Università degli Studi di Padova

Dipartimento di Diritto Pubblico, Internazionale e Comunitario

Corso di Laurea Triennale in

Diritto e Tecnologia (L-14)

Tesi di Laurea

# *I rischi della datificazione per la libertà di espressione: un'analisi dell'hate speech sui social media*

Relatore

Prof. Claudio Sarra

Laureando

Pierpaolo Romagnoli

n° matr. 2012669 / L-14

Anno Accademico 2022/2023

INDICE

## Introduzione

Negli ultimi anni, l'Intelligenza Artificiale (abbreviata IA), il concetto di datificazione e Big Data sono gli argomenti di discussione che più influenzano i governi e la società. L'origine dell'IA, contrariamente a ciò che si crede, non risale ad un tempo recente, bensì a più di settant'anni fa. Si iniziò a parlare di IA attorno al 1950, quando il matematico e informatico britannico Alan Turing scrisse un articolo intitolato *Computing machinery and intelligence*, in cui presentava un test (definito in seguito Test di Turing) secondo cui una macchina poteva essere considerata intelligente se il suo comportamento, osservato da un essere umano, fosse risultato indistinguibile da quello di una persona[1]. Verso gli anni '60 e '70 il campo dell'Intelligenza Artificiale vide una notevole crescita grazie alla creazione di algoritmi e tecniche di programmazione che consentivano alle macchine di apprendere ed elaborare dati. Successivamente nel 1982 venne sviluppato dalla Digital Equipment il primo sistema di IA utilizzato in ambito commerciale, denominato R1. Lo scopo del programma era quello di aiutare a configurare gli ordini di nuovi computer[2]. Esistono diversi tipi di approcci all'IA, Forte[3] e Debole[4]: nel primo caso si ritiene che la macchina non sia più un semplice strumento in quanto, se programmata in maniera opportuna, potrebbe assumere le caratteristiche della mente di un essere umano, se non perfino diventare superiore. Nel secondo caso, invece, si nega che la macchina possa sviluppare coscienza delle attività svolte, nonostante siano molto complesse. Gli obbiettivi dell'IA sono replicare e migliorare l'intelligenza umana nel riconoscimento e nella previsione dei modelli predittivi[5]. Uno dei più grandi problemi del passato riguardava la reperibilità dei dati: una problematica che a oggi è superata. L'utilizzo dell'intelligenza artificiale non è però priva di conseguenze nel mondo del lavoro e nella

[1] A. M. Turing, Mind a quarterly review of psychology and philosophy, I. Computing Machinery and Intelligence, Vol. Lix. No. 236, ottobre 1950, pp. 433-460

[2]John McDermott, R1: a rule-based configurer of computer systems, aprile 1980, Carnegie-Mellon University.
Disponibile al seguente link < https://apps.dtic.mil/sti/tr/pdf/ADA223957.pdf >

[3] Henry Shevlin, Karina Vold, Matthew Crosby & Marta Halina, The limits of machine intelligence, Embo reports, Volume 20 Issue 10, dell'ottobre 2019, pag. 1-5.
Disponibile al seguente link: < https://www.embopress.org/doi/epdf/10.15252/embr.201949177>

[4] George Dvorsky, How Much Longer Before Our First AI Catastrophe? Pubblicato nell'aprile 2013
Articolo disponibile al seguente link: < https://gizmodo.com/how-much-longer-before-our-first-ai-catastrophe-464043243>

[5]Daron Acemoglu e Pascual Restrepo, The wrong kind of AI? Artificial Intelligence and the future of labor demand, Working Paper 25682, National Bureau of Economic Research, marzo 2019, pp.1-10.
Disponibile al seguente link: <https://www.nber.org/system/files/working_papers/w25682/w25682.pdf>

società in generale, risulta perciò necessario creare uno studio che analizzi questo fenomeno nel suo complesso.

Questa ricerca nasce con l'obiettivo di spiegare il ciclo della datificazione e di come possa essere utilizzata per contrastare il fenomeno denominato *hate speech*, definito anche come *discorso d'odio***.** Dopo aver descritto nel primo capitolo che cosa si intende per datificazione e perché è importante, segue nel secondo capitolo un approfondimento sulle sfide etiche e sul contrasto dei discorsi d'odio nel web attraverso l'analisi delle normative vigenti. Infine, nel terzo capitolo vengono esaminati diversi effetti sui soggetti e i vari metodi di risoluzione delle possibili problematiche. Tutte queste tecnologie, se utilizzate nel modo corretto, possono portare a migliorare la vita di ogni singolo individuo, ma se usate nel modo sbagliato, si possono rivelare come una vera e propria arma di battaglia, minando la sicurezza di tutti.

# Capitolo 1

## Il concetto di datificazione e di hate speech

### 1.1 La datificazione: definizione e caratteristiche

Nel corso del tempo, la comunicazione e l'informazione hanno rappresentato per la società e per l'individuo una fonte di controllo sociale da parte sia di Istituzioni sia da parte delle società che offrono servizi via Internet. Con l'avvento dell'accesso a Internet, avvenuto intorno alla fine degli anni Novanta, si verificò un incremento di tale fenomeno[6]: a oggi, infatti, si può affermare che qualsiasi informazione costituisca un dato, il quale può essere classificato come: pubblico, riservato, personale (che a sua volta include il dato sensibile). Per dati pubblici si intendono dati accessibili al pubblico. Sono informazioni che vengono raccolte, conservate e messe a disposizione da enti governativi, organizzazioni di vario genere, istituzioni accademiche o altre fonti affidabili. Questi dati possono essere resi disponibili in vari formati, tra cui testo, tabelle, grafici, file audio o video, e possono riguardare diverse aree, come demografia, economia, ambiente, salute, istruzione, trasporti. Possono essere utilizzati da cittadini, ricercatori, giornalisti, aziende e sviluppatori per condurre analisi, prendere decisioni informate, valutare l'efficacia delle politiche pubbliche, promuovere l'innovazione e risolvere problemi sociali. Alcuni esempi di dati pubblici sono:

1. dati demografici: informazioni sulla popolazione, come età, genere, razza, istruzione, occupazione e reddito;
2. dati economici: statistiche sull'economia di una regione o di un paese, come il PIL, l'occupazione, l'inflazione, il commercio internazionale e le statistiche sulle imprese;
3. dati ambientali: dati sul clima, l'inquinamento atmosferico, l'inquinamento idrico, la biodiversità e altre informazioni relative all'ambiente;
4. dati sanitari: informazioni sulla salute della popolazione, come le statistiche di mortalità, la prevalenza delle malattie, le risorse sanitarie e i dati sull'accesso ai servizi sanitari;

[6] Claudio Sarra, Il mondo-dato, 2019, Cleup SC, pag. 17.

I dati pubblici possono essere accessibili attraverso portali online quali siti web governativi, o possono essere ottenuti tramite richieste specifiche in base alle leggi sull'accesso all'informazione di un paese o di un'organizzazione. È importante tener presente che l'accesso ai dati pubblici può variare da Stato a Stato e può essere soggetto a limitazioni per proteggere la privacy.

I dati riservati sono informazioni sensibili o confidenziali che, a causa del loro potenziale impatto negativo qualora venissero divulgati a soggetti non autorizzati, richiedono una protezione speciale. Questi dati sono spesso associati a questioni personali, finanziarie, mediche, commerciali o di sicurezza[7]. Le Informazioni Personali Identificabili (PII, Personally Identifiable Information) includono nomi, indirizzi, numeri di telefono, numeri di carta di credito utili a identificare in modo univoco gli utenti[8]. La protezione dei dati riservati è fondamentale per prevenire violazioni della privacy, furti di identità, frodi finanziarie, danni reputazionali o altre conseguenze dannose. Per questo motivo le organizzazioni e le Istituzioni che raccolgono e gestiscono dati riservati devono adottare misure di sicurezza adeguate, come l'uso di crittografia, l'accesso limitato ai dati, la sicurezza delle reti, la formazione del personale e il rispetto delle leggi e delle normative sulla privacy.
Per ovviare a questo problema, sono state emanate leggi e regolamenti specifici, come il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea, che stabilisce obblighi e responsabilità per la gestione dei dati riservati e le conseguenze per la violazione di tali norme.

I dati personali includono qualsiasi informazione relativa a una persona fisica identificata o identificabile[9]. Sono di natura personale e possono essere utilizzati per individuare o contattare una persona specifica. Rientrano in questa categoria:

---

[7] Garante per la protezione dei dati personali, Cosa intendiamo per dati personali?
Disponibile al seguente link: <https://www.garanteprivacy.it/home/diritti/cosa-intendiamo-per-dati-personali#:~:text=i%20dati%20rientranti%20in%20particolari,salute%20o%20alla%20vita%20sessuale>
[8] U.S. Department of Labor, Guidance on the Protection of Personal Identifiable Information.
Disponibile al seguente link:
<https://www.dol.gov/general/ppii#:~:text=Personal%20Identifiable%20Information%20(PII)%20is,either%20direct%20or%20indirect%20means>
[9]Vedi nota 7
Disponibile al seguente link: <https://www.garanteprivacy.it/home/diritti/cosa-intendiamo-per-dati-personali#:~:text=una%20persona%20fisica.-,Sono%20dati%20personali%20le%20informazioni%20che%20identificano%20o%20rendono%20identificabile,sua%20situazione%20economica%2C%20ecc>

1. le informazioni di identificazione quali nome, indirizzo, numero di telefono, indirizzo e-mail, numero di previdenza sociale, numero di passaporto, numero di patente di guida, data di nascita, luogo di nascita, sesso, nazionalità e altre informazioni di identificazione simili;
2. i dati finanziari, tra cui conti bancari, carte di credito, informazioni fiscali, reddito, storia del credito e altre informazioni finanziarie personali;
3. i dati di geolocalizzazione come informazioni relative alla posizione di una persona in tempo reale o storica, ad esempio attraverso un dispositivo GPS o dati di localizzazione di un'applicazione;
4. i dati online, ossia informazioni relative all'utilizzo di Internet o di servizi online, come indirizzo IP, cronologia di navigazione, attività sui social media, preferenze di ricerca, cookie.

I dati sensibili, una sottocategoria dei dati personali, hanno la caratteristica di essere particolarmente delicati o critici. Richiedono un'attenzione particolare e una protezione rigorosa per evitare un utilizzo improprio o non autorizzato. Alcuni esempi di dati sensibili includono:

1. informazioni sulla salute mentale o sessuale, che riguardano condizioni o trattamenti medici;
2. origine etnica o razza, considerati sensibili a causa del rischio di discriminazione o pregiudizio;
3. orientamento sessuale o preferenze sessuali, considerate sensibili per proteggere la privacy e prevenire la discriminazione;
4. credenze religiose o filosofiche, considerate sensibili a causa della loro natura personale e del potenziale rischio di persecuzione o discriminazione;
5. dati biometrici, che includono impronte digitali, scansioni della retina, riconoscimento facciale o altre caratteristiche fisiche uniche di un individuo, considerati sensibili perché possono essere utilizzate per l'identificazione personale.[10]

[10] Vedi nota 7
Disponibile al seguente link: <https://www.garanteprivacy.it/home/diritti/cosa-intendiamo-per-dati-personali#:~:text=i%20dati%20rientranti%20in%20particolari,salute%20o%20alla%20vita%20sessuale>

È fondamentale proteggere i dati sensibili implementando misure di sicurezza, come la crittografia, l'accesso limitato ai dati, l'anonimizzazione – che «è un processo mediante il quale i dati personali vengono alterati in modo irreversibile in modo tale che un soggetto interessato non possa più essere identificato direttamente o indirettamente, né dal titolare del trattamento dei dati da solo né in collaborazione con altre parti»[11] – o la pseudonimizzazione dei dati– che «è il trattamento di dati personali in modo tale che i dati personali non possano più essere attribuiti a un soggetto specifico senza l'uso di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e siano soggette a misure tecniche e organizzative per garantire che i dati personali non siano attribuiti a una persona fisica identificata o identificabile»[12]. La violazione dei dati sensibili può avere conseguenze gravi per la privacy delle persone coinvolte e può portare a gravi ripercussioni legali ed etiche.

