evidenziare non solo il fatto che il diritto internazionale, pur non vietando l'*hate speech* in sé e per sé, vieta l'incitamento alla violenza e all'odio, ma altresì per precisare che l'incitamento rappresenta la forma più grave di *hate speech*, in quanto idoneo a produrre conseguenze sul piano pratico e fattuale, generando atti di violenza e d'odio[15].

Anche la *General Recommendation* n. 35 del CERD sul *racist hate speech* possiede dei tratti peculiari, rilevabili, in particolare, per quel che concerne la specifica configurazione del fenomeno ed il suo inquadramento nel sistema delle norme e delle finalità della ICERD. È in quest'ottica che si spiega innanzitutto il significato attribuito al termine *racist* col quale si qualifica il fenomeno dell'*hate speech*: ad avviso del Comitato, infatti, il termine "razziale" deve intendersi nel contesto della *General Recommendation* – così come nel contesto della ICERD – in senso esteso; ne consegue che l'aggettivo *racist* non serve a connotare l'*hate speech* in maniera rigida, posto che, come viene evidenziato, tale fenomeno «can take many forms and is not confined to explicitely racial remarks»[16].

Quanto all'individuazione "of expressions which constitute hate speech" è alle norme della ICERD che occorre fare riferimento[17]: rilevano, a questo proposito, innanzitutto l'articolo 1, che definisce l'ampiezza e l'ambito di applicazione della locuzione *racial discrimination*[18], ma soprattutto l'articolo 4, che individua le condotte rispetto alle quali gli Stati hanno l'obbligo di adottare specifiche misure di contrasto. È quest'ultima norma, infatti, che contiene gli elementi utili a chiarire che cosa debba intendersi per *racist hate speech*. L'articolo 4 menziona espressamente, ponendole sul medesimo piano, attività diverse, consistenti nella diffusione di idee basate sulla superiorità o sull'odio razziale, nell'incitamento alla discriminazione razziale, negli atti di violenza e nell'incitamento a tali atti ed infine nel sostegno a qualsiasi atto di natura razzista[19]: il Comitato ha dunque evidenziato che «Racist

---

15 Si legge che «incitement is a very dangerous form of speech because it explicitly and deliberately aims at triggering discrimination, hostility and violence, which may also lead to or include terrorism or atrocity crimes».

16 CERD, *General Recommendation No. 35*, cit., par. 7.

17 *Ibid*., par. 5.

18 Ai sensi dell'art. 1 «the term "racial discrimination" shall mean any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life». In linea con l'interpretazione ampia della locuzione *racial discrimination* e con l'altrettanto ampia interpretazione delle finalità della ICERD, il Comitato ha richiamato i numerosi atti, adottati con riferimento al tema della discriminazione nelle sue diverse forme, nei quali è stato evidenziato sistematicamente il nesso tra *hate speech* e discriminazione ed il ruolo che lo stesso svolge nell'attivare il meccanismo della discriminazione. Cfr. tra le diverse *General Recommendation* adottate dal Comitato, la n. 25 del 20 marzo 2000, *On gender-related dimensions of racial discrimination*; la n. 27 del 16 agosto 2000, *On discrimination against Roma* e la n. 34 del 3 ottobre 2011, *On racial discrimination against people of African descent*.

19 La norma pone in capo agli Stati l'obbligo specifico di qualificare come reati tutte le condotte che si traducono nella «dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts». Sul punto si tornerà più avanti: cfr. par. 4. Sul tema del *racist hate speech* si rimanda I. MAITRA, M.K. MCGOWAN, *On Racist Hate Speech and the Scope of a Free Speech Principle*, in *Canadian Journal of Law & Jurisprudence*, 2010, p. 343 ss.; C. WEST, *Words That Silence? Freedom of Expression and Racist Hate Speech*, in I. MAITRA, M.K. MCGOWAN, *Speech and Harm: Controversies Over Free Speech*, Oxford 2012, p. 222 ss.; P. THORNBERRY, *International Convention on the Elimination of All Forms of Racial Discrimination: the prohibition of "racist hate speech"*, in T. MCGONAGLE, Y. DONDERS (eds.), *The United Nations and Freedom of*

hate speech addressed in Committee practice has included all the specific speech forms referred to in article 4 directed against groups recognized in article 1 of the Convention»[20]. Se ne deduce, pertanto, che l'*hate speech*, anche considerato nella specifica prospettiva della ICERD, si traduce, ancora una volta, nella diffusione di idee alimentate dal pregiudizio (non solo) razziale, nonché nell'incitamento all'odio e alla violenza, nell'incitamento al compimento di atti violenti per motivi razziali ovvero nel sostegno fornito al compimento degli stessi.

Un'ultima annotazione sembra opportuna, sempre con riferimento all'articolo 4 della ICERD: la norma, come si preciserà nel paragrafo che segue, prevede espressamente che le condotte in esso elencate – che sono quelle cui si è sopra riferimento – devono, dallo Stato, essere qualificate come reati punibili per legge. A tale proposito nella *General Recommendation* n. 35 è contenuta però un'ulteriore indicazione, che serve a ribadire un principio già affermato dalla Raccomandazione CoE n. 20/1997 e poi richiamato anche dagli strumenti specificamente dedicati alla criminalizzazione degli atti di natura razzista e xenofoba che ci si appresta ad esaminare: si intende fare riferimento al fatto che anche il Comitato, pur tenendo conto del portato dell'articolo 4 e degli obblighi da esso imposti agli Stati parte della ICERD, ha però evidenziato che la criminalizzazione del *racist hate speech* deve essere riservata ai casi più gravi e che laddove sia ritenuto necessario fare ricorso alla misura penale, l'applicazione di quest'ultima «should be governed by principles of legality, proportionality and necessity»[21].

Proprio il tema della criminalizzazione delle condotte consistenti nella diffusione di idee fondate sul pregiudizio razziale, nonché sull'incitamento all'odio e alla discriminazione, consente a questo punto di individuare la linea di continuità che collega gli strumenti esaminati finora con altri, cui si è più volte fatto riferimento, che, pur non essendo specificamente intitolati all'*hate speech*, possono ritenersi applicabili a tale fenomeno nella misura in cui l'*hate speech* può essere qualificato, sulla base delle norme in essi contenute, come atto di natura razzista o xenofoba. In quest'ottica, assumono rilievo sia il primo Protocollo addizionale alla Convenzione del Consiglio d'Europa contro la criminalità informatica (il Protocollo), che fu adottato in seno al CoE nel 2001 in materia di criminalizzazione degli atti di natura razzista e xenofoba commessi a mezzo di sistemi informatici[22], sia la decisione quadro 2008/913/GAI (decisione quadro) dell'Unione europea sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale[23]. Merita innanzitutto evidenziare il richiamo, presente nella prima parte di entrambi gli strumenti, alla centralità della libertà di espressione, quale fondamentale parametro di riferimento per l'applicazione delle norme volte a contrastare gli atti di natura razzista, xenofoba o discriminatoria in essi contenute[24]. In secondo luogo, occorre considerare che sia il Protocollo che la decisione

Expression and Information, Cambridge, 2015, pp. 121-144. T. MCGONAGLE, *General Recommenation 35 on combating racist hate speech*, in D. KEANE, A. WAUGHRAY (eds.), *Fifty Years of the International Convention on the Elimination of All Forms of Racial Discrimination: A living Instrument*, Manchester, 2017, p. 246 ss.; L. MANCA, *Sul contrasto al* racial hate speech *nella prassi del Comitato delle Nazioni Unite per l'eliminazione della discriminazione razziale*, in *Ordine internazionale e Diritti umani*, 2018, p.457 ss.

[20] CERD, *General Recommendation No. 35*, cit., par. 6.

[21] *Ibid.* par. 12.

[22] Cfr. CoE, *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer* system, Strasburg, 28.01.2003. Si tratta del primo Protocollo addizionale alla Convenzione di Budapest sui crimini informatici adottata in seno al Consiglio d'Europa il 23 novembre 2001.

[23] Decisione-quadro 2008/913/GAI, GUUE L328/55 del 6.12.2008.

[24] Cfr. il preambolo di entrambi gli strumenti. Nel caso della decisione quadro 2008/913/GAI, è poi l'art. 7 ad affermare testualmente che «L'obbligo di rispettare i diritti fondamentali e i fondamentali principi giuridici

quadro sono stati concepiti in un'ottica penalistica, circostanza, quest'ultima, che serve a spiegare l'enfasi posta sulla necessità che a determinate condotte si debba guardare tenendo conto della loro gravità, perché è in ragione di quest'ultima che si ritiene possa essere giustificata l'eventuale adozione di sanzioni penali. Da questo punto di vista, è innanzitutto il Protocollo a prevedere che gli Stati debbano adottare misure legislative destinate a criminalizzare una serie di condotte, di diversa natura e comunque caratterizzate dal pregiudizio razziale o xenofobo, messe in atto attraverso l'uso del computer. Fra queste è menzionata innanzitutto la diffusione di materiale razzista e xenofobo (art. 3) e, con riferimento al concetto di "materiale razzista e xenofobo" il Protocollo precisa che con questo deve intendersi «any written material, any image or any other representation of ideas or theories which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, colour, descent or national or ethnic origin, as well as religion if used as a pretext for any of these factors»[25]. A tale condotta si aggiungono le altre indicate negli articoli seguenti, dalle minacce di compiere "serious criminal offences" per motivi di natura razzista o xenofoba (art. 4), agli insulti (art. 5), al diniego, supporto o approvazione del genocidio o dei crimini contro l'umanità (art. 6), Trattandosi di condotte che sono messe in atto tramite l'uso del computer, è evidente che vi è una fondamentale componente che le accomuna, ravvisabile nel fatto che tutte le attività considerate integrano forme di "comunicazione" che si connotano per i contenuti discriminatori, razzisti e/o xenofobici e, per quanto concerne specificamente l'attività di diffusione di materiali con contenuti razzisti, discriminatori e xenofobi, implicano forme di incitamento alla discriminazione, all'odio e alla violenza.

Quanto poi alla criminalizzazione delle condotte, il Protocollo dà indicazioni diverse a seconda delle attività considerate. Con riferimento in particolare alla diffusione di materiale razzista e xenofobo, l'articolo 3 prevede ad esempio che sia lasciata allo Stato la facoltà di decidere se rendere penalmente perseguibile la condotta nei casi in cui questa «is not associated with hatred and violence»; in maniera similare, l'articolo 5 prevede che con riferimento all'attività che consiste nel «insulting publicly, through a computer system, (i) persons for the reason that they belong to a group distinguished by race, colour, descent or national or ethnic origin, as well as religion», lo Stato possa sia stabilire che la condotta così descritta deve avere come effetto quello di esporre le persone, verso cui gli insulti sono diretti, all'odio, al disprezzo e al ridicolo, sia stabilire di non ritenersi vincolato, in tutto o in parte, dalla norma in esame, astenendosi dunque dal criminalizzare le condotte in oggetto. Altrettanto deve rilevarsi per quanto concerne il diniego, la banalizzazione o la giustificazione del genocidio e dei crimini contro l'umanità: l'articolo 6 infatti prevede che lo Stato possa decidere che tali condotte sono qualificate come reati nel caso in cui siano poste in essere con l'intento di incitare all'odio e alla violenza nei confronti di individui o gruppi di individui in ragione della loro appartenenza etnica, razziale, religiosa o nazionale.