Codeste tipologie di dati possono essere raccolte ed elaborate per ottenere informazioni in merito a una situazione, un evento o un oggetto. Tale concetto prende il nome di datificazione e si intende «il fenomeno, tipico della società attuale, della traduzione massiva di accadimenti dell'esperienza in rappresentazioni simboliche manipolabili (dati) i quali sono riorganizzati e utilizzati al fine di ricavare forme più sofisticate di conoscenza relative all'andamento del campo così organizzato»[13]. La datificazione è un processo fondamentale nell'era dell'informazione in cui viviamo, dove i dati sono diventati una risorsa di valore inestimabile. Essa rappresenta la trasformazione dei dati non strutturati, ossia informazioni che non seguono un particolare schema predefinito o una struttura organizzata. A differenza dei dati strutturati, questi ultimi non sono organizzati in tabelle, colonne o righe e molto spesso non contengono informazioni facilmente estraibili o

[11]ENISA (European Union Agency for Cybersicurity), Pseudonymisation techniques and best practiesnovembre 2019, pag.9, tradotto dall'inglese «is a process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party».
https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices
[12] Vedi sopra, tradotto dall'inglese «is the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person».
[13]Claudio Sarra, Datificazione e ingegneria del simbolico, in Saggi a margine dei seminari virtuali di Journal of Ethics and Legal Tecnologies, Primiceri Editore, Padova 2022, pag. 155.
Consultabile al seguente link:
<https://www.researchgate.net/publication/362412588_Datificazione_e_ingegneria_del_simbolico>

interpretabili dai computer anche se contengono informazioni potenziali. I dati possono essere anche semi-strutturati, che spesso sono considerati alla stregua dei non strutturati, ma che contengono alcune caratteristiche di quelli strutturati[14] in una forma organizzata e analizzabile.

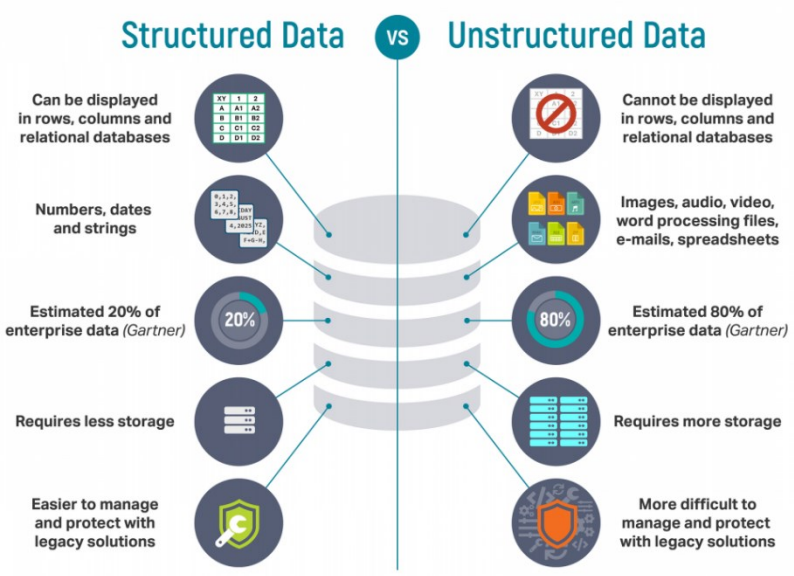


*Figura 1: differenza tra dati strutturati e non strutturati*

La trasformazione dei dati non strutturati consente di sfruttarne appieno il potenziale, possono infatti essere utilizzati per prendere decisioni informate, ottenere informazioni significative e scoprire nuovi modelli.

Il processo di datificazione si articola in diverse fasi: la prima è la raccolta dei dati, in cui vengono identificate le fonti pertinenti, come documenti, database, file di testo o dati provenienti da dispositivi. Una volta che i dati vengono raccolti devono essere pre-processati in modo da renderli adatti per l'elaborazione successiva. Durante questo processo vengono effettuati diversi passaggi per garantirne la qualità e la gestibilità. In primo luogo, vengono eliminati i dati errati, incompleti o non leggibili; a seguire i dati

[14] Mohamed Y. Eltabakh, Mayuresh Kunjir, Ahmed Elmagarmid, Mohammad Shahmeer Ahmad, Cross Modal Data Discovery over Structured and Unstructured Data Lakes, Proceedings of the VLDB Endowment Volume 16 Issue 11, agosto 2023, pagg.1-17.
Disponibile al seguente link:
<https://www.researchgate.net/publication/373369439_Cross_Modal_Data_Discovery_over_Structured_and_Unstructured_Data_Lakes>

vengono “normalizzati”, o semplicemente ridotti di dimensione, soprattutto nel caso in cui siano in formato analogico, al fine di renderli più gestibili.

La fase successiva si concentra sulla trasformazione dei dati e ha l'obiettivo di organizzarli in una forma strutturata. Questo viene fatto comunemente utilizzando modelli o schemi specifici. Ad esempio, i dati testuali vengono suddivisi in parole o frasi, mentre i dati numerici vengono organizzati in tabelle o matrici. Durante questa fase è possibile applicare varie tecniche, come l'estrazione delle caratteristiche o la riduzione della dimensionalità, che consentono di individuare le informazioni più rilevanti o di ridurre la complessità dei dati, al fine di semplificarne l'analisi e l'elaborazione successiva. [15] Per ottenere informazioni significative basterà applicare una serie di tecniche ai dati ormai strutturati, tra i quali metodi statistici, algoritmi di apprendimento automatico, tecniche di visualizzazione dei dati. Queste metodologie consentono di identificare modelli, tendenze o relazioni nascoste all'interno dei dati e di formulare previsioni o suggerimenti basati sui risultati ottenuti.

L'ultima fase del processo di datificazione riguarda la presentazione e l'utilizzo dei risultati dell'analisi. Le informazioni ottenute possono essere presentate attraverso dashboard, grafici, report o altre forme di visualizzazione, che facilitano la comprensione e l'interpretazione dei risultati. Queste informazioni possono quindi essere utilizzate per prendere decisioni, pianificare nuove strategie aziendali o addirittura migliorare le prestazioni o sviluppare nuovi prodotti e servizi.

Il termine datificazione ha un significato totalmente diverso dal termine digitalizzazione, la quale invece consiste nel prendere contenuti analogici, tra cui libri, film, fotografie, e convertirli in informazioni digitali, ossia in una sequenza binaria di 1 e 0 che i computer sono in grado di decifrare. La datificazione è un'attività molto più ampia: prende tutti gli aspetti della vita quotidiana e li trasforma in dati.[16] Per questo motivo offre numerosi vantaggi e opportunità, permettendo di trarre valore dai dati che altrimenti rimarrebbero inutilizzati o difficili da comprendere. Aiuta, inoltre, a identificare problemi o opportunità

[15] Giorgio Grossi professore dell’Università degli Studi di Milano Bicocca, La nostra esistenza “datificata”: ecco la nuova era del digitale, Agenda Digitale, 2023.
Disponibile al seguente link < https://www.agendadigitale.eu/cultura-digitale/la-principale-conseguenza-iperevolutiva-della-rivoluzione-digitale-lesistenza-datificata/>
[16] Kenneth Cukier and Viktor Mayer-Schoenberger, The Rise of Big Data, MAY/JUNE 2013, Published by: Council on Foreign Relations, p.35

nascoste, a ottimizzare processi, a migliorare la precisione delle previsioni e a prendere decisioni basate su dati concreti.

Alcuni esempi di datificazione possono essere trovati nel mondo delle Big Tech: Facebook ha raccolto dati sulla nostra rete di amicizie e conoscenze, nonché sui contenuti di nostro gradimento a cui abbiamo messo "mi piace"; Google ha registrato dati relativi alle ricerche effettuate e alla cronologia delle attività; Waze e Google Maps hanno tracciato i percorsi di guida e i luoghi che abbiamo visitato.

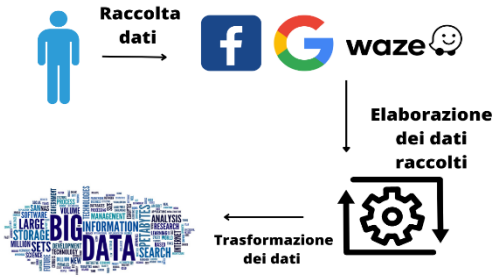


*Figura 2: rappresentazione di come vengono trasformati i dati grezzi*

Un altro concetto chiave riguarda il termine Big Data, il quale è strettamente correlato con la datificazione. Essi sono grandi quantità di informazioni provenienti da varie fonti, come transazioni online, e-mail, ricerche su Internet, video etc. Sono caratterizzati dalle cosiddette tre "V", Volume, Velocità, Varietà[17], dove:

1. il Volume si riferisce alla grande quantità di informazioni che non possono essere raccolte utilizzando le tecnologie tradizionali;
2. la Velocità non è altro che la celerità con cui si registrano i dati;
3. la Varietà riguarda le diverse tipologie di dati disponibili. [18]

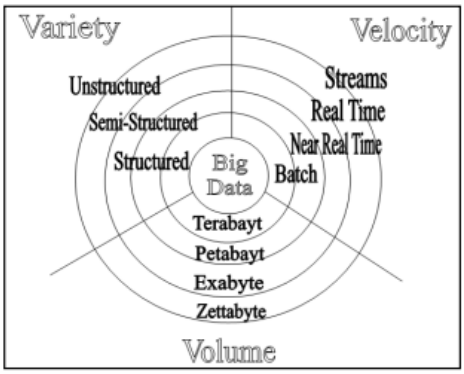


*Figura 3: rappresentazione delle 3 V dei big Data*

A questa tripartizione classica, possono essere aggiunte la Veridicità che si riferisce alla qualità dei dati, e il Valore riferito sia all'utilità dei dati che alla redditività delle informazioni recuperate dall'analisi dei Biga Data[19].

[17] Seref SAGIROGLU and Duygu SINANC, Big Data: A Review, May 2013, Published by: Gazi University Department of Computer Engineering, Faculty of Engineering, Ankara, Turkey,
[18] C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P.C. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw-Hill Companies, 2011, pag.26. Disponibile al link: < https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/I111025E.pdf >
[19] Ceylan Onay and Elif Öztürk, A review of credit scoring research in the age of Big Data, Journal of Financial Regulation and Compliance Vol. 26 No. 3, 2018, pag.382.

I Big data consentono di acquisire informazioni molto più approfondite rispetto ai meri dati tradizionali, grazie, appunto, alla loro dimensione e diversità. In questo ambito la datificazione assume un ruolo fondamentale, in quanto rappresenta il processo di trasformazione delle informazioni grezze in dati strutturati che possono essere in seguito utilizzati per l'analisi.

### 1.2 L'importanza della datificazione nel contrasto all'hate speech.

Con l'espressione *hate speech* si intende «l'istigazione, la promozione o l'incitamento alla denigrazione, all'odio o alla diffamazione nei confronti di una persona o di un gruppo di persone, o il fatto di sottoporre a soprusi, molestie, insulti, stereotipi negativi, stigmatizzazione o minacce tale persona o gruppo, e comprende la giustificazione di queste varie forme di espressione, fondata su una serie di motivi, quali la "razza", il colore, la lingua, la religione o le convinzioni, la nazionalità o l'origine nazionale o etnica, nonché l'ascendenza, l'età, la disabilità, il sesso, l'identità di genere, l'orientamento sessuale e ogni altra caratteristica o situazione personale»[20]. Questa è la definizione ufficiale data dalla Commissione parlamentare Jo Cox (istituita in Italia nel 2016 dalla Camera dei deputati e presieduta dalla Presidente della Camera Laura Boldrini). La relazione stipulata ha dimostrato l'esistenza della cosiddetta "Piramide dell'Odio", al cui vertice si trovano i comportamenti considerati "più gravi" e alla base si trovano quelli valutati "meno gravi". Si suddivide in quattro categorie:

1. crimini d'odio: atti di violenza fisica, fino all'omicidio, perpetrati contro persone in base a qualche caratteristica come il sesso, l'orientamento sessuale, l'etnia, il colore della pelle, la religione o altro.
2. linguaggio d'odio: minacce e/o incitamento alla denigrazione e alla violenza contro una persona o gruppi di persone identificate in base ad una qualche

Disponibile al seguente link: < https://www.emerald.com/insight/content/doi/10.1108/JFRC-06-2017-0054/full/pdf?title=a-review-of-credit-scoring-research-in-the-age-of-big-data>

[20]Camera dei deputati, La piramide dell'odio in Italia, Commissione "Jo Cox" su fenomeni di odio, intolleranza, xenofobia e razzismo, luglio 20216.
Disponibile al link:
<https://www.camera.it/application/xmanager/projects/leg17/attachments/shadow_primapagina/file_pdfs/000/007/099/Jo_Cox_Piramide_odio.pdf>

caratteristica come il sesso, l'orientamento sessuale, l'etnia, il colore della pelle, la religione o altro.

3. discriminazioni: comportamenti negativi riguardanti il lavoro, l'alloggio, la scuola, le relazioni sociali
4. stereotipi: classificazione in base a false rappresenta, rappresentazioni false o fuorvianti, insulti, linguaggio ostile "normalizzato" o banalizzato.

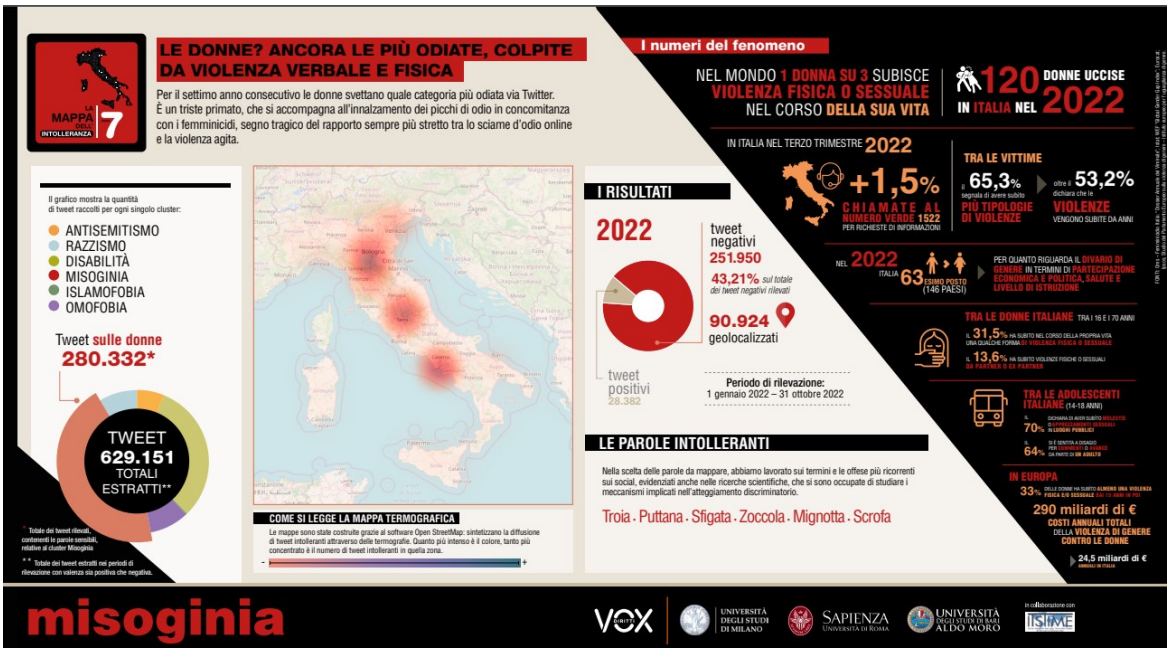


*Figura 4: indagine Vox 2022 rileva come le donne siano le più colpite da violenza verbale sui social e non solo.*[21]

Come può essere definito l'odio? Per rispondere a tale quesito bisogna ricorrere a una delle prime definizioni ideate da Gordon Allport[22], noto psicologo statunitense che nel 1954 elaborò la *Allport's Scale of Predjudice and Discrimination,* ossia una scala dove vengono elencati cinque comportamenti dettati dai pregiudizi[23]:

[21] VOX (Osservatorio italiano sui diritti), Mappa dell'intolleranza 7.0, 2022
Disponibile al link: <https://www.retecontrolodio.org/cmswp/wp-content/uploads/2023/01/Mappa-dellIntolleranza-7.pdf>
[22] Gordon Allport (Montezuma, 1897 – 1967)
[23] G.W. Allport, The nature of Prejudice, 1954, pag.49

1. Anticolution: è il primo livello, descrive come un gruppo di maggioranza stereotipizzi in maniera negativa un gruppo di minoranza (cd. "outgroup") tramite espressioni offensive o basate su pregiudizi.
2. Avoidance: è il secondo livello, descrive il modo in cui le persone del gruppo di maggioranza evitano volontariamente i soggetti appartenenti a un gruppo di minoranza, per cui il danno arrecato si concretizza nell'isolamento delle persone e nella loro conseguente esclusione[24].
3. Discrimination: è il terzo livello, riguarda il comportamento di coloro che vogliono danneggiare un gruppo di persone impedendogli di raggiungere i propri obbiettivi, ottenere l'accesso a un adeguato percorso scolastico o lavorativo. [25]
4. Physical attack: è il quarto livello, conosciuto anche con il termine *hate crime* e si riferisce a quei comportamenti violenti motivati dall'odio.
5. Extermination: è il quinto livello, prevede l'eliminazione di un gruppo attraverso il genocidio o la pulizia etnica.

I comportamenti descritti dal primo al terzo punto, sono classificati da Allport come anticipatore alla violenza, a differenza dei punti quarto e quinto punto che descrivono comportamenti di violenza vera e propria. Nel 2017 l'associazione *SOS Racisme* e *l'Institut de Detrs Humans de Cataluya*, con il finanziamento del Municipio di Barcellona, compiendo un'operazione simile a quella di Allport, ha tentato di classificare alcuni comportamenti utili per descrivere i discorsi d'odio. È stato formulato *l'Iceberg dell'odio*, suddividendo le condotte in visibili e invisibili. [26]

A partire dalla base della piramide, ordinato in modo crescente, troviamo:

1. gli stereotipi;
2. i pregiudizi;
3. la discriminazione;
4. la violenza fisica e verbale;

[24] G. Ziccardi, L'odio online, Milano: Raffaello Cortina, 2016, cit. pag. 21.
[25] Peter Watson, Psychology and race, New Brunswick, NJ: Aldine Transaction, 2007, p.46
[26] Edoardo Bazzaco, Ana García Juanatey, Jon Lejardi, Anna Palacios y Laia Tarragona, ¿Es odio? Manual práctico para reconocer y actuar frente a discursos y delitos de odio, pubblicato a Barcellona, novembre 2017, pag. 7-42.
Disponibile al link:
< https://www.idhc.org/arxius/recerca/1517393506-ES_ODIO__Manual_practico_vF.pdf>

5. la violenza fisica rivolta alla persona.

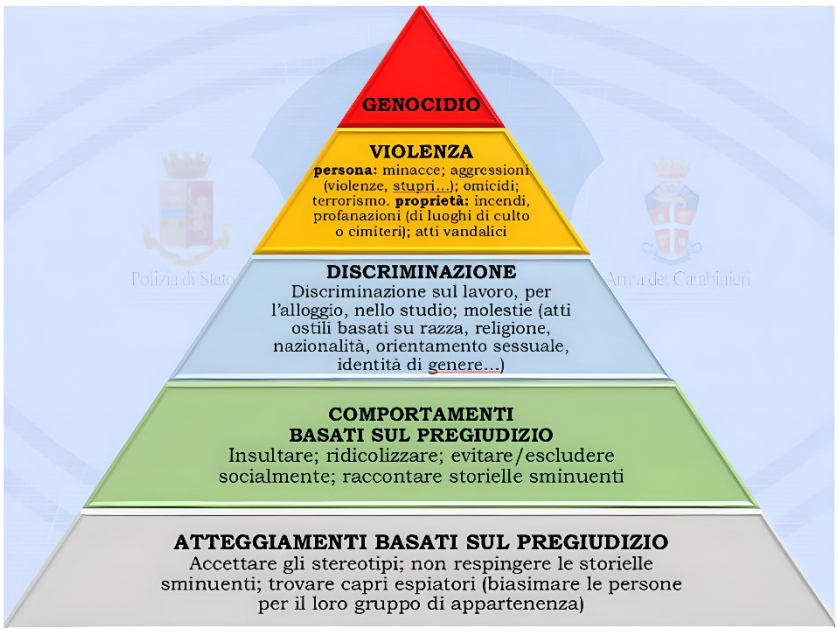


*Figura 5: Piramide dell'odio*[27]

L'*hate speech* online è materia di studio negli Stati Uniti già a partire dal 1999. Lì vennero pubblicati degli studi in cui si evidenziava l'uso di Internet, nello specifico nei primi forum, come mezzo per promuovere i discorsi d'odio e più in generale per incitare alla discriminazione e alla violenza.[28] Nel corso degli ultimi anni si è verificato un incremento significativo di questo fenomeno. L'odio online presenta tre peculiarità:

1. la permanenza nel tempo: il discorso d'odio può restare online per molto tempo, in differenti formati e su più piattaforme. Più a lungo resta accessibile, maggiore è il suo potenziale in termine di danni.
2. Itinerante e ricorrente: i contenuti d'odio si possono propagare all'infinito tra le diverse piattaforme, grazie alla loro peculiare dinamica di diffusione delle

[27] Oscad (Osservatorio per la Sicurezza Contro gli Atti Discriminatori), Commissione straordinaria per il contrasto dei fenomeni di intolleranza, razzismo, antisemitismo e istigazione all'odio e alla violenza, Audizione del Prefetto Vittorio Rizzi, 22 giugno 2021, pag.1
Disponibile al seguente link:
<https://www.senato.it/application/xmanager/projects/leg18/attachments/documento_evento_procedura_commissione/files/000/382/401/Osservatorio_per_la_sicurezza_contro_gli_atti_discriminatori_OSCAD_.pdf>

[28] L'hate speech e la violenza verbale.
Disponibile al seguente link:
<https://www.dirittodellinformatica.it/ict/web/lhate-speech-e-la-violenza-verbale-online.html/>

informazioni. Un contenuto rimosso, infatti, può apparire sotto un altro nome o un altro titolo sulla stessa piattaforma o altrove.