Sembra dunque che anche nel caso del Protocollo si sia voluta affermare l'idea che si debba – o si possa, tenuto conto del margine di manovra lasciato allo Stato – distinguere tra atti, che comunque possono ricondursi in quella magmatica categoria definibile *hate speech*, ma che sono diversamente qualificabili in ragione della loro gravità.

---

sanciti dall'articolo 6 del trattato sull'Unione europea, tra cui la libertà di espressione e di associazione, non è modificato per effetto della presente decisione quadro» (par 1) e che «La presente decisione quadro non ha l'effetto di imporre agli Stati membri di prendere misure che siano in contrasto con i principi fondamentali riguardanti la libertà di associazione e la libertà di espressione….» (par. 2).

[25] *Additional Protocol*, cit., art. 2.

L'idea cui si è appena fatto riferimento sarà ribadita, come ci si accinge ad evidenziare, anche nel contesto della decisione quadro 2008/913/GAI. Quest'ultima, nel definire i reati di stampo razzista, si focalizza innanzitutto sulla condotta che consiste nell'«istigazione pubblica[26] alla violenza e all'odio nei confronti di individui o gruppi di persone sulla base del loro colore e/o della loro appartenenza etnica, razziale o religiosa»[27]; ad ulteriore specificazione precisa che l'istigazione deve essere qualificata come reato anche quando avvenga «mediante la diffusione e la distribuzione pubblica di scritti, immagini o altro materiale» (par. b). In maniera similare al Protocollo, la decisione quadro allarga poi lo spettro delle fattispecie penalmente perseguibili inserendo fra esse il negazionismo, la trivializzazione o la minimizzazione del genocidio, dei crimini di guerra e dei crimini contro l'umanità[28].

Per quanto poi concerne il problema della criminalizzazione di tali condotte, la decisione quadro attribuisce allo Stato la facoltà di decidere se rendere penalmente sanzionabili solo «i comportamenti atti a turbare l'ordine pubblico o che sono minacciosi, offensivi o ingiuriosi»[29]. Diversamente dal Protocollo che contiene, come si è rilevato, indicazioni diverse a seconda delle condotte considerate, la decisione quadro utilizza invece un unico parametro che risulta applicabile con riferimento ad attività, diverse dal punto di vista materiale, ma caratterizzate dalla natura discriminatoria, razzista e/o xenofoba. Allo Stato è dunque lasciato un margine di discrezionalità che può produrre effetti sotto diversi punti di vista. Innanzitutto, sul piano legislativo, perché consente al legislatore di qualificare come reati sia quelle condotte che consistono nell'incitamento alla discriminazione, all'odio e alla violenza perché potenzialmente idonee a dar luogo ad atti di violenza e odio, sia quelle condotte la cui idoneità a turbare l'ordine pubblico è strettamente legata alla loro specifica

---

[26] Merita segnalare che l'utilizzo di termini differenti – incitamento/istigazione – nelle lingue in cui la decisione quadro è stata redatta, consente di individuare dei margini entro i quali il legislatore e il giudice nazionale possono orientarsi quanto alla qualificazione dell'*hate speech* come atto/comportamento penalmente rilevante e dunque perseguibile. Nel testo francese, spagnolo e inglese compaiono i termini *incitement*, *incitación* e *inciting* rispettivamente, mentre nel testo italiano si utilizza il termine istigazione. Per quanto specificamente concerne l'ordinamento italiano, è noto che proprio il concetto di istigazione presenta aspetti problematici legati all'uso legislativo di termini affini, ma non considerabili come sinonimi, quali istigare, indurre e incitare. Con riferimento in particolare al problema dell'istigazione a delinquere per motivi di discriminazione razziale o religiosa, l'art. 604-*bis* c.p. prevede la criminalizzazione di condotte diverse che vanno dalla propaganda di idee fondate sulla superiorità razziale o etnica, all'istigazione a commettere (o commissione di) atti di discriminazione per motivi razziali, etnici o religiosi (comma 1a), all'istigazione a commettere violenza o atti di provocazione alla violenza (comma 1b) alla partecipazione o supporto ad organizzazioni e associazioni che abbiano tra i propri fini l'incitamento alla discriminazione o alla violenza per motivi etnici, razziali, nazionali o religiosi (comma 2). Sul più generale tema degli *hate crimes* si rimanda a J. GARLAND, N. CHAKRABORTI, *Divided by a Common Concept? Assessing the Implications of Different Conceptualizations of Hate Crime in the European Union*, in *European Journal of Criminology*, 2012, p. 38 ss.; N. CHAKRABORTI, J. GARLAND, *Hate Crime. Impact, Causes, and Responses*, Los Angeles-London, 2015;
J. GARLAND, C. FUNNELL, *Defining Hate Crime Internationally: Issues and Conundrum*, in J. SCHWEPPE, M.A. WALTERS, *The Globalization of Hate: Internationalizing Hate Crime?* Oxford, 2016, p. 15 ss.; L. GOISIS, *"Hate crimes": perché punire l'odio. Una prospettiva internazionale, comparatistica e politico-criminale*, in *Rivista italiana di diritto e procedura penale*, 2018, p. 2010 ss.; E. KIRS, *Hate crimes and international institutions: A literature review*, in *Hungarian Journal of Legal Studies*, 2020, p. 285 ss.

[27] Decisione quadro 2008/913/GAI, art. 1.

[28] *Ibid.*, parr. c,d. La decisione-quadro fa espressamente riferimento sia al dettato degli articoli 6, 7 e 8 dello Statuto della Corte penale internazionale, sia all'articolo 6 dello Statuto del Tribunale militare internazionale, allegato all'accordo di Londra dell'8 agosto 1945.

[29] *Ibid.* art. 1, par. 2.

connotazione ideologica, al di là, dunque, del verificarsi di episodi di odio e/o violenza[30]. In secondo luogo, quel margine di discrezionalità produce effetti sul piano giurisdizionale, perché il giudice potrà valutare se, ed in che misura, una determinata condotta possa essere considerata suscettibile di turbare l'ordine pubblico o avere una natura offensiva e ingiuriosa e possa pertanto richiedere l'imposizione di una sanzione penale a carico dell'autore.

Restando in tema di criminalizzazione degli atti di natura razzista e xenofoba, sembra infine opportuna qualche rapida considerazione in merito ad alcune osservazioni contenute nel testo delle linee guida, adottate nel 2024 nell'ambito dell'OSCE che, se da un lato sono focalizzate sul tema degli *hate crime*, dall'altro affrontano tale tema considerando anche il punto di intersezione tra *hate crimes* e *hate speech*[31]. In generale, può rilevarsi che il tema degli *hate crime* è più ampio rispetto al tema dell'*hate speech* ma occorre verificare in che termini sia stato, dai redattori delle linee guida, individuato un possibile collegamento o, meglio, una possibile sovrapposizione, fra le due fattispecie. Per quanto infatti le linee guida muovano da una prospettiva differente perché il focus è posto sul concetto di crimine motivato dall'odio su base etnica, razziale, religiosa e nazionale – crimine che evidentemente può essere commesso a prescindere da una qualsiasi preliminare manifestazione del pensiero qualificabile come *hate speech* – ciò che merita sottolineare è il fatto che all'*hate speech* è comunque riconosciuto un ruolo come elemento che potenzialmente può attivare la commissione di un reato. Le linee guida evidenziano che, sebbene si tratti, anche concettualmente, di fenomeni diversi, il problema del distinguere l'*hate crime* dall'*hate speech* si pone perché si registra una tendenza del legislatore interno a criminalizzare l'*hate speech*, in conformità a quanto richiesto dalle norme internazionali, nelle forme più diverse: sia quando per esempio in esso si ravvisa una forma di incitamento all'odio e alla violenza (articolo 20 del Patto), sia quando assume i tratti dell'ideologia negazionista o di quella diretta a sminuire la portata del genocidio e dei crimini internazionali (decisione quadro). La conseguenza è che qualificare la condotta, in ciascuno specifico caso, con precisione e ricondurla alle fattispecie previste come penalmente perseguibili, può risultare tutt'altro che semplice. Mentre infatti nel caso dell'*hate crime* la condotta posta in essere sarebbe qualificabile come reato anche nell'ipotesi in cui venisse a mancare quel pregiudizio che invece muove l'autore e che dunque permette di qualificare il reato come *hate crime*, nel caso dell'*hate speech*, ci si trova invece, ad avviso dei redattori delle linee guida, dinanzi ad un cosiddetto *inchoate crime*[32]: ciò significa che

---

[30] In questa logica rientrano evidentemente anche le norme, presenti in alcuni ordinamenti nazionali che attribuiscono rilevanza penale all'apologia del nazismo o del fascismo. È il caso, innanzitutto, del Codice penale tedesco (art. 130, c. 4) e dell'ordinamento italiano; quest'ultimo, come è noto, vieta e considera penalmente perseguibile l'apologia del fascismo (legge "Scelba" del 1952, art. 4 e, sulla base dell'interpretazione dell'art. 1, anche ai sensi della legge n. 205/1993, c.d. legge "Mancino"). È noto, peraltro, che un disegno di legge, attualmente fermo in Senato – il c.d. disegno "Fiano" – prevede l'introduzione nel Codice penale di un articolo, l'art. 293-*bis*, specificamente dedicato al reato di propaganda del regime fascista.

[31] OSCE, *Hate Crimes Prosecution and the Intersection of Hate Crime and Criminalized "Hate Speech": A Practical Guide*, Warsaw, 2024. Va peraltro evidenziato che le line guida sono state adottate con l'intento di completare e integrare il contenuto delle linee guida adottate, sempre in tema di *hate crime* nel 2022 (cfr. OSCE, *Hate Crime Laws: A Practical Guide*, 2 ed., 2022).

[32] *Hate Crimes Prosecution*, cit. p. 25. Le linee guida richiamano a questo proposito le indicazioni date dall'ufficio dell'Alto Commissario delle Nazioni Unite per i diritti umani in merito al tema dell' "incitement to hatred" (cfr. OHCHR, *One-pager on "incitement to hatred"*, https://www.ohchr.org/sites/default/files/Rabat_threshold_test.pdf.). Tali indicazioni riprendono, a loro volta, quelle contenute nel "Rabat Plan of Action" adottato il 4-5 ottobre 2012 da un gruppo di esperti convocato sotto l'egida dell'Alto Commissario delle Nazioni Unite ed incaricato di redigere i principi applicabili per garantire un adeguato bilanciamento tra tutela della libertà di espressione e divieto di incitamento all'odio e

perché l'*hate speech* sia considerato penalmente rilevante non è necessario che produca come risultato la commissione di un reato, per cui è la valutazione di un insieme di circostanze e di fattori che può e deve condurre chi giudica a stabilire che all'*hate speech* debba/possa invece essere attribuita rilevanza penale. Da questo punto di vista, le linee guida, richiamando le indicazioni fornite dall'Alto Commissario delle Nazioni Unite per i diritti umani, affermano che compito del giudice è «to determine that there was a reasonable probability that the speech would succeed in inciting actual action against the target group, recognizing that such causation should be rather direct»[33]: ad assumere rilievo sono un insieme di fattori – dal contesto nel quale quei contenuti (scritti, verbali o di altro tipo) sono diffusi, alla posizione/status ricoperta dall'autore, al grado di diffusione, all'effettiva intenzione dell'autore e, soprattutto, alla probabilità che l'incitamento si traduca in un'azione concreta[34] – ai quali, come si avrà modo di considerare a breve, il giudice ha dovuto fare riferimento ogniqualvolta si è dovuto pronunciare in tema di *hate speech* o, per meglio dire, ha dovuto individuare il punto di equilibrio tra diritto alla libertà di espressione e tutela di diritti altrui.