3. Nascosti dietro a un monitor: si presenta un'idea di anonimato e di impunibilità associato all'utilizzo della rete internet.[29]

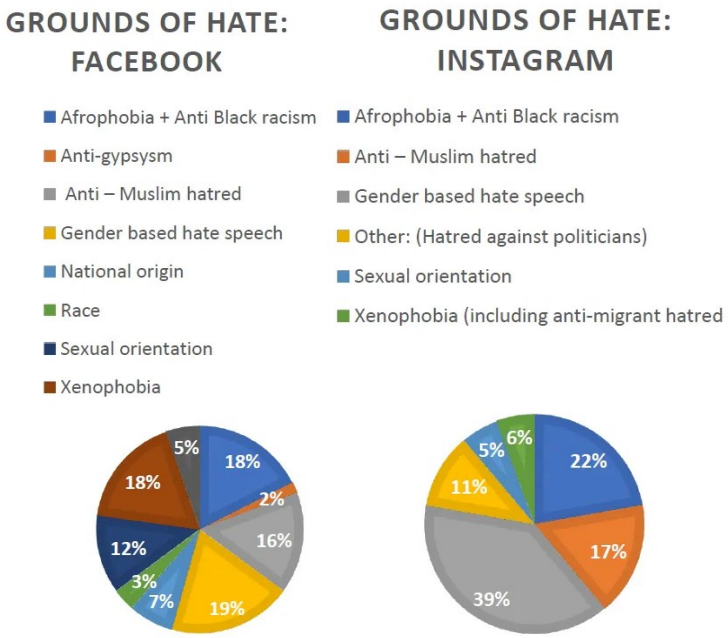


*Figura 6: tipologia d'odio segnalato su diverse piattaforme social*[30]

La datificazione, se usata dalle diverse piattaforme social, potrebbe aiutare e influenzare la gestione dei discorsi d'odio nel web, in quanto hanno la possibilità di raccogliere e analizzare una vastissima quantità di dati dei loro utenti, compresi i contenuti condivisi, i commenti e le interazioni. Grazie a questo strumento di analisi massiva di dati, si è assistito un notevole miglioramento da parte degli algoritmi nell'identificare i contenuti d'odio. La datificazione, inoltre, può aiutare a comprendere meglio il contesto in cui si verificano i discorsi d'odio, oltre che a individuare i responsabili di tali azioni, e di conseguenza rappresenta uno strumento utile per riuscire a contrastarli, grazie agli algoritmi e alle segnalazioni degli utenti, si

[29] Amnesty International Sezione Italiana, Hate Speech conoscerlo e contrastarlo guida breve per combattere i discorsi d'odio online, 2019, pag.19
Disponibile al link: <https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2019/05/13104653/HATE-SPEECH-CONOSCERLO-E-CONTRASTARLO_web-version.pdf>
[30] CESIE, sCAN: Sfide e risultati del 4°esercizio di monitoraggio in base al Codice di Condotta sottoscritto da Commissione Europea e piattaforme, IT.
Disponibile al seguente link:
<https://cesie.org/studi/scan-monitoring-exercise/>

riesce in maniera più veloce e automatica a eliminare i post che incitano all'odio. Nonostante questi vantaggi, però, la datificazione cela anche delle difficoltà, che riguardano il contesto e l'interpretazione. I discorsi d'odio possono essere scritti in una forma tale che l'algoritmo non è ancora in grado di comprendere: l'*hate speech* può evolversi in maniera molto veloce attraverso nuove parole chiave o nuove metafore che vengono introdotte. Queste analisi richiedono notevoli infrastrutture tecnologiche, oltre che ingenti risorse finanziarie e competenze tecniche avanzate. Nel 2022 le piattaforme social sono riuscite, grazie ai feedback degli utenti, a migliorare gli algoritmi per la ricerca di parole illecite. Secondo lo studio condotto dalla Commissione Europea, denominato *7th evaluation of the Code of Conduct*, nel 64,4% dei casi le aziende IT hanno valutato le segnalazioni dei post da parte degli utenti in meno di 24 ore, rimuovendo il 63,6% dei contenuti. Questi risultati però sono più bassi rispetto al 2020 quando si era registrata una percentuale di rimozione pari al 70%[31].

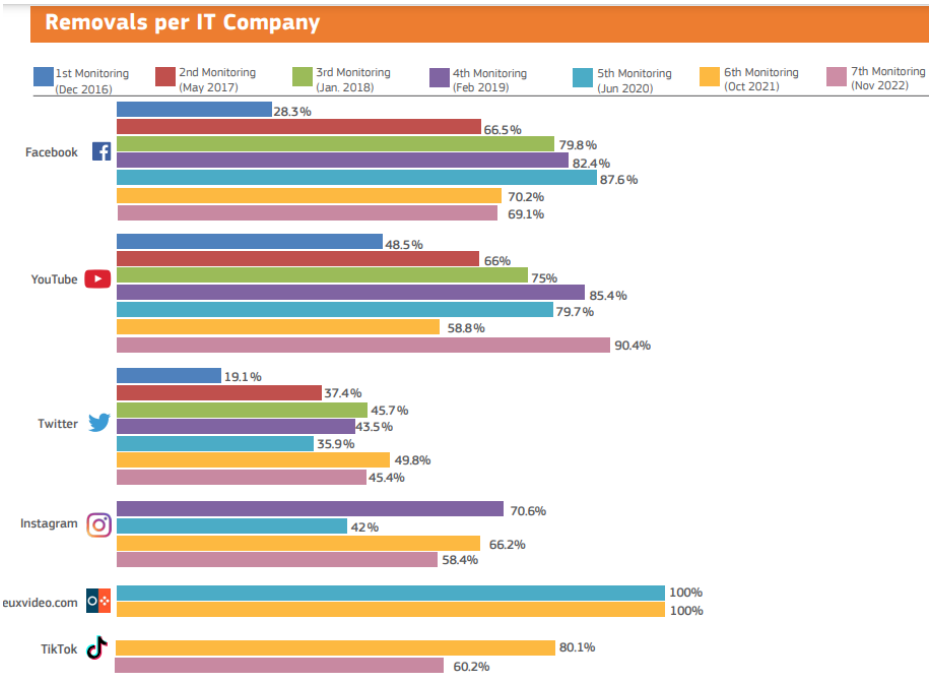


*Figura 7: tasso di rimozione dei contenuti dal 2016 al 2022*[32]

[31] European Commission,7th evaluation of the Code of Conduct, novembre 2022, pag.1 Disponibile al link: <https://commission.europa.eu/system/files/2022-12/Factsheet%20-%207th%20monitoring%20round%20of%20the%20Code%20of%20Conduct.pdf>

[32] Vedi sopra, pag.3

### 1.3. Strumenti e tecniche di analisi dei dati per il contrasto all'hate speech

*"Il connubio di odio e di tecnologia è il massimo pericolo che sovrasti l'umanità. E non mi riferisco alla sola grande tecnologia della bomba atomica, mi riferisco anche alla piccola tecnologia della vita di ogni giorno: conosco persone che stanno per ore davanti al televisore perché hanno disimparato a comunicare tra di loro".*

*S. Wiesenthal*

La prima domanda che ci si pone è come mai la popolazione abbia "disimparato a comunicare", soprattutto nel mondo online. A questo quesito vi sono molteplici risposte e una delle tante, forse la più importante, è l'immediatezza della comunicazione, dove per cui non si guarda più alla forma o al contenuto del messaggio, a causa anche della presenza dei cd. "troll". Nel gergo delle comunità virtuali vengono definiti come «Chi pubblica un messaggio deliberatamente provocatorio in un gruppo di discussione o in una bacheca con l'intenzione di causare il massimo disturbo e discussione»[33]. I *troll* sfruttano l'*hate speech* come mezzo principale per raggiungere i loro obiettivi provocatori, utilizzando commenti offensivi o linguaggio discriminatorio per alimentare rabbia e indignazione tra le persone che interagiscono con loro. È importante notare che non tutti i *troll* promuovono l'*hate speech*, ma vi è una significativa sovrapposizione tra i due fenomeni poiché entrambi si basano su comportamenti negativi online e possono contribuire a favorire un ambiente digitale non sano. Per contrastare entrambi i fenomeni, gli Internet Service Provider (ISP, ossia fornitori di servizi internet) adottano diverse tecniche. La più importante il *Text mining*[34] e utilizza il Natural Language Processing (NLP, ossia linguaggio naturale) per estrarre informazioni significative e analizzare il contenuto dei messaggi. L'NLP viene utilizzato per classificare i testi oltre che a rilevare le espressioni offensive.[35] Questo processo si struttura in tre fasi: l'indicizzazione, il

[33] Tradotto dall'inglese: «One who posts a deliberately provocative message to a newsgroup or message board with the intention of causing maximum disruption and argument».
Disponibile al link:
<https://www.urbandictionary.com/define.php?term=troll>

[34] Definizione: « Text Mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources».
Vishal Gupta, A Survey of Text Mining Techniques and Applications, Lecturer Computer Science & Engineering, University Institute of Engineering & Technology, Journal of emerging technologies in web intelligence, VOL. 1, NO. 1, AUGUST 2009

[35] Cineca, Text mining.
Vedasi informazioni al seguente link:

*mining* e la valutazione. Nella prima fase si effettua l'analisi linguistica, nella seconda si utilizzano algoritmi di *Data Mining* per raggiungere i vari obbiettivi, ad esempio un algoritmo di *Machine Learning* può classificare o individuare nuove parole potenzialmente offensive. Nell'ultima fase si valutano e interpretano i risultati ottenuti. La seconda metodologia utilizzata per contrastare i *troll* e l'*hate speech* è il *Machine Learning*: tramite l'addestramento di questi algoritmi di apprendimento automatici possono essere utilizzati per il riconoscimento dell'*hate speech*. Questi modelli si possono utilizzare per eseguire un'attenta analisi automatizzata su una grande quantità di testi in modo da identificare i vari contenuti discriminatori.[36]

Altri metodi per contrastare i fenomeni d'odio in rete sono la creazione di filtri per i contenuti, come nel caso di Instagram che ha introdotto un nuovo filtro per rilevare e bloccare in maniera automatica i messaggi contenenti un linguaggio offensivo. Questi strumenti sono in grado di analizzare testi, immagini e video per identificare i contenuti discriminatori. Si può, inoltre, ricorrere alla classica moderazione umana: le piattaforme possono avvalersi di personale per monitorare ed eventualmente cancellare i contenuti. Questa tecnica è di difficile applicazione in quanto è impensabile l'analisi di milioni di testi manualmente. Le istituzioni sia nazionali che europee hanno attribuito con fermezza la responsabilità alle piattaforme della diffusione di contenuti illeciti. Nel 2016 l'Unione Europea ha cercato di coinvolgere, in maniera attiva, le maggiori aziende informatiche attraverso la stipulazione di un Codice di Condotta contro l'incitamento all'odio online. Venne firmato all'inizio da Facebook, Microsoft, YouTube e Twitter e successivamente nel 2018 venne sottoscritto da Instagram e da Snapchat, per poi essere siglato da TikTok nel settembre del 2020 e Linked 2021[37]. Grazie al Codice di Condotta anche queste

<https://www.cineca.it/text-mining>

[36] Guansong Pang, Jitendra Singh Malik, Anton van den Hengel, Deep Learning for Hate Speech Detection: A Large-scale Empirical Evaluation, Singapore Management University, Research Collection School Of Computing and Information Systems, 2022, pagg. 1-4
Disponibile al seguente link:
< https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=8018&context=sis_research>

[37] Codice di condotta per lottare contro le forme illegali di incitamento all'odio online, 30 giugno 2016, Commissione Europea, ottrobre 2019.
Disponibile al link
<https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en>

multinazionali condividono con agli Stati dell'UE l'impegno di contrastare le forme illegali di incitamento all'odio.

Ulteriori strumenti di contrasto all'*hate speech* sono le campagne di sensibilizzazione per contribuire a sviluppare i principi di uguaglianza e il rispetto della dignità umana, oltre al principio di non discriminazione[38], come nel caso del progetto del Governo italiano *Lascia l'odio senza parole* che si colloca all'interno del progetto della Commissione Europea *Innovative Monitoring Systems and Prevention Policies of Online Hate Speech*[39].

Il progetto IMSyPP si propone lo «sviluppo di modelli e tecniche per la *detection* automatizzata dell'*hate speech* in diverse lingue, al fine di identificare i fattori determinanti, le raccomandazioni più efficaci sul piano delle narrazioni, e le più opportune proposte di policy in un'ottica europea»[40].

[38] Art. 3 Cost*.* «Tutti i cittadini hanno pari dignità sociale e sono eguali davanti alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali.
È compito della Repubblica rimuovere gli ostacoli di ordine economico e sociale, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese».
[39] Governo italiano Presidenza del Consiglio dei ministri, Campagna di comunicazione "Lascia l'odio senza parole, maggio 2023.
Disponibile al link:
<https://www.governo.it/it/media/campagna-di-comunicazione-lascia-l-odio-senza-parole/22520>
Video del progetto:
https://www.youtube.com/watch?v=72RsJvGYCnI&t=27s
[40] AGCOM, Progetto IMSyPP "Innovative Monitoring Systems and Prevention Policies of Online Hate Speech", 2020.
Disponibile al link:
<https://www.agcom.it/956>

## Capitolo 2

## Le sfide etiche della datificazione nel contrasto all'hate speech

### 2.1 Normativa in merito ai discorsi d'odio

*Tutti i cittadini hanno pari dignità sociale e sono eguali davanti alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali.*[41]

Il Comitato dei Ministri del Consiglio d'Europa, con la Raccomandazione n.5 del 2010, ha invitato tutti gli Stati membri di «adottare misure adeguate per combattere qualsiasi forma di espressione, in particolare nei mass media e su internet, che possa essere ragionevolmente compresa come elemento suscettibile di fomentare, propagandare o promuovere l'odio o altre forme di discriminazione nei confronti delle persone lesbiche, gay, bisessuali o transessuali»[42], preoccupandosi però che tali misure rispettino «il diritto fondamentale alla libertà di espressione, conformemente all'art. 10 della Convenzione europea[43] dei diritti dell'uomo e alla giurisprudenza della Corte»[44]. A livello nazionale non esiste ancora una regolamentazione che contrasti direttamente il fenomeno dell'*hate speech:* il vuoto normativo viene riempito dall'applicazione di norme relative ai discorsi d'odio che fanno riferimento alla Legge Mancino del 1993[45], la quale mira a contrastare l'incitamento all'odio, le discriminazioni o violenze per motivi razziali, etnici, nazionali o religiosi. Inoltre, la Legge prevede un'aggravante per i reati commessi con motivazioni razziste, xenofobe o legate all'intolleranza religiosa. Ulteriore norma riguarda la Dichiarazione dei Diritti in Internet, emanata nel 2015 dalla Camera dei deputati

[41] Art. 3 Costituzione
[42] Consiglio d'Europa, Raccomandazione CM/2010(5), parte I, lett. B, para. 6
Disponibile al link: <https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016804c6add>
[43] Art. 10 CEDU, «Ogni persona ha diritto alla libertà d'espressione. Tale diritto include la libertà d'opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera. Il presente articolo non impedisce agli Stati di sottoporre a un regime di autorizzazione le imprese di radiodiffusione, cinematografiche o televisive.
2. L'esercizio di queste libertà, poiché comporta doveri e responsabilità, può essere sottoposto alle formalità, condizioni, restrizioni o sanzioni che sono previste dalla legge e che costituiscono misure 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, per impedire la divulgazione di informazioni riservate o per garantire l'autorità e l'imparzialità del potere giudiziario».
[44] Vedi nota 43
[45]DECRETO-LEGGE 26 aprile 1993, n. 122, convertito poi in legge il 25 giugno 1993 n.205
Disponibile al link: <https://www.gazzettaufficiale.it/eli/id/1993/06/26/093A3644/sg>

presieduta dalla Presidente Laura Boldrini. In questo documento, in particolare negli artt. 2 (Diritto di accesso) e 3 (Diritto alla conoscenza e all'educazione in rete) viene sottolineata «l'importanza dell'accesso a Internet come diritto fondamentale della persona e condizione per il pieno sviluppo individuale oltre ad avere il diritto ad essere posta in condizione di acquisire e di aggiornare le capacità necessarie ad utilizzare Internet in modo consapevole per l'esercizio dei propri diritti e delle proprie libertà fondamentali.»[46] Questo testo, inoltre, garantisce la tutela dei dati personali all'art. 5, il diritto all'identità all'art. 9, oltre la protezione dell'anonimato della persona all'art. 10.

Più recente è il *Regolamento recante disposizioni in materia di rispetto della dignità umana e del principio di non discriminazione e di contrasto all'Hate Speech* del 15 maggio 2019, emanato dall'AGCOM (Autorità per le Garanzie nelle Comunicazione), nel quale vengono stabilite delle nuove normative di comportamento per i fornitori di servizi radiofonici, audiovisivi e online. Particolare attenzione va rivolta all'art. 5 comma 2 dove viene sancito: «I fornitori di servizi di media audiovisivi e radiofonici, anche privati, sono invitati a promuovere iniziative aventi ad oggetto i temi dell'inclusione e della coesione sociale, della promozione delle diversità e dei diritti fondamentali della persona[47].». Tale Regolamento mira a impedire i comportamenti che incitano all'odio basato sull'etnia, religione, sesso, etc. Difatti, l'obbiettivo principale, oltre che a contrastare tutti i fenomeni di discriminazione, è quello di impedire che vengano minati i principi fondamentali della tutela della persona e della dignità umana.

A seguito dell'aumento sproporzionato degli atti di bullismo e cyberbullismo [48], è stata introdotta e adottata una normativa volta a tutelare i minori, inserendo degli strumenti

[46]Camera dei deputati, Dichiarazione dei diritti in Internet, 2015.
Disponibile al link:
<https://www.camera.it/application/xmanager/projects/leg17/commissione_internet/dichiarazione_dei_diritti_internet_pubblicata.pdf>

[47]AGCOM, Delibera n.159/19/CONS, Regolamento recante disposizioni in materia di rispetto della dignità umana e del principio di non discriminazione e di contrasto all'hate speech.
Disponibile al link: <https://www.agcom.it/documents/10179/13511391/Allegato+23-5-2019+1558628852738/5908b34f-8c29-463c-a7b5-7912869ab367?version=1.0>

[48] Secondo i dati della Sorveglianza Health Behaviour in School-aged Children - HBSC Italia 2022 (Indagine HBSC 2022 presentata all'Istituto Superiore di Sanità) affermano che gli undicenni vittime di bullismo sono il 18,9 % dei ragazzi e il 19,8% delle ragazze; nella fascia di età di 13 anni sono il 14,6% dei maschi e il 17,3% delle femmine; gli adolescenti (15 anni) sono il 9,9% dei ragazzi e il 9,2% delle ragazze. Per quanto riguarda il cyberbullismo la fascia di età 11 anni risultano vittime il 17.2% dei maschi e il 21,1% delle femmine; i tredicenni coinvolti sono il 12,9% dei ragazzi e il 18,4% delle ragazze; gli adolescenti di 15 anni sono il 9,2% dei maschi e l'11,4% delle femmine.

utili per la rimozione dalla Rete di contenuti lesivi (Legge n. 71 del 29 maggio del 2019) [49].

L'Italia, insieme a 177 Parti, ha aderito alla International Convention on the Elimination of All Forms of Racial Discrimination (ICERD, ossia la Convenzione Internazionale sull'eliminazione di ogni forma di discriminazione razziale) stipulata il 21 dicembre 1965 a New York ed entrata in vigore il 4 gennaio 1969. Tale Convenzione obbliga gli Stati che l'hanno sottoscritta a perseguire, mediante qualsiasi mezzo, una politica tendente all'eliminazione di ogni tipo di discriminazione, favorendo l'intesa tra tutte le razze e punendo ogni comportamento lesivo per la dignità del soggetto. Trova applicazione all'art.4 punto a. «gli Stati Parte considereranno reato punibile per legge ogni diffusione di idee basate sulla superiorità o sull'odio razziale, ogni incitamento alla discriminazione razziale, nonché ogni atto di violenza o incitamento a tali atti, rivolti contro qualsiasi razza o gruppo di individui di diverso colore o origine etnica, così come ogni assistenza ad attività razzistiche, compreso il loro finanziamento[50].»

Il Codice penale italiano punisce diversi reati legati all'*hate speech*, come l'istigazione alla discriminazione all' art. 604-bis (Propaganda e istigazione a delinquere per motivi di discriminazione razziale etnica e religiosa)[51]. La norma in oggetto tutela in maniera diretta la dignità della persona e il principio di uguaglianza, sanzionando l'apologia del fascismo e del nazismo, l'incitamento alla violenza e all'odio razziale o religioso, come nel recente caso che vede coinvolto il partito politico Forza Nuova e il colosso Facebook, il quale ne ha oscurato il profilo social, perché diffondevano contenuti contrari alle linee guida, oltre che a violare le norme nazionali e sovranazionali. Tutto ciò ha portato il Giudice a ribadire

---

Disponibile al link: <https://www.epicentro.iss.it/hbsc/indagine-2022-nazionali-convegno-8-febbraio-2023>

[49]Legge 29 maggio 2017, n. 71.
Vedasi informazioni al link:
<https://www.gazzettaufficiale.it/eli/id/2017/06/3/17G00085/sg>

[50]Convenzione Internazionale sull'eliminazione di ogni forma di discriminazione razziale, New York 1965.
Disponibile al link:
<https://fedlex.data.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/1995/1164_1164_1164/20220421/it/pdf-a/fedlex-data-admin-ch-eli-cc-1995-1164_1164_1164-20220421-it-pdf-a.pdf>

[51]Articolo 604-bis completo al seguente link:
<https://www.gazzettaufficiale.it/atto/serie_generale/caricaArticolo?art.versione=1&art.idGruppo=60&art.flagTipoArticolo=1&art.codiceRedazionale=030U1398&art.idArticolo=604&art.idSottoArticolo=2&art.idSottoArticolo1=10&art.dataPubblicazioneGazzetta=1930-10-26&art.progressivo=0#:~:text=E'%20vietata%20ogni%20organizzazione%2C%20associazione,%2C%20etnici%2C%20nazionali%20o%20religiosi.>

come nel caso di specie vi fosse stata una violazione delle regole della community, ma soprattutto delle norme imperative poste proprio a contrasto del fenomeno dell'*hate speech*.[52] Si cita anche l'art. 604-ter cp. (Circostanza aggravante) il quale prevede per l'appunto una circostanza aggravante generica, applicabile a qualsiasi reato commesso con finalità di discriminazione razziale, religiosa o etnica. Il codice di procedura penale all'art. 90quater (Condizione di particolare vulnerabilità) valuta se il reato è stato commesso con violenza alla persona o con odio razziale.