4. *Il problema del rapporto tra* hate speech *e libertà di espressione*

La questione di fondo, che sottende in generale il dibattito sull'*hate speech* e che, come si è potuto constatare, è stata sistematicamente richiamata nel testo degli strumenti dedicati al problema dell'*hate speech* è quella del rapporto tra *hate speech* e libertà di espressione[35]. In essi compare infatti un riferimento costante alla necessità di conciliare due obiettivi giustapposti: da un lato il contrasto all'*hate speech* e, dall'altro, la tutela di un basilare diritto quale è il diritto alla libertà di espressione, il cui esercizio va tutelato anche quando assume forme che possono risultare scioccanti, disturbanti, o offensive[36]. Il tema del contrasto all'*hate speech* ha posto dunque e continua a porre il problema della compatibilità delle misure eventualmente adottate dallo Stato per combattere tale fenomeno con gli obblighi, di diversa natura, imposti al medesimo da alcune norme di diritto internazionale: rilevano, a tale proposito, non solo le

alla violenza. Cfr. *Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred*, A/HRC/22/17/Add.4, 11 gennaio 2013.

[33] *Hate Crimes Prosecution*, cit., p. 25, nonché OHCHR, *One-pager on "incitement to hatred"* cit.

[34] Si tratta del cosiddetto "threshold test" definito dal sopracitato "Rabat Plan of Action", Doc. A/HRC/22/17/Add.4, par. 29. Con riferimento alla valutazione della probabilità che all'incitamento consegua un'azione concreta a danno di specifici individui, il Rabat Plan utilizza l'espressione "likelihood, including imminence".

[35] Per i diversi profili si rimanda a M. SPATTI, *Hate Speech e negazionismo tra restrizioni alla libertà di espressione e abuso del diritto*, in *Studi sull'integrazione europea*, 2014, p. 341 ss.; A. BROWN, *What is Hate Speech? Part 2: Family Resemblances*, in *Law and Philosophy*, 2017, p. 561 ss.; J.W. HOWARD, *Free Speech and Hate Speech*, in *Annual Review of Political Science*, 2019, p. 93 ss.; P. DE SENA, M. CASTELLANETA, *Hate Speech e messaggi discriminatori: riflessioni internazionalistiche sul caso* CasaPound c. Facebook, in *Quaderni di SIDIBlog*, 2020, p. 391 ss.; A. HAREL, *Hate Speech*, in A. STONE, F. SCHAUER, *The Oxford Handbook of Freedom of Speech*, Oxford, 2021, p. 455 ss.; E. ASWAD, D. KAYE, *Convergence & Conflict: Reflections on Global and Regional Human Rights Standards on Hate Speech*, in *Northwestern Journal of Human Rights*, 2022, p.186 ss.; M. LUGATO, *Il «discorso d'odio»: le coordinate giuridiche del ragionamento internazionalistico*, in *Rivista di diritto internazionale*, 2022, p. 959 ss.; A. CLOONEY, A. GARDOLL, *Hate Speech*, in A. CLOONEY, D. NEUBERGER (eds.), *Freedom of Speech in International Law*, Oxford, 2024, spec. p. 172 ss.

[36] In questo senso si è espressa la Corte europea dei diritti umani in un caso ben noto quanto risalente (ECtHR, *Handyside v. United Kingdom*, n. 5493/72, 7 dicembre 1976, par. 49). L'espressione è poi ripresa nel testo della Raccomandazione 20/1997 del CoE (Preamblo e principle 2), della Raccomandazione 16/2022 del CoE (preambolo, art.3 (28) e della Raccomandazione 15/2016 dell'ECRI (parr. 8,10).

norme che tutelano il diritto alla libertà di opinione ed espressione, ma altresì quelle che impongono il divieto di discriminazione, contenute in diversi strumenti vincolanti adottati, a livello universale e regionale, in materia di diritti umani. Si tratta di norme che in diverso modo prevedono la possibilità che lo Stato intervenga limitando l'esercizio della libertà di manifestazione del pensiero e delle opinioni, sia nell'ottica di combattere il fenomeno della discriminazione nelle sue varie forme, sia nell'ottica di tutelare diritti altrui o superiori esigenze di carattere pubblico.

Sotto il primo profilo vengono in esame gli articoli 19 e 20 del Patto delle Nazioni Unite sui diritti civili e politici del 1966 (d'ora in poi il Patto), nonché le corrispondenti norme presenti negli strumenti regionali in materia di tutela dei diritti umani, quali la CEDU (articolo 10), la Convezione americana dei diritti dell'uomo (1969 – articolo 13), la Carta dei diritti fondamentali dell'Unione europea (2007 – articolo 11). Sotto il secondo profilo, vengono invece in esame le norme contenute nella già citata ICERD ed in particolare l'articolo 4, cui si è fatto riferimento nel paragrafo precedente.

Prendendo le mosse innanzitutto dalle norme contenute nel Patto, merita ricordare che il dibattito che aveva condotto all'elaborazione del testo dell'articolo 19 era stato lungo ed articolato ed era apparso incentrato da un lato sulla natura di un diritto fondamentale – quello alla libertà di espressione[37] – posto a cardine di qualsiasi società democratica e concepito come libero da interferenze governative, dall'altro sulla necessità che occorresse individuare dei limiti alle manifestazioni del pensiero e delle opinioni nell'ottica di un bilanciamento tra diritti del singolo, tutela dell'interesse pubblico e tutela di altri diritti egualmente garantiti[38]. Quello stesso dibattito aveva così condotto gli estensori della norma a concepirne il dettato in modo da precisare che l'esercizio di quel diritto comporta anche precise responsabilità, innanzitutto quelle relative al rispetto dei diritti altrui; in quest'ottica è stata contemplata la possibilità che possano essere apposte restrizioni alla libertà di espressione, giustificabili alla luce della necessità di garantire il rispetto dei diritti o della reputazione altrui, nonché la salvaguardia della sicurezza nazionale, dell'ordine pubblico, della sanità e della morale pubblica[39]. A corredo della norma in esame, e ad ulteriore precisazione del portato e dei limiti che contraddistinguono l'esercizio di tale fondamentale diritto, si pone l'altra norma, di cui all'articolo 20, che oltre a stabilire un divieto di propaganda di guerra, impone in capo agli Stati l'obbligo di adottare misure legislative dirette a sanzionare qualsiasi appello all'odio nazionale, razziale o religioso che costituisca incitamento alla discriminazione, ostilità o violenza.

Sul piano regionale un percorso similare aveva condotto all'affermazione del diritto alla libertà di espressione così come formulato nell'articolo 10 della CEDU e nell'articolo 13 della Convenzione americana sui diritti umani[40]; quest'ultima, poi, a differenza della CEDU e

---

[37] La norma tutela il diritto alla libertà di opinione ed espressione. È poi in particolare il par. 2 dell'art. 19 a stabilire che «Ogni individuo ha il diritto alla libertà di espressione; tale diritto comprende la libertà di cercare, ricevere e diffondere informazioni e idee di ogni genere, senza riguardo a frontiere, oralmente, per iscritto, attraverso la stampa, in forma artistica o attraverso qualsiasi altro mezzo di sua scelta».

[38] Cfr. *Draft International Covenants on Human Rights. Report of the Third Committee*, UN Doc. A/5000, 5 dicembre 1961.

[39] Patto sui diritti civili e politici, art. 19, par. 3.

[40] Entrambe le norme, infatti, prevedono che il diritto alla libertà di espressione possa essere sottoposto a condizioni e restrizioni «necessarie in una società democratica alla sicurezza nazionale, all'integrità territoriale o alla pubblica sicurezza, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, alla protezione della reputazione o dei diritti altrui (…)» (CEDU, art. 10 par. 2), ovvero per tutelare il

analogamente a quanto previsto dall'art. 20 del Patto, prevede al paragrafo 5 dell'articolo 13 il divieto di propaganda di guerra e il divieto di incitamento all'odio e alla discriminazione[41]. Merita inoltre fare cenno all'articolo 11 della Carta dei diritti fondamentali dell'Unione europea la cui formulazione riprende quella dell'articolo 10 della CEDU, senza tuttavia che compaia il riferimento alle eventuali limitazioni che possono essere apposte all'esercizio del diritto, che la norma della CEDU invece prevede[42]. Va però precisato che anche laddove, come nel caso dell'articolo 10 della CEDU e dell'articolo 11 della Carta dei diritti fondamentali dell'Unione europea, il divieto di propaganda o di istigazione all'odio e alla discriminazione non sia espressamente enunciato, il fatto che la norma preveda che possano essere apposte restrizioni all'esercizio del diritto per tutelare altri diritti protetti dal medesimo strumento, ovvero superiori interessi della collettività (ordine pubblico, sicurezza dello Stato), può comunque consentire allo Stato di intervenire, laddove lo ritenga necessario, al fine di vietare la propaganda o l'incitamento all'odio e alla discriminazione. Nel sistema delineato dalla CEDU è poi l'articolo 17 che può svolgere un ulteriore ruolo per contrastare eventuali fenomeni di "abuso del diritto": tale norma, infatti, consente alla Corte europea dei diritti dell'uomo di rigettare le istanze di chi lamenta la violazione di diritti protetti dalla Convenzione nel caso in cui proprio i comportamenti posti in essere dal ricorrente integrino la violazione o siano contrari ad altri diritti egualmente garantiti dalla medesima Convenzione[43].

In un'ottica invece di lotta alla discriminazione e dunque di tutela di diritti basilari, ma al contempo di un superiore interesse collettivo è concepita anche la formulazione del menzionato articolo 4 della ICERD; quest'ultimo peraltro non riconosce una facoltà allo Stato, ma piuttosto prevede – come già rilevato – l'obbligo di adottare misure legislative volte a qualificare come reati quelle manifestazioni del pensiero che servano a diffondere idee basate sulla superiorità razziale, nonché ogni atto di violenza, o incitamento a tali atti, diretti contro ogni razza o gruppo di individui di colore diverso o di diversa origine etnica[44].

Le norme citate rappresentano dunque il parametro fondamentale per quanto concerne l'ambito di esercizio del diritto alla libertà di espressione: tuttavia, hanno, nel tempo, costituito l'oggetto di un dibattito costante. Da un lato, infatti, si è posta la questione relativa alla natura non assoluta del diritto alla libertà di espressione, che si evince dalle possibili restrizioni che lo Stato è autorizzato ad apporre, richiamate nel testo delle norme citate; dall'altro quella della portata degli obblighi imposti allo Stato sia dall'articolo 20 del

---

rispetto dei diritti e della reputazione di altri, nonché la protezione della sicurezza nazionale, dell'ordine pubblico o della salute o della morale pubblica (Convenzione americana dei diritti umani, art. 13 par. 2(a,b).