Al livello europeo non si può che cominciare dalla Universal Declaration of Human Rights (UDHR, ossia la Dichiarazione dei Diritti Umani) del 1948, approvata dall'Assemblea Generale delle Nazioni Unite. Innanzitutto, la Dichiarazione sancisce all'art. 2 il principio di non-discriminazione e di uguaglianza ed evidenzia i cosiddetti *prohibited grounds of discrimination* (motivi di discriminazione vietati)*:* «Ad ogni individuo spettano tutti i diritti e tutte le libertà enunciate nella presente Dichiarazione, senza distinzione alcuna, per ragioni dirazza, di colore, di sesso, di lingua, di religione, di opinione politica o di altro genere, di origine nazionale o sociale, di ricchezza, di nascita o di altra condizione»[53]. La Dichiarazione pone dei vincoli con gli artt. 7[54] e 29[55] per quanto riguarda la libertà d'espressione (art. 19) in quanto un esercizio illimitato potrebbe portare a degli abusi. L'*hate speech*, infatti, può costituire un possibile caso di abuso della libertà d'espressione.

---

[52] Tribunale Roma, N. R.G. 64894/2019, Sezione diritti della persona e immigrazione civile,23/02/2020. Sentenza disponibile al seguente link:
<https://www.asgi.it/wp-content/uploads/2020/02/Ordinanza-RG-648942019-Forza-Nuova-art700.pdf>

[53] Assemblea Generale delle Nazioni Unite, Dichiarazione Universale dei Diritti Umani, dicembre 1948. Disponibile al link:
<https://www.ohchr.org/sites/default/files/UDHR/Documents/UDHR_Translations/itn.pdf>

[54] Art. 7 UDHR.
*Tutti sono uguali davanti alla legge e hanno diritto, senza alcuna discriminazione, a un'uguale tutela da parte della legge. Tutti hanno diritto a un'eguale protezione contro qualsiasi discriminazione in violazione della presente Dichiarazione e contro qualsiasi incitamento a tale discriminazione.*

[55] Art. 29 UDHR.
*Ognuno ha dei doveri verso la comunità nella quale solo è possibile il libero e pieno sviluppo della sua personalità.*
*Nell'esercizio dei suoi diritti e delle sue libertà, ognuno sarà soggetto solo alle limitazioni determinate dalla legge al solo scopo di garantire il dovuto riconoscimento e il rispetto dei diritti e delle libertà altrui e di soddisfare le giuste esigenze della moralità, dell'ordine pubblico e del benessere generale in una società democratica.*
*Tali diritti e libertà non possono in nessun caso essere esercitati contrariamente agli scopi e ai principi delle Nazioni Unite.*

La libertà di esprimere il proprio pensiero e di diffonderlo rappresenta uno dei pilastri della società democratica. Tale libertà, infatti, viene riconosciuta come un diritto fondamentale e «coessenziale al regime di libertà garantito dalla Costituzione»[56], «pietra angolare dell'ordinamento democratico»[57], nonché «cardine di democrazia nell'ordinamento più generale»[58]. Queste definizioni vanno attuate sempre, purché non si vada a ledere la libertà personale degli altri individui. Come sancisce l'art. 30 della UDHR sul divieto dell'abuso dei diritti: «Nulla nella presente Dichiarazione può essere interpretato nel senso di implicare un diritto di un qualsiasi Stato, gruppo o persona di esercitare un'attività o di compiere un atto mirante alla distruzione di alcuno dei diritti e delle libertà in essa enunciati.»[59]

Sempre a livello sovranazionale, la Commissione per i Diritti Umani era un organo che ha cessato di esistere nel 2006 ed è stato sostituito dal Consiglio per i Diritti Umani delle Nazioni Unite[60]. La Commissione per i Diritti Umani era incarica di definire degli standard per il rispetto della Dichiarazione Universale e allo stesso tempo contribuire alla stesura e poi all'emanazione del Patto Internazionale sui Diritti Civili e Politici (ICPRR) del 1996. Nella disciplina della libertà di espressione, l'ICCPR sancisce le condizioni per limitarla:

1. la riserva di legge;
2. l'esistenza di uno scopo di legittimità «rispetto dei diritti o della reputazione, salvaguardia della sicurezza nazionale, dell'ordine pubblico, della salute o la morale pubblica»[61];

---

[56] Rif. Sentenza n. 11 del 1968.
I Diritti Fondamentali nella Giurisprudenza della Corte costituzionale, pag.5.
Disponibile al link:
<https://www.cortecostituzionale.it/documenti/convegni_seminari/STU185_principi.pdf>
[57] Rif. Sentenza n. 84 del 17 aprile 1969, pag.1378
Disponibile al link:
<https://www.jstor.org/stable/pdf/23159431.pdf?refreqid=excelsior%3A2671fc5c465a5e459e118ada8e215072&ab_segments=&origin=&initiator=&acceptTC=1>
[58] Vedi sopra
[59] Art. 30 della Dichiarazione dei Diritti Umani.
Disponibile al link:
<https://www.ohchr.org/sites/default/files/UDHR/Documents/UDHR_Translations/itn.pdf>
[60] Non va confusa con il Comitato per i diritti umani (cd. CCPR) che invece è un organo formato da esperti creato dall'ICCPR.
[61] United Nations, Promotion and protection of the right to freedom of opinion and expression, del 2019 pag.5
Disponibile al link:<https://digitallibrary.un.org/record/3833657>

3. la restrizione deve rispettare i requisiti di proporzionalità e necessità.[62] In particolare, nei casi in cui si voglia difendere la reputazione di un individuo, il diritto internazionale riguardante i diritti umani è maggiormente propenso a preferire una sanzione amministrativa o civile rispetto a una sanzione penale.[63]

I discorsi basati sull'odio sono vietati, inoltre, dalla Convenzione Europea dei Diritti dell'Uomo (CEDU). Firmata nel 1950 dal Consiglio d'Europa, essa mira a salvaguardare le persone dalla violazione dei diritti umani. Innanzitutto, bisogna tener presente l'articolo 14 che sancisce il divieto di discriminazione «Il godimento dei diritti e delle libertà riconosciuti nella presente Convenzione deve essere assicurato senza nessuna discriminazione, in particolare quelle fondate sul sesso, la razza, il colore, la lingua, la religione, le opinioni politiche o quelle di altro genere, l'origine nazionale o sociale, l'appartenenza a una minoranza nazionale, la ricchezza, la nascita od ogni altra condizione»[64]. Come soprammenzionato, anche nell'ambito della protezione dei diritti umani, vi è un esplicito divieto di discriminazione sulla base dei cosiddetti *prohibited grounds* (motivi vietati) che includono la razza, il sesso, inclusa la gravidanza, l'identità di genere, la lingua, l'etnia, il credo religioso, età, l'orientamento sessuale etc.

Viste le nuove modalità di diffusione dei messaggi d'odio, nel 2001 il Consiglio d'Europa ha redatto la Convenzione di Budapest[65]: una guida per elaborare una legislazione atta a combattere la criminalità informatica, il cui obbiettivo principale è perseguire i reati commessi online, come la violazione del diritto d'autore, le frodi informatiche, i crimini d'odio e tutte le violazioni della sicurezza online, oltre alla pornografia infantile. Successivamente, il 28 gennaio 2003 venne firmato dal Consiglio Europeo il Protocollo addizionale alla Convenzione di Budapest sulla criminalità informatica, riguardante la criminalizzazione degli atti di razzismo e xenofobia commessi a mezzo di sistemi informatici[66], il quale entrò in vigore a livello internazionale il 1° marzo 2006. Tale

[62] Vedi nota sopra
[63] Alessandro Di Rosa, Hate speech e discriminazione, Mucchi Editore, maggio 2020, pag. 28.
[64] Convenzione Europea dei Diritti dell'Uomo.
Disponibile al link: <https://www.echr.coe.int/documents/d/echr/convention_ita>
[65] Consiglio d'Europa, Convention on Cybercrime, European Treatry Series – No. 185, Budapest, novembre 2001.
Disponibile al link: <https://rm.coe.int/1680081561>
[66] STE no.189.
Disponibile al link: <https://www.coe.int/it/web/conventions/full-list2?module=treaty-detail&treatynum=189>

Protocollo ha la finalità di ampliare la portata della Convenzione di Budapest e di estendere i reati legati alla propaganda a sfondo xenofobo e razzista, oltre che a prevedere delle sanzioni penali qualora si verificassero tali illeciti.[67] Si cita anche la Carta dei diritti fondamentali dell'Unione europea, la quale vieta all'art. 21 qualsiasi forma di discriminazione basata sul sesso, razza, colore, etnia origini, disabilità etc., includendo anche l'art. 23 sull'uguaglianza tra uomo e donna in qualsiasi area, come il lavoro, la paga o l'occupazione[68]. Sfortunatamente, all'interno delle numerose Regolamentazioni vengono menzionate solo alcune delle categorie a potenziale rischio di *hate speech*, per questo motivo il Parlamento Europeo, su comunicazione della Commissione Europea[69] a seguito degli studi condotti dall' European Commission against Racism and Intolerance (ECRI), ha chiesto di revisionare la decisione-quadro 2008/913/GAI[70] allo scopo di introdurre nella classifica delle espressioni contenenti messaggi d'odio anche quelle che presentano contenuti transfobici/omofobi.

Da menzionare è il recentissimo Digital Service Act presentato dalla Commissione Europea, sia al Parlamento Europeo che al Consiglio dell'Unione Europea, il 15 dicembre 2020 e approvato il 19 ottobre 2022[71]. Lo scopo principale è quello di creare uno spazio digitale sicuro per tutti gli utenti e di regolamentare le attività delle varie piattaforme online, oltre che a creare un sistema per la segnalazione e rimozione dei contenuti illeciti da parte di qualsiasi utente[72].