[41] Più precisamente si legge che «Qualunque propaganda in favore della guerra e qualunque richiamo all'odio nazionale, razziale o religioso che costituisca incitamento alla violenza illegale o ad ogni altra azione simile contro qualunque persona o gruppo di persone per qualsiasi ragione, compresi motivi di razza, colore, religione, lingua o origine nazionale o sociale, deve essere considerato dalla legge come reato».

[42] Va tuttavia considerato che il mancato riferimento ai limiti che possono essere apposti all'esercizio del diritto va interpretato alla luce del disposto dell'art. 52 della stessa Carta: quest'ultimo non solo prevede che eventuali limitazioni all'esercizio dei diritti garantiti dalla Carta devono essere previste per legge e devono rispondere a finalità di interesse generale o di tutela di diritti altrui, ma prevede altresì che laddove le norme della Carta tutelino diritti corrispondenti a quelli garantiti dalla CEDU, «il significato e la portata degli stessi sono uguali a quelli conferiti dalla suddetta convenzione».

[43] La norma è pensata nell'ottica di preservare quell'assetto democratico su cui tutto il sistema creato dalla CEDU si basa. Sul punto cfr. ECtHR, *Refah Partisi et al. v. Turkey*, nn. 41340/98, 41342/98,41343/98, 41344/98, 13 febbraio 2003, par. 99; *Ždanoka v. Latvia*, n. 58278/00, 16 marzo 2006, par. 100; *Petropavlovskis v. Latvia*, n. 44230/06, 1 giugno 2015, parr. 71-72.

[44] Cfr. *supra*, nota n. 18.

Patto, sia dalle norme corrispondenti contenute negli strumenti regionali, in particolare il citato articolo 13 (par. 5) della Convenzione americana. Da questo punto di vista, già l'apposizione di riserve all'articolo 20 del Patto, considerato dai rappresentanti di alcuni Stati in contrasto con il diritto alla libertà di espressione, nonché di dichiarazioni interpretative, volte a precisare che l'applicazione della norma non avrebbe comportato modifiche legislative all'interno degli ordinamenti nazionali[45], testimonia del fatto che fosse già allora ben viva la preoccupazione che di quelle facoltà/obblighi di cui agli articoli 19 e 20, gli Stati avrebbero potuto fare un uso distorto e strumentale rispetto all'obiettivo di circoscrivere e comprimere un diritto considerato fondamentale per il funzionamento di qualsiasi ordinamento democratico.

Lungi dall'essere stato superato, il dibattito intorno al ruolo dello Stato e al margine di discrezionalità di cui esso gode nel definire in che termini e in quale misura eventuali restrizioni alla libertà di espressione possano essere apposte per assicurare un adeguato bilanciamento sia tra diritti egualmente garantiti, sia tra l'esercizio di un diritto fondamentale e le esigenze di carattere collettivo, appare tutt'altro che sopito. Per quanto, infatti, in linea teorica il diritto alla libertà di espressione venga solitamente qualificato come un diritto fondamentale, posto a cardine di qualsiasi società democratica, in realtà ingerenze e restrizioni sono frequentemente apposte al suo esercizio anche nell'ambito di ordinamenti liberali e democratici[46].

Nell'ambito di tale dibattito si inserisce, evidentemente, la questione relativa all'*hate speech*: quest'ultimo è infatti fenomeno che va letto e interpretato alla luce di quelle norme che, pur tutelando la libertà di espressione, consentono che la stessa possa essere, in presenza di certe circostanze, limitata e compressa. Proprio quelle norme hanno rappresentato il necessario parametro di riferimento non solo per gli estensori degli strumenti, sopra richiamati, di contrasto all'*hate speech*, ma altresì per gli organi di controllo chiamati a pronunciarsi in merito alle asserite violazioni, da parte dello Stato, del diritto alla libertà di espressione. Se infatti combattere l'*hate speech* significa garantire il necessario bilanciamento di diritti egualmente riconosciuti, nonché contrastare tutte le forme di incitamento all'odio e alla discriminazione su base nazionale, razziale, etnica religiosa o di altro tipo è però indubbio che in assenza di definizioni chiare ed inequivoche circa i contenuti della fattispecie *hate speech*, è rimesso all'organo di controllo il compito di valutare in che misura certe manifestazioni del pensiero che hanno contenuti offensivi, diffamatori, denigratori, così come quelle definibili razziste, xenofobe, negazioniste e/o revisioniste possono essere considerate compatibili con l'esercizio del diritto alla libertà di espressione ovvero devono essere ritenute lesive di diritti altrui o di interessi di natura collettiva[47]. Altrettanto deve dirsi per quanto concerne la

[45] Si vedano le riserve e le dichiarazioni apposte da vari paesi occidentali (fra gli altri, Australia, Belgio, Danimarca, Finlandia, Lussemburgo, Malta, Nuova Zelanda, Regno Unito, Stati Uniti). Cfr. https://ccprcentre.org/files/media/List_of_ICCPR_reservations.pdf.

[46] Cfr. *Article 19*, *Global Expression Report 2024*, https://www.globalexpressionreport.org/.

[47] Si veda per es. quanto affermato dal CERD nel caso *The Jewish community of Oslo et al. v. Norway*, n. 30/2003, 22 agosto 2005, CERD/C/67/D/30/2003, par. 10.5, in cui il Comitato ribadendo le proprie posizioni di cui alla *General recommendation* 15, afferma che «the prohibition of all ideas based upon racial superiority or hatred is compatible with the right to freedom of opinion and expression»; ne derivava dunque che le affermazioni fatte da un appartenente ad un gruppo neonazista nel corso di una manifestazione organizzata per commemorare Rudolf Hesse, in cui si era glorificato l'operato di Hitler e si erano diffusi proclami antisemiti, dovevano considerarsi «as incitement at least to racial discrimination, if not to violence», per cui l'assoluzione da parte dell'Alta Corte norvegese della persona imputata era considerarsi alla stregua di una violazione dell'art. 4 della ICERD. Si veda altresì quanto affermato dal CCPR nel noto caso *Faurisson v. France*, n. 550/1993, 8 novembre 1996, CCPR/C/58/D/550/1993.

valutazione relativa al rispetto, da parte dello Stato che abbia apposto eventuali restrizioni all'esercizio della libertà di espressione, dei parametri di necessità e di proporzionalità tra le misure – di natura penale o civile – adottate e gli obiettivi da perseguire in ciascun singolo caso[48]. In questo senso, il giudice e più in generale l'organo di controllo internazionale hanno dovuto, caso per caso, tracciare il *discrimen* tra i comportamenti qualificabili come semplicemente idonei a sollecitare sentimenti che possono essere ostili o finanche odiosi ed i comportamenti che devono essere stigmatizzati perché suscettibili di istigare alla violenza e all'odio e potenzialmente idonei a tradursi in azioni[49]. Se da un lato l'orientamento affermatosi è quello per cui l'incitamento all'intolleranza, alla violenza e all'odio devono essere considerati come «l'une des limites à ne dépasser en aucun cas dans le cadre de l'exercice de la liberté d'expression»[50], dall'altro appare evidente che la posizione assunta dalla giurisprudenza riflette bene quell'idea, affermata anche nel testo degli strumenti internazionali adottati in materia di *hate speech*, secondo la quale l'adozione di sanzioni penali deve essere riservata ai casi più gravi ed applicata sulla base di norme formulate in maniera chiara ed inequivoca[51]. È

[48] Come ha precisato il CCPR, *General Comment 34, on Article 19: Freedoms of opinion and expression*, del 29 luglio 2011, CCPR/C/GC/34, par. 21 «(…) when a State party imposes restrictions on the exercise of freedom of expression, these may not put in jeopardy the right itself. The Committee recalls that the relation between right and restriction and between norm and exception must not be reversed (…) ». La prassi del Comitato offre conferma di tale impostazione: lo stesso ha avuto modo di affermare che «freedom of opinion and freedom of expression constitute the foundation of every free and democratic society. Any restrictions on the exercise of those freedoms must conform to the strict tests of necessity and proportionality and must be applied only for those purposes for which they were prescribed and must be directly related to the specific need on which they are predicated» (*Pazel Kovlov v. Belarus*, n. 1949/2010, 7 maggio 2015, CCPR/C/113/D/1949/2010, par. 7.6). Cfr. altresì *Seok-ki Lee, Hong-yeol Kim et al. V. Republic of Korea*, n. 2809/2016, 25 marzo 2021, CCPR/C/130/D/2809/2016, parr. 7.2, 7.3, 7.9; *Dina Baydildayeva v. Kazakhstan*, n. 2545/2015, 22 maggio 2023, CCPR/C/137/D/2545/2015, parr. 8.3, 8.4; *Gennady Fedynich v. Belarus*, n. 2913/2016, 6 dicembre 2022, CCPR//C/136/D/2913/2016, parr. 7.8, 7.9; *Shin v. Republic of Korea*, n. 926/2000, 25 aprile 2000, CCPR/C/80/D/926/2000, parr. 7.2, 7.3. Nella stessa direzione si è mossa la Corte europea dei diritti dell'uomo: quest'ultima ha infatti avuto modo di affermare che l'adozione di misure che implichino la compressione della libertà di opinione va motivata alla luce di "pressing social needs" la cui effettiva sussistenza è valutata dallo Stato ma che è comunque, in ultima istanza, oggetto del sindacato della Corte (*Handyside v. United Kingdom*, cit., par. 50; cfr altresì *Sürek v. Turkey*, n. 1, 26682/95, 9 luglio 1999, par. 58; *Nilsen et Johnsen v. Norway*, n. 23118/93, 25 novembre 1999, par. 4; *Erbakan v. Turkey*, n. 59405/00, 6 ottobre 2006, par. 55; *Perinçek v. Switzerland* n. 27510/08, 15 ottobre 2015; *Magyar Helsinki Bizottság v. Hungary*, n. 18030/2011, 8 novembre 2016, par. 187. Per quanto più specificamente concerne gli aspetti legati alla necessità e proporzionalità delle misure limitative della libertà di espressione rispetto alla necessità di tutelare la sicurezza pubblica e/o i diritti altrui cfr. *Barthold v. Germany*; n. 8734/79, 25 marzo 1985, parr. 58,59; *Zana v. Turkey*, n. 69/1996/688/880), 25 novembre 1997, parr. 59-62; *Magyar Helsinki Bizottság v. Hungary*, cit., parr. 196-200; *Zemmour v. France*, n. 63539/19, 20 dicembre 2022, par. 65.

[49] Cfr. CCPR, *Faurisson v. France*, cit., laddove il Comitato afferma che le restrizioni apposte al diritto alla libertà di espressione «served the respect of the Jewish community to live free from fear of an atmosphere of anti-semitism» (par. 9.6).

[50] ECtHR, *Zemmour v. France*, cit., par. 50.