---

[67] Vedi sopra
[68] Carta dei Diritti Fondamentali dell'Unione Europea.
Testo consultabile al seguente link:
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT>
[69] Un'Europa più inclusiva e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio, Bruxelles 9.12.2021.
Disponibile al link: <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021DC0777&from=EN#footnote123>
[70] DECISIONE QUADRO 2008/913/GAI DEL CONSIGLIO sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale
Disponibile al link:<https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex%3A32008F0913>
[71]Regolamento UE 2022/2065
Disponibile al link:
<https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32022R2065&from=EN>
[72] Commissione Europea, Digital Services Act package.
Disponibile al link:
<https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

### 2.2 Responsabilità degli ISP

Le espressioni d'odio non sono legate solo al mondo digitale. Esse sono esistite fin dall'antichità, sia in maniera verbale sia in forma scritta. Citando Alessandro Spena "Odiare, per vero, si è sempre odiato"[73]. Guardando invece ai nuovi strumenti di comunicazione è innegabile notare come i social network abbiano accentuato le forme di intolleranza e di discriminazione. Alex Ingrassia (dottore di ricerca in diritto penale), nel suo trattato sul ruolo dell'ISP nel ciberspazio[74], identifica tre paradigmi di responsabilizzazione penale:

1. primo paradigma è caratterizzato dalla massimizzazione della libertà di espressione e comunicazione: l'Internet Service Provider è posto sullo stesso piano degli utenti, di conseguenza non ha alcun dovere di controllo rispetto alle condotte altrui, né di obblighi di denunciare o di oneri di collaborazione con le varie autorità. Gli viene riconosciuto semplicemente il ruolo di cittadino comune e, come tale, dal punto di vista penale la responsabilità è limitata al concorso commissivo doloso nell'altrui condotta criminosa.
2. Secondo paradigma invece vuole assicurare la tutela della comunità e dei soggetti terzi: l'ISP è obbligato a tenere un'attività di verifica e di censura preventiva di tutto il materiale che viene caricato, assumendo di conseguenza il ruolo di controllore. Sul piano penale la responsabilità è quella del reato omissivo improprio.
3. Terzo e ultimo paradigma: l'ISP si pone in una posizione intermedia tra i primi due punti, la repressione dei reati avviene solo *ex post* e quindi gli viene imposto l'obbligo di denunciare tutti gli illeciti di cui viene a conoscenza, ha degli oneri di collaborazione con le autorità e nell'individuazione di tutti gli autori degli illeciti, oltre all'obbligo di rimozione dei vari contenuti ritenuti proibiti. Va a limitare in maniera parziale la libertà di espressione, oltre a ridurre l'anonimato

[73] A. SPENA, La parola (-) odio. Sovraesposizione, criminalizzazione e interpretazione dello Hate Speech, cit., p. 577

[74] Definizione di Ciberspazio, Treccani, «spazio virtuale nel quale utenti (e programmi) connessi fra loro attraverso una rete telematica (v., per es., internet) possono muoversi e interagire per gli scopi più diversi» Disponibile al link: <https://www.treccani.it/vocabolario/ciberspazio/>

degli utenti. Dal punto di vista della responsabilità si configura il reato omissivo proprio[75].

Gli ISP giocano un ruolo centrale nella diffusione dell'*hate speech* online in quanto i messaggi d'odio divulgati in rete sono potenzialmente più pericolosi e dannosi rispetto a quelli offline, in quanto i contenuti discriminatori online sono permanenti nel tempo, in quanto possono essere più volte ripubblicati, basti pensare al sistema degli hashtag di Twitter o Instagram; dunque, possono creare dei danni irreversibili oltre ad essere raggiunti più facilmente da milioni di utenti. Nonostante lo scopo dei social network non sia quello di essere dei veicoli di diffusione di odio e discriminazione, purtroppo si sono rilevati dei perfetti mezzi di diffusione ti tali fenomeni, tanto che hanno costretto i Governi a collaborare insieme per contrastarli. Ad oggi, sono molteplici i messaggi d'odio online che si sono trasformati in atti di violenza nella vita reale: ne è un esempio la sparatoria avvenuta nel 2019 a Christchurch in Nuova Zelanda, dove un uomo armato ha fatto irruzione dentro a due moschee facendo fuoco, il tutto in diretta streaming su Facebook[76]. E ancora l'omicidio del politico tedesco Walter Lübcke sempre nel 2019, da parte di un neonazista, a seguito del quale il Governo tedesco ha rafforzato la legislazione vigente, il NetzDG del 2017, includendo delle specifiche disposizioni per contrastare l'*hate speech* e le altre forme di incitamento all'odio. Tale legge per la tutela dei diritti sui social network in Germania impone, alle piatteforme social che abbiano più di cento segnalazioni annuali di redigere un rapporto sulla gestione delle segnalazioni di tutti i contenuti ritenuti illeciti. Il rapporto deve essere pubblicato sulla propria home page entro un mese dalla fine del semestre e deve essere facilmente individuabile, disponibile e direttamente accessibile[77]. A volte, però, sono gli stessi gestori dei social network come

[75] Alex Ingrassia, Il ruolo dell'ISP nel ciberspazio: cittadino, controllore o tutore dell'ordine? pagg. 5-6
Disponibile al link:
<https://archiviodpc.dirittopenaleuomo.org/upload/1351711435Il%20ruolo%20del%20ISP%20nel%20cyberspazio%20DPC.pdf>

[76] Nuova Zelanda, strage in diretta video in due moschee a Christchurch: 49 i morti, 2019.
Vedasi informazione:
<https://www.ansa.it/sito/notizie/mondo/oceania/2019/03/15/nuova-zelanda-strage-in-due-moschee-a-christchurch-49-i-morti_e6b34ef9-9876-4644-8f1c-8d1801681b6b.html>

[77]NetzDG (Netzwerkdurchsetzungsgesetz)
Disponibile al link: <https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html>
Report di luglio 2023, relativo al semestre gennaio - giugno, di Facebook al cui interno si possono trovare tutte le tabelle contenti i vari tipi di trasgressioni e il numero di post rimossi.
Disponibile al link: < https://scontent.fvce2-1.fna.fbcdn.net/v/t39.8562-6/364139226_306330431844353_783773684531438147_n.pdf?_nc_cat=107&ccb=1-

Facebook o Twitter a sospendere in maniera temporanea o definitiva i vari account, come nel caso dell'ex Presidente degli Stati Uniti d'America Donald Trump, il quale, a causa del suo comportamento di incitamento all'odio, non poté accedere ai suoi account per circa due anni. A seguito di tutte questi avvenimenti, il gruppo Meta (società proprietaria di Facebook, Instagram e WhatsApp, che ad oggi contano più di due miliardi di utenti attivi) non consente a organizzazioni o persone che rivendicano, proclamano missioni violenti o comunque azioni aggressive di poter utilizzare le sue piattaforme. Valuta tutte queste entità sia online sia offline e le classifica in tre categorie:

1. primo livello, sono appartenenti a questo gruppo tutte le entità che commettono gravi atti di violenza nel mondo reale, sostengano o organizzino azioni violente contro i cittadini. All'interno di questa classificazione sono incluse le organizzazioni terroristiche, di odio e criminali e Meta rimuove qualsiasi contenuto che elogia, supporti tali gruppi o le azioni commesse, come crimini d'odio, omicidi seriali, atti di violenza. Il livello 1 include anche qualsiasi organizzazione segnalata dal Governo degli Stati Uniti, come boss del narcotraffico, le Foreign Terrorist Organizations (FTO, organizzazioni terroristiche straniere) o i Specially Designated Global Terrorists (SDGT, terroristi globali con designazione speciale);
2. secondo livello, appartengono a questo gruppo i soggetti non statali violenti, cioè tutte quelle entità che commettono azioni violente contro soggetti statali e militari. A differenza delle entità classificate al livello 1, queste non prendono di mira i civili e anche in questo caso ne vengono rimossi i contenuti che vanno a supportare tali azioni;
3. terzo livello, appartengono a questo gruppo quelle entità che potrebbero commettere ripetute violazioni delle normative sull'incitamento all'odio o sulle organizzazioni pericolose della piattaforma. Sono incluse in questo livello, inoltre, le entità che dimostrano l'intento di commettere, in un futuro, atti di violenza, anche se non hanno ancora commesso alcun evento dannoso. Sono inclusi tutti i movimenti sociali militarizzati e le reti cospirazioniste[78].

7&_nc_sid=ae5e01&_nc_ohc=KBVZoheTET8AX_7qDDg&_nc_ht=scontent.fvce2-1.fna&oh=00_AfCosX_XgVHpbA_w-me50ZqkV_BOZSD2zu4iOOqM_I-5IQ&oe=64F40D40>

[78]Transparency Center Facebook, Organizzazioni e persone pericolose.

Il gruppo Meta negli ultimi anni, a seguito delle varie multe ricevute, ha fissato degli standard molto elevati e stringenti per quanto riguarda i contenuti che possono ledere i diritti della dignità umana. Questi standard sono basati sui feedback ricevuti dagli utenti tramite la segnalazione e dagli esperti del settore, oltre alle normative vigenti. Gli obbiettivi che si è posto sono il rispetto della libertà d'espressione (purché non vada in contrasto con le leggi in vigore) e la sicurezza, in quanto Facebook, secondo la Corte Suprema di Cassazione con la sentenza n.37596 del 2014, definisce la piattaforma come «"luogo" virtuale aperto all'accesso di chiunque utilizzi la rete»[79] e in quanto tale deve contribuire a rimuovere i contenuti che potrebbero portare a un rischio di violenza. Meta si è posta come obbiettivo, inoltre, il rispetto della dignità della persona e la privacy.[80]

In conclusione, i Provider sono soggetti a sistemi legali molto vincolanti, che impongono loro di rimuovere obbligatoriamente i contenuti ritenuti lesivi, oltre alla certezza e alla brevità delle tempistiche in cui farlo. Tutti i recenti sforzi a livello europeo hanno portato a una maggiore collaborazione con le varie piattaforme e a una elaborazione di nuovi codici di condotta.

Vedasi le informazioni:
<https://transparency.fb.com/it-it/policies/community-standards/dangerous-individuals-organizations/>
[79] Definizione ripresa nella sentenza numero 11679 del 2023 della Corte di Cassazione a pagina 5. Disponibile al link:
<https://www.italgiure.giustizia.it/xway/application/nif/clean/hc.dll?verbo=attach&db=snpen&id=./20230320/snpen@s50@a2023@n11679@tS.clean.pdf >
[80]Standard della community di Facebook.
Vedasi le informazioni: <https://transparency.fb.com/it-it/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2F>


*Figura 8: esempio di come un Social Network rimuova immediatamente un contenuto prima di essere pubblicato.*

### 2.3 Il rischio di discriminazioni e stereotipi dovuti alla datificazione

L'Intelligenza Artificiale ha assunto un ruolo centrale nella società moderna e questo può comportare dei gravi rischi per privacy degli utenti. Tutti i servizi online, che all'apparenza sembrano gratuiti, in realtà non lo sono poiché i gestori di tali piattaforme chiedono e ottengono i dati personali come prezzo da pagare per usufruire dei loro servizi. Una buona parte dei dati servono per poter profilare gli utenti. La profilazione viene descritta dall'art. 4 del GDPR come «*qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di*