[51] Come rilevato dalla Corte europea dei diritti dell'uomo, «criminal conviction is a serious sanction (….) although sentencing is in principle a matter for the national courts, the imposition of a prison sentence for an offence in the area of a debate on an issue of legitimate public interest will be compatible with freedom of expression as guaranteed by Article 10 of the Convention only in exceptional circumstances, notably where other fundamental rights have been seriously impaired (…)». Cfr. *Sürek v. Turkey*, cit., par. 68; *Savva Terentyev v. Russia*, n. 10692/09, 4 febbraio 2019, parr. 83, 85; *Dmitriyevskiy v. Russia,* n. 42168/06, 29 gennaio 2018, parr. 82-83*; Otegi Mondragon v. Spain,* n. 2034/07, 15 marzo 2011, parr. 59-60. Nello stesso senso cfr. CCPR, *Eglė Kusaitė v. Lithuania*, n. 2716/2016, 24 settembre 2019, CCPR/C/126/D/2716/2016, par. 8.10; con riferimento poi ai casi di diffamazione il Comitato ha affermato che «States parties should consider the decriminalization of defamation, and, in any case, the application of the criminal law should only be countenanced in the most

stato dunque, ed è tutt'ora compito del giudice internazionale (o dell'organo di monitoraggio) valutare, di volta in volta, tenendo conto di una serie di elementi, la "gravità" del caso, che è quanto dire la portata e l'impatto delle manifestazioni e delle espressioni sul singolo destinatario ovvero sulla comunità e sul contesto sociale, al fine di verificare poi se le misure adottate dallo Stato, che comportano una compressione della libertà di espressione, siano davvero giustificabili alla luce della necessità di tutelare diritti fondamentali o esigenze di carattere pubblico[52]. Tale valutazione serve ad accertare la presenza dei presupposti richiesti non solo perché possa essere giustificata l'adozione di sanzioni penali, ma altresì perché possa essere dimostrato che qualora lo Stato abbia posto limiti all'esercizio della libertà di espressione, lo abbia fatto nel rispetto dei fondamentali parametri di necessità in una società democratica[53] e di proporzionalità tra misure adottate ed obiettivo da perseguire: esigenza primaria è infatti, e resta, quella di evitare che la discrezionalità di cui lo Stato gode non sia così ampia da lasciare spazio ad abusi e ad un'applicazione selettiva delle norme.

5. *Il contrasto all'* hate speech *online e il ruolo degli Internet* provider: *dai codici di condotta al recente regolamento (UE) 2022/2065*

Le considerazioni fatte finora in merito al ruolo svolto dalle norme che tutelano la libertà di espressione e vietano l'incitamento all'odio e alla discriminazione, quali parametri di riferimento per l'individuazione delle regole applicabili nel contrasto all'*hate speech*, si ritiene

serious of cases and imprisonment is never an appropriate penalty. If defamation should never result in a penalty of deprivation of liberty being imposed on the grounds that it is not an appropriate penalty, then a fortiori no detention based on charges of defamation may ever be considered either necessary or proportionate» (*Lydia Cacho Ribeiro v. Mexico*, n. 2767/2016, 29 agosto 2018, CCPR/C/123//D/2767/2016, par. 8.10). Nello stesso senso *Rafael Marques de Morais v. Angola*, n. 1128/2002, 18 aprile 2005, CCPR/C/83/D/1128/2002, par. 6.1.

[52] Cfr. ad es. i casi in cui la Corte europea dei diritti umani si è pronunciata in materia di diffamazione a mezzo stampa pervenendo a conclusioni diverse tenuto conto delle specifiche circostanze del caso (cfr. ECtHR, *Cumpănă e Mazăre v. Romania*, ricorso n. 33348/96, 17 dicembre 2004, par. 115; *Ruokanen and Others v. Finland*, n. 45130/06, 6 aprile 2010, parr. 51-52). In altri casi la Corte ha ritenuto proporzionate le misure adottate nei confronti di alcuni giornalisti autori di articoli che contenevano espressioni lesive della dignità delle popolazioni di etnia non russa in alcune regioni della Russia (*Atamanchuk v. Russia*, n. 4493/11, 11 febbraio 2020, par.72), perché ha ritenuto che quelle affermazioni fossero idonee a provocare o indurre l'odio e l'intolleranza su base religiosa, etnica o razziale (nello stesso senso cfr. *Norwood v. United Kingdom*, n. 23131/03, 16 novembre 2004; *E. S. v. Austria*, n. 38450/12, 25 ottobre 2018, par. 57; *Jersild v. Denmark*, n.15890/89, 23 settembre 1994, par. 35). Altrettanto può dirsi per quanto concerne l'orientamento della Corte rispetto alle affermazioni volte a negare fatti storici chiaramente accertati, come l'olocausto, o a diffondere idee e stereotipi antisemiti (*B.H., M.W, H.P. and G K. V. Austria*, n. 12774/87, 12 dicembre 1989, *passim*; *Künen v. Germany*, n. 12194/86, 12 maggio 1988, passim; *Garaudy v. France*, n. 65831/01, 24 giugno 2003, *passim*). Nello stesso cfr. CCPR, *J. R. T. and the W. G. Party v. Canada*, n. 104/1981, 6 aprile 1983, CCPR/C/OP/2, par. 8(b); *Faurisson v. France*, cit., parr. 9.5, 9.6; *Ross v. Canada*, n. 736/1997, 26 ottobre 2000, CCPR/C/70/D/736/1997, parr. 11.1, 11.5, 11.6.

[53] Come ha affermato la Corte europea dei diritti umani (*Zemmour v. France*, cit., par. 51) «(…) la tolérance et le respect de l'égale dignité de tous les êtres humains constituent le fondement d'une société démocratique et pluraliste. Il en résulte qu'en principe on peut juger nécessaire, dans les sociétés démocratiques, de sanctionner voire de prévenir toutes les formes d'expression qui propagent, incitent à, promeuvent ou justifient la haine fondée sur l'intolérance (y compris l'intolérance religieuse), si l'on veille à ce que les "formalités", "conditions", "restrictions" ou "sanctions" imposées soient proportionnées au but légitime poursuivi. Il reste loisible aux autorités compétentes d'adopter, en leur qualité de garantes de l'ordre public institutionnel, des mesures, même pénales, destinées à réagir de manière adéquate et non excessive à de pareils propos».

siano riferibili anche al cosiddetto *hate speech online*: come precisato in apertura, si parte infatti dal presupposto che non ci si trova dinanzi a due fenomeni distinti se non per quanto attiene alle modalità, *rectius* agli strumenti, con cui vengono veicolate idee ed opinioni e si diffondono contenuti che possono essere qualificati come *hateful*. È proprio tenuto conto di tali modalità e strumenti, che è quanto dire delle peculiarità che caratterizzano il sistema della comunicazione digitale e del ruolo svolto da Internet in tale contesto[54], che si ritiene di dover dedicare le ultime riflessioni al tema, più specifico, dell'*hate speech online*. Rispetto a tale fenomeno si è reso infatti necessario definire un insieme di regole che tenesse conto delle specificità che contraddistinguono il funzionamento del sistema: soprattutto nella sua seconda fase di sviluppo denominata *web 2.0*, Internet si caratterizza infatti non solo per il fatto che chi usufruisce dei servizi può al contempo essere l'autore dei contenuti immessi in rete, ma anche e soprattutto per il ruolo svolto, nell'ambiente digitale, dai *provider* che, come precisato in apertura, si pongono come intermediari tra chi crea e diffonde contenuti e chi ne fruisce. Si è dunque posta la necessità di individuare dei meccanismi di contrasto attivabili a monte, vale a dire prima che certi contenuti, classificabili come *hate speech* si propaghino in rete; in quest'ottica è apparso inevitabile dover disciplinare il ruolo dei *provider*, questione che fin dapprincipio si è posta in termini diversi da quelli che caratterizzano la regolamentazione della tradizionale attività editoriale[55]. Tenuto conto delle peculiarità del sistema Internet, l'operazione si è rivelata tutt'altro che semplice e solleva alcune fondamentali problematiche che restano sullo sfondo nonostante siano state individuate possibili soluzioni: si intende fare riferimento alla dimensione transnazionale dell'attività svolta dai *provider*, nonché al fatto che quella basilare attività di controllo sui contenuti che devono essere qualificati come *hate speech* viene non solo demandata a soggetti privati[56], ma soprattutto affidata al funzionamento dello

---

[54] ECtHR, *Times newspaper L.t.d. v. The United Kingdom*, n. 3200/03 e 23676/03, 10 giugno 2009, par. 27 in cui si afferma che «In the light of its accessibility and its capacity to store and communicate vast amounts of information, the Internet plays an important role in enhancing the public's access to news and facilitating the dissemination of information in general». Tale assunto è ribadito in *Ahmet Yldirim v. Turkey*, n. 3111/10, 18 marzo 2013, par. 48; *Cengiz and Others v. Turkey*, n. 48226/10 e 14027/11, 1 marzo 2016, par. 52. In *Delfi AS v. Estonia*, n. 64569/09, 16 giugno 2015, la Corte, ribadendo la propria posizione e dunque riconoscendo il fondamentale ruolo di internet nel fornire «an unprecedented platform for the exercise of freedom of expression» (par. 110), evidenzia che «however, alongside these benefits, certain dangers may also arise. Defamatory and other types of clearly unlawful speech, including hate speech and speech inciting violence, can be disseminated like never before, worldwide, in a matter of seconds, and sometimes remain persistently available online» (*ibid.*); nello stesso senso cfr. *Annen v. Germany*, n. 3690/10, 26 febbraio 2016, par. 67; *Savva Terentyev v. Russia*, cit., par. 79.

[55] Come ha evidenziato la Corte europea dei diritti dell'uomo (*Editorial Board of Pravoye Delo and Shtekel v. Ukraine*, n. 33014/05, 5 agosto 2011, par. 63) «It is true that the Internet is an information and communication tool particularly distinct from the printed media, especially as regards the capacity to store and transmit information. The electronic network, serving billions of users worldwide, is not and potentially will never be subject to the same regulations and control. The risk of harm posed by content and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press. Therefore, the policies governing reproduction of material from the printed media and the Internet may differ. The latter undeniably have to be adjusted according to the technology's specific features in order to secure the protection and promotion of the rights and freedoms concerned». Cfr. altresì *Annen v. Germany*, cit., par. 72.

[56] Cfr. J.M BALKIN, *Free Speech is a Triangle*, in *Columbia Law Review*, 2018, p. 2011 ss.; G. RUOTOLO, *A Little Hate Worldwide! Di libertà d'opinione e discorsi politici d'odio* on-line *nel diritto internazionale ed europeo*, in *Diritti umani e diritto internazionale*, 2020, p. 549 ss.; G. ZICCARDI, *Le espressioni d'odio sulle piattaforme digitali: alcune considerazioni informatico-giuridiche*, in M. D'AMICO, M. BRAMBILLA, V. CRESTANI, N. FIANO (a cura di), *Il linguaggio dell'odio. Fra memoria e attualità*, Milano, 2021, p. 159 ss.; M. HUSOVEC, *(Ir)Responsible Legislature? Speech Risks under the EU's Rules on Delegated Digital Enforcement*, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3784149; G.

strumento algoritmico e dunque ad un meccanismo di "moderazione automatizzata" dei contenuti immessi in rete[57].

Guardando dunque alla tipologia degli strumenti di contrasto all'*hate speech online* che sono stati delineati, si evince come questi abbiano inizialmente fatto perno sulla definizione di un insieme di principi utili ad indirizzare i *provider* verso una corretta "gestione" delle proprie piattaforme: è quanto dire del controllo che il *provider* deve esercitare sull'utilizzo che sia fatto delle proprie piattaforme, al fine di garantire che attraverso le medesime non siano trasmessi, veicolati e propagati contenuti che possono avere un contenuto qualificabile come *hateful*. Un primo tentativo di regolamentazione di questo tipo è individuabile nell'avvio, nel 2008, della cosiddetta *Global Network Initiative* (GNI): si tratta di un forum costituito da un gruppo ampio e vario di *stakeholder* che, nell'ambito di un'attività a supporto del diritto alla libertà di espressione, ha adottato un documento in cui sono enunciati i principi diretti a garantire e dare impulso alla libertà di pensiero e di espressione nel settore delle tecnologie e dell'informazione a livello globale[58]. Nel testo del documento compare il significativo richiamo alle norme internazionali sui diritti umani che, nell'ottica dei redattori, assurgono a parametro di riferimento rispetto al quale i *provider* devono uniformare la propria attività. Da questo punto di vista, ai *provider* è richiesto di svolgere un duplice ruolo. Innanzitutto, un ruolo di controllo, finalizzato ad evitare che le proprie piattaforme si prestino ad essere il luogo in cui alcuni fondamentali diritti possono essere violati e a garantire la rimozione dei contenuti che risultano essere in contrasto con le norme e gli standard internazionali ovvero l'adozione di misure alternative[59]; in secondo luogo, un ruolo di controllo e di vigilanza sull'attività dello Stato, al fine di minimizzare o, meglio, contrastare «the adverse impact of governments demands, laws or regulations» sull'effettivo godimento di quegli stessi diritti[60].

VASINO, *Censura "private" e contrasto all'* hate speech *nell'era delle* Internet Platforms, in *Federalismi.it*, 2023, p. 130 ss.; G. E. VIGEVANI, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritto costituzionale*, 1/2023, pp. 41-54.

[57] In merito al problema dell'utilizzo (e ai rischi connessi) dell'algoritmo nell'individuazione di *harmful content*, cfr. M. FINCK, *Artificial Intelligence and Online Hate Speech. Issue Paper*, January 2019, https://cerre.eu/wp-content/uploads/2020/05/CERRE_Hate-Speech-and-AI_IssuePaper.pdf.; R. GORWA *et al.*, *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*, in *Big Data and Society*, 2020, p. 1 ss.; E. LLANSÓ, J. VAN HOBOKEN, J. HARAMBAM, *Artificial intelligence, content moderation, and freedom of expression*, 2020, https://pure.uva.nl/ws/files/190771414/AI-Llanso-Van-Hoboken-Feb-2020.pdf; European Union Agency for Fundamental Rights, *Bias in Algorithms. Artificial Intelligence and Discrimination*, Vienna, 2022; P. DUNN, *Moderazione automatizzata e discriminazione algoritmica: il caso dell'*hate speech, in *Rivista italiana di informatica e diritto*, 2022, p. 133 ss.; B. FARRAND, *"Is This a Hate Speech?" The Difficulty in Combating Radicalisation in Coded Communications on Social Media Platforms*, in *European Journal on Criminal Policy and Research*, 2023, p. 477 ss.

[58] Cfr. *GNI Principles on Freedom of Expression and Privacy*, https://globalnetworkinitiative.org/wp-content/uploads/2018/04/GNI-Principles-on-Freedom-of-Expression-and-Privacy.pdf.

[59] Si legge che «ICT companies should comply with all applicable laws and respect internationally recognized human rights, wherever they operate»; le norme e i principi cui si fa riferimento sono quelli contenuti nella Dichiarazione universale dei diritti dell'uomo del 1948, nei Patti delle Nazioni Unite sui diritti civili e politici e sui diritti economici, sociali e culturali, nonché nei principi-guida delle Nazioni Unite su imprese e diritti umani. Cfr. altresì GNI, *Implementation Guidelines for the Principles on Freedom of Expression and Privacy*, https://globalnetworkinitiative.org/wp-content/uploads/2018/08/Implementation-Guidelines-for-the-GNI-Principles.pdf .

[60] È infatti posto in evidenza che il ricorso da parte dell'autorità statale a misure limitative della libertà di espressione deve essere motivato da circostanze eccezionali e le misure adottate devono essere proporzionate allo scopo che si persegue Si veda più di recente l'adozione, nel 2020, del *Content Regulation & Human Rights Policy Brief* che, anche tramite l'analisi delle iniziative legislative adottate in vari Stati, evidenzia le criticità connesse con il ricorso a strumenti che nell'ottica di contrastare l'*hate speech* e l'*hate speech online*, possono tuttavia

Natura simile ai principi guida stabiliti nel contesto della GNI ha anche il Codice di condotta per la prevenzione e il contrasto alle forme illegali di incitamento *online* adottato nel 2016 dalla Commissione europea di concerto con alcuni dei più importanti *provider*[61]. Il codice di condotta prevede che questi ultimi si impegnino ad adottare procedure idonee a segnalare ed eventualmente rimuovere messaggi e contenuti illegali, qualificabili come incitamento all'odio e alla violenza, che siano veicolati *online*. La natura illegale dei medesimi è peraltro valutata alla luce di quanto previsto dalla decisione quadro 2008/913/GAI: quest'ultima costituisce dunque il necessario parametro di riferimento per le squadre specializzate che i *provider* devono provvedere a costituire, al fine di garantire la necessaria sorveglianza e la conseguente tempestiva rimozione dei contenuti illegali. Ulteriore impegno richiesto è quello relativo alla predisposizione di meccanismi di informazione e sensibilizzazione degli utenti sulle tipologie di contenuti vietati.

L'adozione del codice di condotta sembra aver favorito il progressivo allineamento della *policy* dei più importanti *provider* agli standard e ai principi indicati[62]. Tale allineamento è stato in qualche modo indotto anche dal fatto che il problema del ruolo che gli stessi svolgono nel controllo dei contenuti qualificabili come *hate speech* è diventato sovente oggetto di un contenzioso sul quale il giudice internazionale si è in più circostanze espresso[63]. Il

---

comportare effettive e non del tutto prevedibili limitazioni al diritto alla libertà di espressione. https://globalnetworkinitiative.org/resources/content-regulation-human-rights/.
Sul problema del rispetto da parte dei *provider* delle fondamentali norme sui diritti umani nell'ambito delle attività dagli stessi svolte e dei servizi offerti cfr. E.M. ASWAD, *To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?* in *Washington and Lee Law Review*, 2020, p. 609 ss.

[61] Cfr. *Code of Conduct on Countering Illegal Hate Speech Online*, file:///C:/Users/utente/Downloads/code_of_conduct_on_countering_illegal_hate_speech_online_en_C08AC7D9-984D-679D-CAEF129AD536E128_42985%20(3).pdf.

[62] È stato calcolato che nell'ultimo quadrimestre del 2024, facebook ha provveduto a rimuovere 6,4 milioni di post: si registra una flessione, dunque, sia rispetto al quadrimestre precedente (7, 2 milioni), sia rispetto al quadrimestre aprile-giugno 2021, periodo durante il quale erano stati rimossi 21 milioni di post (https://www.statista.com/statistics/1013804/facebook-hate-speech-content-deletion-quarter/). Per quanto invece concerne X, nel primo quadrimestre 2024 sono stati rimossi circa 4,5 milioni di post a causa del "hateful content" (la percentuale più alta, seguita dai post con contenuti violenti o definiti come "abuse & harassement". Cfr. https://www.socialmediatoday.com/news/data-shows-x-suspending-far-fewer-users-hate-speech/728136/): nel terzo quadrimestre del 2023 Instagram ha provveduto a rimuovere circa 7 milioni di post, cifra inferiore rispetto a quella relativa al quadrimestre precedente dello stesso anno (9,8 milioni. Cfr. https://www.statista.com/statistics/1275933/global-actioned-hate-speech-content-instagram/). Il monitoraggio effettuato nel 2022 circa l'applicazione del codice di condotta testimonia comunque di una flessione delle percentuali relative alla rimozione dei contenuti *hateful* da parte delle VLOP (fatta eccezione per youtube) rispetto al precedente periodo del 2020 (63% di contenuti rimossi rispetto alle segnalazioni ricevute contro il precedente 71%). La decisione di rimuovere i contenuti dipende dalla gravità degli stessi: in media, il 69.6% dei contenuti che incitano all'omicidio o alla violenza nei confronti di specifici gruppi è stato rimosso, mentre i messaggi che utilizzano parole o immagini diffamatorie sono stati rimossi nel 59.1% dei casi. Cfr. European Commission, *7th Evaluation of the Code of Conduct*, novembre 2022, /commission.europa.eu/document/download/5dcc2a40-785d-43f0-b806-f065386395de_en?filename=Factsheet%20-%.

[63] Cfr. Corte europea dei diritti umani, *Delfi AS v. Estonia*, cit., (parr. 52, 146) in cui la Corte si allinea, richiamandola, alla giurisprudenza della Corte di giustizia dell'Unione europea e in particolare alla sentenza *Google France SARL and Google Inc.* (joined cases C-236/08), nella parte in cui la Corte per stabilire l'eventuale responsabilità del *provider* quanto ai contenuti diffusi tramite la propria piattaforma, aveva attribuito un'importanza centrale al concetto di neutralità del ruolo svolto dal *provider*. Era stato precisato che per ruolo neutrale dovesse intendersi un ruolo di tipo meramente tecnico con la conseguenza che la responsabilità del *provider* poteva essere esclusa – ai sensi dell'art. 14 della direttiva sul commercio elettronico (98/48/EC) – solo nel caso in cui il *provider* si fosse limitato a svolgere un ruolo di tal tipo e non avesse esercitato alcun tipo di

ricorso, tuttavia, a strumenti il cui funzionamento è imperniato su logiche volontaristiche non si è rivelato idoneo a soddisfare l'esigenza sempre più pressante di contenimento dei rischi legati all'utilizzo della rete: si è così ritenuto necessario operare per rafforzare il quadro normativo ed in questo senso si ritiene debba essere interpretata l'adozione, nel 2022 da parte dell'Unione europea, del regolamento 2022/2065, meglio noto come *Digital Services Act* (DSA)[64]. Seppure concepito in un quadro più ampio, in quanto destinato a riformare il sistema già definito dalla direttiva 2000/31/CE sul commercio elettronico, il regolamento contiene norme destinate a disciplinare l'attività dei *provider* (definiti "prestatori di servizi intermediari"), anche se basati fuori dal territorio dell'Unione europea, il cui comportamento responsabile «è essenziale per un ambiente online sicuro, prevedibile e affidabile»[65]. Passaggio importante, contenuto nel preambolo del regolamento, è quello relativo al concetto di "contenuto illegale": se infatti da un lato si afferma sia che per garantire un ambiente online sicuro, diventa centrale definire il concetto di "contenuto illegale", sia che è necessario venga assicurata la corrispondenza tra il concetto di contenuti illegali *online* e di contenuti illegali *offline*, dall'altro però non è data una specifica definizione di ciò che debba considerarsi illegale. A tale proposito, il regolamento fa riferimento – operando così un rimando a quanto previsto non solo dalle norme internazionali ma anche e soprattutto dagli ordinamenti nazionali – ad alcune categorie di atti definiti illegali, o perché considerati "di per sé illegali" oppure perché "riguardano attività illegali". Significativo, appare in tal senso il richiamo alla necessità che il termine "illegale" sia inteso in senso lato: i riferimenti, contenuti nel testo, alle attività che possano essere in tal modo qualificate hanno evidentemente carattere esemplificativo e non certo esaustivo[66].

Rispetto ai contenuti illegali viene poi delineato un sistema di controllo, nell'ambito del quale i *provider* sono chiamati a svolgere un ruolo centrale. Tale sistema fa perno sul meccanismo delle segnalazioni: quanto al suo funzionamento, da un lato è previsto che ciascun *provider* predisponga degli strumenti appositi, di facile utilizzo, che consentano agevolmente a qualunque utente di segnalare la presenza di contenuti illegali[67]; dall'altro è altresì previsto che per il tramite del coordinatore (nazionale) dei servizi digitali, all'interno di ciascuno Stato membro possano essere individuati i cosiddetti "segnalatori attendibili" (*trustflagger*), cioè soggetti cui è rimesso il compito di individuare e segnalare la presenza di contenuti illegali[68]. L'attivazione dei meccanismi di segnalazione diventa centrale in quanto

controllo sui contenuti immagazzinati nella propria piattaforma: diversamente invece, se una volta messo al corrente della natura illegale dei contenuti trasmessi attraverso i propri servizi, non avesse poi provveduto a rimuovere prontamente tali contenuti. Cfr. altresì *Comité de rédaction de Pravoye Delo et Shtekel v. Ukraine*, Comité n. 33014/05, 5 maggio 2011, par. 63; *Cicad v. Suisse*, ricorso n. 17676/09, 7 giugno 2016, par. 59; *M.L. and W.W. v. Germany*, n. 60798/10-6599/10, 28 giugno 2018, par. 91.

[64] Regolamento UE 2022/2065 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali), GUUE L 277, 27.10.2022.

[65] *Id.*, art.1. L'art. 2 specifica invece, con riferimento ai prestatori di servizi intermediari, che il regolamento sarà applicabile ai servizi offerti ai destinatari il cui luogo di stabilimento si trova nell'UE, indipendentemente da quale sia il luogo di stabilimento del prestatore di servizi intermediari.

[66] Sono menzionate, fra le altre, l'illecito incitamento all'odio o i contenuti terroristici illegali e i contenuti discriminatori illegali. *Id.*, preambolo, punto 12.

[67] *Id.* art. 16.

[68] Ai sensi dell'art. 22, la qualifica di segnalatore attendibile «viene riconosciuta, su richiesta di qualunque ente, dal coordinatore dei servizi digitali dello Stato membro in cui è stabilito il richiedente al richiedente che abbia dimostrato di soddisfare tutte le condizioni seguenti: a) dispone di capacità e competenze particolari ai fini dell'individuazione, dell'identificazione e della notifica di contenuti illegali; b) è indipendente da qualsiasi fornitore di piattaforme online; c) svolge le proprie attività al fine di presentare le segnalazioni in modo diligente,

elemento capace di modulare l'eventuale responsabilità dei *provider* nel caso vengano veicolati contenuti o messaggi illegali tramite i propri servizi. Da questo punto di vista, si delinea un sistema nel quale da un lato è confermato il principio dell'irresponsabilità/neutralità dei *provider*, posto che sugli stessi non grava un generale obbligo di sorveglianza[69]; dall'altro, però, tale neutralità ovvero assenza di responsabilità può essere fatta valere nella misura in cui sia dimostrato che il *provider* non era a conoscenza della presenza di contenuti illegali[70]. È proprio rispetto all'attività di segnalazione che si modula infatti l'intervento del *provider* diretto ad apporre restrizioni alle attività di condivisione di messaggi e contenuti: nel caso in cui questi ultimi siano ritenuti illegali, le restrizioni includono la limitazione della visibilità dei medesimi, la loro rimozione o ancora la disabilitazione dell'accesso a tali contenuti[71]. In un'ottica poi di stretta collaborazione con le autorità giudiziarie degli Stati membri, in capo ai *provider* è imposto l'obbligo di notifica all'autorità giudiziaria qualora si sia in possesso di informazioni da cui si desume che sia stato commesso o stia per essere commesso un reato[72]. Ulteriori adempimenti riguardano poi gli obblighi di trasparenza: i *provider* sono infatti tenuti a rendere note in termini chiari, inequivoci e facilmente accessibili, le misure e le procedure utilizzate – ivi comprese quelle relative alla gestione dei reclami – per effettuare il controllo sui contenuti[73].

Obblighi più stringenti sono poi previsti a carico dei *provider* di grandi dimensioni, si tratti di piattaforme (c.d. VLOP- *Very Large Online Platform*) o motori di ricerca online (c.d.

---

accurato e obiettivo». È peraltro previsto che nell'individuazione dei segnalatori attendibili sia data preferenza ad enti – pubblici e non – e non a persone, che hanno dimostrato, tra l'altro, di disporre di capacità e competenze particolari nella lotta ai contenuti illegali e di svolgere le proprie attività in modo diligente, accurato e obiettivo, impegnati sul fronte della segnalazione e notifica dei contenuti razzisti e xenofobi, nonché sul contrasto a fenomeni come terrorismo e pedofilia. Cfr. Preambolo, punto 61. Per quanto riguarda l'Italia, merita rilevare che con delibera n. 26/25/CONS del 22 gennaio 2025, l'AGCOM (Autorità garante per le garanzie nelle comunicazioni) ha provveduto a rilasciare la prima qualifica di segnalatore attendibile alla società Argo Business Solutions S.r.l. Cfr. https://www.agcom.it/sites/default/files/provvedimenti/delibera/2025/Delibera%2026%2025%20CONS_Argo%20Business%20Solutions_segnalatore%20attendibile.pdf.

[69] Regolamento UE 2022/2065, cit., art. 8.

[70] *Id.*, art. 16. Indicazioni in questo senso erano già state date dalla Corte europea dei diritti umani: cfr. *Delfi v. Estonia*, cit., *supra* note 44 e 52, par. 159; *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, 2947/13, 2 maggio 2016, par. 91. Nel caso *Sanchez v. France*, n. 45581/2015, 15 maggio 2023, la Corte con riferimento ai commenti pubblicati su un account Facebook per i quali l'autore aveva subito dalle autorità francesi una condanna penale, ribadisce che «(…) Nevertheless, to exempt producers from all liability might facilitate or encourage abuse and misuse, including hate speech and calls to violence, but also manipulation, lies and disinformation. In the Court's view, while professional entities which create social networks and make them available to other users necessarily have certain obligations (..), there should be a sharing of liability between all the actors involved, allowing if necessary for the degree of liability and the manner of its attribution to be graduated according to the objective situation of each one» (par. 185). In dottrina cfr. L. BRUNER, *The Liability of an Online Intermediary/Third Party Content: the Watchdog becomes the Monitor: Intermediary Liability After Delfi*, in *Human Rights Law Review*, 2016, p. 163 ss.; I. GASPARINI, *L'odio ai tempi della rete: le politiche europee di contrasto all' "online hate speech"*, in *Jus*, 3/2017, pp. 505-532; T. LONGKE, *On an Internet Service Provider's Content Management Obligation and Criminal Liability*, in *Journal of Eastern-European Criminal Law*, 2019, p.145 ss.; G. RUOTOLO, *Le proposte europee di riforma della responsabilità dei fornitori di servizi su Internet*, in *Rivista italiana di informatica e diritto*, 2022, p. 19 ss.; G. FROSIO, C. GEIGER, *Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime*, in *European Law Journal*, 2023, p. 31ss.;

[71] Regolamento UE 2022/2065, cit., art. 23.

[72] *Id.*, art. 18.

[73] *Id.*, art. 14.

VLOSE- *Very Large Online Search Engines*)[74]: tali obblighi sono concepiti nell'ottica di mitigare i maggiori rischi sistemici che, date le dimensioni dei VLOP/VLOSE, sono connessi con la progettazione ed il funzionamento dei sistemi e dei servizi offerti. Su tale categoria di soggetti grava dunque l'obbligo di operare una valutazione periodica di tali rischi fra i quali rientrano quelli legati sia alla diffusione di contenuti illegali tramite i propri servizi, sia agli eventuali effetti negativi che potrebbero prodursi sul godimento di certi diritti garantiti dalla Carta dei diritti e delle libertà fondamentali dell'Unione europea, fra cui il diritto al rispetto della vita privata e familiare, il diritto a non essere discriminati, ma anche il diritto alla libertà di espressione e di informazione[75]. È altresì previsto che i soggetti pubblici possano chiedere di avere accesso ai dati così da verificare se il sistema ed i meccanismi approntati sono in linea con quanto previsto dal regolamento.

Merita da ultimo fare cenno a quanto previsto dall'articolo 45 del DSA, ai sensi del quale la Commissione e il Comitato dei servizi digitali[76] incoraggiano e agevolano l'adozione, da parte dei *provider*, di codici di condotta intesi a garantire la corretta applicazione del regolamento «tenendo conto in particolare delle sfide specifiche connesse alla lotta ai diversi tipi di contenuti illegali e ai rischi sistemici (…)».

È a tale disposizione che si ricollega l'adozione, nel 2024, del *Code of conduct on countering illegal hate speech online*+ da parte degli iniziali sottoscrittori del precedente codice di condotta siglato nel 2016, cui sono andate ad aggiungersi, di recente, altre VLOP[77]. Il recente codice di condotta ha il dichiarato intento di rafforzare la portata del previgente codice di condotta sulla base delle disposizioni contenute nel DSA, nell'ottica di conferire maggiore efficienza alla lotta contro l'*hate speech* online. Occorre sottolineare, preliminarmente, che il codice fa riferimento ad una definizione molto ampia di *hate speech* quale risulta dal dettato della decisione quadro 2008/913/GAI, nonché dalle previsioni legislative in vigore nei diversi Stati, che hanno dato attuazione alla decisione quadro, ma altresì dai «possible forthcoming updates to this Framework Decision, where relevant»[78]. Da questo punto di vista, è intuibile che ad integrare il quadro normativo potrebbero intervenire atti eventualmente adottati in accoglimento della proposta della Commissione relativa all'inserimento dell'*hate speech* e degli *hate crimes* nella lista dei reati di cui all'articolo 83 del TFUE[79], così come pure le modifiche legislative introdotte negli ordinamenti nazionali a seguito, per esempio, della direttiva 2024/1385 in materia di contrasto alla violenza di genere, per quel che specificamente concerne in particolare il *gender-based hate speech*. Per quanto invece concerne più specificamente la tipologia degli impegni assunti, il Codice prevede innanzitutto che i *provider*, in ottemperanza a quanto disposto dall'articolo 16 del DSA, predispongano i meccanismi

---

[74] Regolamento UE 2022/2065, cit., art. 33, parr. 1 e 4 ai sensi del quale sono denominati piattaforme online o motori di ricerca online di grandi dimensioni le piattaforme o i motori di ricerca che hanno un numero medio mensile di destinatari attivi del servizio nell'Unione, pari o superiore a 45 milioni.

[75] *Id*., art. 34.

[76] *Id*., art. 61, per quel che concerne la natura, le funzioni e la composizione del Comitato.

[77] Cfr. *Code of Conduct on Countering Illegal Hate Speech Online* +. Il codice adottato nel 2024 è stato sottoscritto oltre che dagli iniziali firmatari del Codice del 2016 (Facebook, Instagram, YouTube, X-Twitter e Microsoft) anche dalle piattaforme che successivamente avevano aderito al Codice, più precisamente Instagram, Dailymotion, Snap Inc., Jeuxvideo.com, TikTok, LinkedIn, Rakuten Viber and Twitch. file:///C:/Users/utente/Downloads/Code_of_Conduct_o_Countering_Illegal_Hate_Speech_Online__QM9XoUhsQ701NJ5pKu6RgV8gFpk_111777%20(2).pdf.

[78] *Id*., nota 4.

[79] Comunicazione della Commissione al Parlamento europeo e al Consiglio, *Un'Europa più inclusiva e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio*, 9 dicembre 2021, COM(2021)777.

necessari per rendere possibile, sia ai *trusted flagger* che a qualunque altro utilizzatore sul territorio dell'Unione europea, la segnalazione della presenza di contenuti qualificabili come *hate speech*; in secondo luogo è previsto che i *provider* intervengano in modo tempestivo – entro le 24 ore deve essere esaminato almeno il 50% dei contenuti segnalati – diligente e non arbitrario, per rimuovere i contenuti o per disabilitare l'accesso ai medesimi; in terzo luogo è previsto l'impegno a valutare, sulla base della metodologia concordata con la Commissione europea, la conformità delle misure adottate, per quanto concerne l'intervento sui contenuti segnalati, agli impegni assunti e a monitorare nel tempo le tendenze in atto. Da ultimo, sono individuati specifici impegni intesi sia a garantire la creazione di un quadro di cooperazione finalizzato a rafforzare l'interscambio tra le VLOP e gli attori della società civile, sia a rafforzare il ruolo delle VLOP nella ricerca di strumenti innovativi e nella promozione di campagne di informazione, di educazione e di *awarness raising* attraverso le proprie piattaforme.

6. *Considerazioni conclusive*

La strategia di contrasto all'*hate speech* si è caratterizzata negli anni per avere fatto perno soprattutto sulla lettura e l'interpretazione delle norme internazionali poste a garanzia del diritto alla libertà di espressione e, contestualmente, delle norme che affermano il fondamentale divieto di discriminazione nelle sue diverse forme. In questa direzione si sono mossi gli organi di controllo sulla tutela dei diritti umani fin dall'epoca precedente l'adozione di strumenti specifici dedicati all'*hate speech* ed in assenza, soprattutto, di una definizione univoca ed unitaria di cosa debba intendersi con tale espressione.

Dall'analisi condotta è emerso che i problemi di tipo definitorio e semantico, per quel che concerne l'*hate speech*, continuano a sussistere: i tratti caratterizzanti del fenomeno sono ricavabili dal contenuto di diversi strumenti internazionali, specie di *soft law*, adottati in materia, senza che però si sia pervenuti ad incasellare entro confini certi e precisi un fenomeno che resta multiforme e talvolta difficile da individuare. L'assenza di una definizione di *hate speech*, univoca e generalmente accolta, si riflette nell'approccio assai eterogeneo al tema che caratterizza gli ordinamenti interni: più correttamente può anzi dirsi sia, da tale eterogenea attitudine, fondamentalmente motivata.

Il quadro normativo internazionale sembra caratterizzarsi dunque per la presenza di strumenti non vincolanti dedicati all'*hate speech* e di strumenti vincolanti, peraltro non numerosi, pensati per combattere atti, di diversa tipologia, che abbiano natura razzista e xenofoba. Sia gli uni che gli altri – i primi per definizione, gli altri per il modo in cui sono stati concepiti gli obblighi imposti allo Stato – tendono a lasciare a quest'ultimo un margine di discrezionalità nella scelta delle misure da adottare. L'applicazione di queste ultime – siano esse il prodotto dell'adattamento alle norme internazionali ovvero il frutto dell'iniziativa del legislatore interno – ha condotto, in molti casi, ad una compressione del diritto alla libertà di espressione, alimentando il contenzioso interno e creando altresì i presupposti perché fosse richiesto l'intervento del giudice internazionale o dell'organo di monitoraggio sul rispetto dei diritti umani. È ad essi, infatti, che è stato rimesso il compito di individuare volta per volta, quali tipologie di discorsi o di manifestazioni di pensiero ed opinioni potessero essere equiparate ad un esercizio del diritto alla libertà di espressione idoneo a compromettere diritti altrui, egualmente tutelati, ovvero interessi della collettività. La giurisprudenza ha così

costantemente offerto indicazioni che, se non sono servite a colmare una lacuna normativa, hanno però contribuito a delineare contorni e contenuti di quel fenomeno definito *hate speech*: non è azzardato ritenere che proprio quelle indicazioni abbiano costituito la base di riferimento per l'elaborazione delle definizioni di *hate speech* che si ritrovano nel testo degli atti adottati in materia.

Al giudice internazionale o all'organo di monitoraggio, che non può fare riferimento a dei parametri normativi certi, entro i quali incasellare il fenomeno dell'*hate speech*, è stato ed è tutt'ora richiesto uno sforzo evidente per garantire il contemperamento di esigenze teoricamente contrapposte: da un lato, quella di tutelare la libertà di espressione, dall'altra quella di contrastare le manifestazioni del pensiero e delle opinioni, nella misura in cui queste non appaiono semplicemente offensive, denigratorie o anche odiose nei contenuti, ma sono idonee a ledere diritti altrui o fondamentali interessi della comunità. Garantire la tutela di un diritto fondamentale, "posto a cardine di qualsiasi società democratica", ha dunque significato e significa stabilire se ed in che misura l'intervento dello Stato diretto a contrastare l'*hate speech* sia capace di tradursi nell'adozione di misure che, comprimendo la libertà di espressione inibiscono l'esercizio del diritto di opinione ed espressione producendo quello che la Corte europea dei diritti umani ha definito, in più occasioni, il *chilling effect*. Di volta in volta, dunque, sono stati individuati non solo i limiti entro i quali il diritto alla libertà di espressione può essere riconosciuto e garantito, ma contestualmente si è provveduto a definire l'ampiezza del margine di apprezzamento del singolo Stato nella scelta delle misure da adottare, idonee a comprimere quel medesimo diritto: da questo punto di vista, l'attività dell'organo di controllo ha contribuito – in linea con la tendenza affermatasi allorché sono stati adottati strumenti di *soft law* specificamente dedicati al tema dell'*hate speech* – a tracciare una linea di demarcazione tra le diverse manifestazioni del pensiero, riconducibili a quella magmatica fattispecie qual è l'*hate speech*, sulla base di un criterio che fa perno sulla gravità della condotta. Tale valutazione ha tenuto conto, caso per caso, di un insieme di elementi: dal contesto e dalle circostanze, alla necessità di assicurare un bilanciamento tra diritti egualmente garantiti, che è quanto dire dell'idoneità di certe manifestazioni del pensiero e delle opinioni a compromettere il godimento di altri diritti; dal possibile impatto di certi messaggi e/o contenuti sia sul contesto sociale sia nei confronti degli eventuali specifici destinatari, all'impatto sulla sicurezza e sulla morale pubblica. L'impostazione così descritta ha caratterizzato anche l'approccio ai casi che hanno avuto ad oggetto la dimensione digitale dell'*hate speech*: il giudice e l'organo di controllo hanno fornito alcune indicazioni che hanno poi trovato riscontro non solo negli strumenti di *soft law* adottati con riferimento allo specifico fenomeno dell'*hate speech online* e dunque nei codici di condotta, ma altresì nello strumento normativo – il DSA – col quale il legislatore europeo ha cercato di definire un assetto normativo idoneo ad affrontare le specifiche problematiche poste dal fenomeno nella sua dimensione digitale. Rispetto a tale intervento possono essere fatte diverse considerazioni. La prima è che nemmeno nel testo del DSA è dato rinvenire una definizione di *hate speech*: uno strumento, pensato per garantire "un ambiente *online* sicuro, prevedibile e affidabile" non definisce i contenuti del fenomeno, idoneo a rendere l'ambiente *online* insicuro, imprevedibile e inaffidabile, che intende contrastare e rimanda, per l'individuazione di quello che può essere classificato come "contenuto illegale" a quanto previsto dal diritto applicabile. Non solo, ma il problema dell'individuazione dei contenuti illegali, che è quanto dire della perimetrazione del diritto alla libertà di espressione, quando l'esercizio del medesimo avvenga attraverso Internet, è risolto demandando ai *provider* il compito di vigilare sul modo in cui le piattaforme vengono utilizzate al fine di individuare ed eventualmente rimuovere i contenuti illeciti:

proprio con riferimento a questo aspetto, l'adozione del primo codice di condotta del 2016 aveva sollevato non pochi interrogativi sui risvolti connessi con il conferimento – in qualche modo la delega – di poteri di controllo a soggetti privati. Con l'adozione del DSA, l'attività di controllo continua ad essere demandata ai *provider*: si è cercato però di apportare dei correttivi al sistema, innanzitutto delineando un sistema di regole che mira ad incardinare l'azione dei *provider* e in particolare delle piattaforme di grandi dimensioni all'interno di un quadro di principi e di norme poste a salvaguardia di diritti fondamentali. Al di là ed oltre l'iniziativa che i *provider* possono intraprendere su base volontaria, così come previsto dai codici di condotta, si pongono dunque le norme volte ad imporre obblighi specifici di diversa natura, che oltre a rendere possibile l'esercizio dell'attività di controllo, incanalano l'attività dei *provider* in un quadro più ampio di sinergia e collaborazione con l'autorità giudiziaria dei singoli Stati nel contrasto ai contenuti illeciti. Tuttavia, malgrado lo sforzo normativo, alcune criticità permangono, specie se si considera che il portato degli obblighi posti in capo ai *provider* quanto all'individuazione di contenuti che potrebbero avere un impatto negativo su diritti garantiti in particolare dalla Carta dei diritti fondamentali dell'Unione europea, è suscettibile di generare la tendenza alla sopravvalutazione del rischio, con una conseguente indebita compressione della libertà di espressione. È chiaro infatti che, a differenza dello Stato che si vede riconoscere un margine di apprezzamento nel valutare in che misura l'esercizio della libertà di espressione possa essere compresso (sul piano legislativo) e dunque nell'applicare (sul piano giudiziario) le misure limitative previste dalla legge, il *provider* è obbligato ad operare, entro il perimetro definito dalle norme del regolamento, in maniera pressoché automatica. Se infatti da un lato sussiste per essi l'obbligo di immediata segnalazione all'autorità giudiziaria per i casi in cui vi sia il sospetto che un reato sia stato commesso o stia per esserlo, per tutti gli altri casi il *provider* è tenuto a individuare i contenuti illeciti avendo quale parametro di riferimento le norme applicabili, sia quelle internazionali, recepite nell'ordinamento interno, sia quelle che sono il frutto dell'iniziativa del legislatore nazionale. Tale parametro non è tuttavia uniforme, considerato che, come più volte, osservato, gli ordinamenti interni si pongono sia rispetto al problema dell'individuazione di contenuti illegali, sia rispetto al problema della definizione di *hate speech*, in maniera differente. Da qui il ricorso al meccanismo automatizzato e all'algoritmo cui è affidata l'individuazione dei contenuti illeciti, con tutto che ciò che può conseguirne in termini di falsi positivi, di mancata individuazione di contenuti che realmente illeciti e di sostanziale incertezza quanto alla capacità del sistema automatizzato di rispondere alle esigenze di tutela di diritti fondamentali. I rischi cui si è fatto sopra cenno permangono, lasciando dunque ancora aperto non solo il dibattito, ma anche lo spazio per il contenzioso futuro.