*detta persona fisica*»[81]. In sintesi, la profilazione serve per categorizzare i vari comportamenti, le preferenze e i gusti di una persona fisica. Può essere utilizzata per molteplici fini, quello più noto è relativo alla personalizzazione delle pubblicità (cosiddetta Pubblicità mirata). L'analisi dei dati personali serve per alimentare gli algoritmi e renderli sempre più precisi e performanti, però non sono sempre infallibili, in quanto non utilizzano tutte le informazioni ma solo un numero determinato di dati e ciò può comportare a dei gravi vizi. Gli algoritmi possono essere alterati da *bias* (pregiudizi) di vario genere e il caso più eclatante è quello di Loomis negli Stati Uniti. Eric Loomis, nel 2013, era alla guida di un'auto che era stata utilizzata in una sparatoria nel Wisconsin, quando venne fermato dalle autorità e gli furono addebitati cinque capi d'imputazione. Per determinare la pena i giudici si affidano al software COMPAS, uno strumento di valutazione per prevedere il rischio di recidività di un soggetto e la sua pericolosità sulla base di diversi fattori, come le condizioni economiche, sociali, la presenza di precedenti giudiziari etc. Il tribunale di La Crosse (Wisconsin) lo condannò a 6 anni di reclusione e 5 di libertà vigilata, la difesa del Sig. Loomis chiese di poter esaminare il codice sorgente, in quanto violava la norma costituzionale del giusto processo. A seguito degli approfondimenti degli esperti e dello studio da parte di Propublica[82], si arrivò ad affermare che vi fosse un'elevata probabilità dell'algoritmo a giungere a conclusioni errate nel caso di imputati afroamericani, dichiarandoli ad alto tasso di recidività, nonostante non avessero commesso alcun tipo di reato nei precedenti due anni. L'algoritmo tendeva a giudicare i soggetti bianchi meno rischiosi di quanto lo erano effettivamente. Gli algoritmi possono essere viziati sia da pregiudizi che da discriminazioni come nel caso appena citato, possono anche violare la privacy degli utenti in quanto una raccolta massiva di dati potrebbe portare a violare gli stessi diritti degli utenti. I dati possono essere stereotipati in partenza, come nel caso della divergenza salariale tra uomo e donna; quindi, l'algoritmo prende un *bias* e se il soggetto che dovrebbe prendere che la decisione in merito non ne è consapevole o lascia decidere al sistema, si va ad automatizzare un processo che prenderà sempre delle scelte sbagliate;

[81] Articolo 4 GDPR
[82] Propublica è un'organizzazione a non scopo di lucro statunitense che mira a produrre giornalismo investigativo di interesse pubblico.
Vedasi le informazioni:
<https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis>

tutto questo processo prende il nome di *feedback loop negativo*[83]. Per far fronte al processo decisionale automatizzato, tramite l'applicazione dell'art.22 GDPR[84] il legislatore tutela il soggetto contro le decisioni prese dall'algoritmo. L'Unione Europea individua nella Legge sull'Intelligenza Artificiale[85] del 21 aprile 2021 i settori nei quali l'utilizzo dell'IA merita una particolare attenzione, come nel settore dell'occupazione, della gestione dei lavoratori e, facendo particolare attenzione all'assunzione, alla selezione di un soggetto e al monitoraggio delle performance lavorative. La datificazione, infine, può portare a un'eccessiva generalizzazione riguardante gruppi di individui, basata su delle caratteristiche comuni. Questo può generare stereotipi dannosi, oltre che a prendere scelte imprecise sulle decisioni che riguardano le persone singole. Per mitigare i rischi bisognerebbe adottare un approccio etico e responsabile, valutando gli algoritmi e testandoli al fine di individuare le eventuali discriminazioni o gli effetti negativi. È fondamentale mantenere sempre una trasparenza adeguata ai vari tipi di sistema IA utilizzati. Tale precisazione trova riferimento nel principio di proporzionalità della Legge sull'Intelligenza Artificiale, secondo cui bisogna fornire delle spiegazioni sempre chiare e accessibili sulle decisioni adottate dagli algoritmi. Ultimo punto da tenere in considerazione per poter limitare i danni è agire sulla base dell'etica e della morale: bisogna considerare molto attentamente le implicazioni sociali delle decisioni basate sulla datificazione e agire in modo responsabile al fine di non mettere a rischio la dignità dell'individuo.

[83] Alessando Mariani, Data Science Ethics: scovare le discriminazioni nei dati, Politecnico di Milano. Vedasi informazioni:
<https://www.frontiere.polimi.it/data-science-ethics-scovare-discriminazioni-dati/>
[84] Art. 22 GDPR «L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».
[85] Commissione Europea, Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione, Bruxelles, 21.4.2021, pag.29.
<https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0006.02/DOC_1&format=PDF>

### 2.4 L'impatto della tecnologia sulla libertà di espressione

Il diritto di esprimere i propri pensieri senza censure venne riconosciuto con l'entrata in vigore della Costituzione, più precisamente con l'articolo 21, il quale enuncia il diritto di manifestare liberamente il proprio pensiero con la parola, lo scritto o qualsiasi altro mezzo di diffusione. Nella definizione di libertà di pensiero non rientra solamente il pensiero strettamente inteso, ma anche tutte le manifestazioni di emozioni, sensazioni, stati d'animo o incitamenti all'azione.

Le innovazioni tecnologiche hanno consentito lo sviluppo di Internet, la cui operatività si basa su diverse caratteristiche:

1. la Globalità, in quanto offre un accesso immediato a qualsiasi informazione proveniente da qualsiasi parte del mondo;
2. la Decentralizzazione, in quanto essendoci un numero sempre maggiore di server distribuiti in tutto il mondo ciò comporta una difficile forma di controllo da parte delle autorità;
3. l'Apertura, in quanto l'accesso a Internet ha pochissime barriere, il servizio generalmente viene offerto a un prezzo accessibile da chiunque;
4. la Vastità: in quanto l'informazione può essere distribuita mediante qualsiasi mezzo di comunicazione, dal personal computer al telefono mobile;
5. l'Interattività, in quanto Internet si basa sul concetto di rendere possibili infiniti flussi di comunicazione da una parte all'altra del mondo.[86]

I social network nel corso del tempo hanno assunto un ruolo sempre di più centrale e significativo all'interno della nostra società, nelle relazioni fra soggetti, fino ad arrivare al mondo della politica. La stessa società Meta nel rapporto annuale relativo all'anno 2022 inviato alla Securities and Exchange Commission (SEC, l'ente federale statunitense garante del mercato azionario), ha evidenziato come la compagnia abbia effettivamente avuto un impatto sulle pratiche di e-commerce, sull'applicazione dei Regolamenti inerenti alla protezione dei dati, sulla tutela dei minori e sulle questioni

[86] Paolo Emanuele Rozo Sordini, LA LIBERTÀ DI ESPRESSIONE NELL'ERA DIGITALE: DISCIPLINA INTERNAZIONALE E PROBLEMATICHE, Working Paper n.52, cit. pag. 5-6, ottobre 2013.
Disponibile al link: <https://www.ispionline.it/sites/default/files/pubblicazioni/wp_52_2013.pdf>

legate ai diritti civili, oltre che alla moderazione dei contenuti[87]. A tutti questi aspetti bisogna aggiungere la questione relativa alla libertà d'espressione. I Social hanno il "potere" di rimuovere determinati tipi di contenuti dalla propria piattaforma e per poter far questo i gestori hanno elaborato, in maniera sempre più progressiva, delle linee guida sempre più stringenti per quanto riguarda la tutela della dignità delle persone. Si vede il caso di Facebook, che in caso di bullismo o intimidazioni, in primo luogo, fa una distinzione tra personaggi pubblici e privati. Con *personaggi pubblici* intende i funzionari governativi a livello statale e a livello nazionale, i candidati politici, gli utenti con oltre un milione di *followers* (utenti che seguono altri utenti) o soggetti che con la loro esposizione ricevono una copertura mediatica elevata. Per evitare di reprimere la libertà d'espressione e consentire il dibattito, la piattaforma rimuove solo le offese gravi o alcune offese in cui i personaggi vengono *taggati* (menzionati) in un post o commento. Gli *individui privati*, invece, sono gli utenti che utilizzano la piattaforma, con particolare riguardo per coloro che hanno un'età compresa tra i 13 e 18 anni, la quale riconosce una protezione maggiore[88] offrendo degli strumenti utili per proteggersi[89]. Mentre, per quanto riguarda l'incitamento all'odio, fortemente vietato da parte della piattaforma: Meta proibisce l'uso di stereotipi offensivi, la discriminazione in base alla razza, religione, etnia, nazionalità, disabilità etc. Questi comportamenti vengono puniti con la rimozione del post e, nei casi più gravi, con la sospensione dell'account[90]. Gli standard della *community* servono per determinare quali contenuti sono ammissibili e quali invece no, oltre che a descrivere quali sono le conseguenze in caso di violazione. Occorre fare una distinzione tra due ipotesi riguardanti il tema della libertà d'espressione: la prima ipotesi riguarda i contenuti che possono essere considerati reato, come nel caso della diffamazione o altri reati contro l'onore commessi nei social media. In questi casi,

[87] United States Securities and Exhange Commission, Annual report Meta Platform Inc. for fiscal year ended, 31 dicembre 2022. Rif. pag. 40.
Disponibile al link: <https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/e574646c-c642-42d9-9229-3892b13aabfb.pdf>
[88] Transparency Center Facebook, Bullismo e Intimidazioni.
Vedasi informazioni:
<https://transparency.fb.com/it-it/policies/community-standards/bullying-harassment/>
[89]Meta collabora con diverse ONG, esperti per la prevenzione e il contrasto di determinati comportamenti
Vedasi informazioni: <https://about.meta.com/actions/safety>
[90]Transparency Center Facebook, Incitamento all'odio.
Vedasi informazioni:
<https://transparency.fb.com/it-it/policies/community-standards/hate-speech/>

una volta che viene accertata la natura criminosa della condotta (processo non sempre semplice) come viene più volte dimostrato dalla nostra giurisprudenza. A tal proposito, con la recente sentenza n.14 del Tribunale di Frosinone del 19.1.2022, è stato sancito che «la diffusione di un messaggio diffamatorio attraverso l'uso della bacheca Facebook non può dirsi posta in essere con il mezzo della stampa, non essendo i social network destinati ad una attività di informazione professionale diretta al pubblico ed avendo essi una cassa di risonanza tendenzialmente più circoscritta; non per questo, tuttavia, detta diffusione è dotata di minore potenzialità negativa, anche perché, a differenza di quella a mezzo stampa, non è oggetto di controlli specifici ed al contempo è considerata quasi come un luogo, virtuale, in cui poter dire tutto ciò che si pensa» [91].Tale sentenza è in contrasto con quanto descritto dalla Cassazione penale sez. V, n.4239 del 21.10.2021: «Il delitto di diffamazione può essere commesso anche a mezzo di Internet, con uso dei social e tale ipotesi integra l'aggravata di cui al comma 3 dell'art. 595 c.p.. La riferibilità della diffamazione può basarsi anche su indizi, a fronte della convergenza, pluralità e precisione di dati quali il movente, l'argomento del forum su cui avviene la pubblicazione, il rapporto tra le parti, la provenienza del post dalla bacheca virtuale dell'imputato, con utilizzo del suo nickname, anche in mancanza di accertamenti circa la provenienza del post di contenuto diffamatorio dall'indirizzo IP dell'utenza telefonica intestata all'imputato medesimo»[92]. La protezione della libertà di espressione non è più un problema. In sintesi, se il contenuto viola una norma penale allora la rimozione del contenuto dal social network non solleva particolari questioni riguardo alla protezione della libertà di espressione. Tuttavia, ci possono essere altre questioni da affrontare in seguito alla rimozione del contenuto ritenuto lesivo, come la responsabilità del gestore della piattaforma o i tempi o i modi con cui è avvenuta la rimozione stessa. Queste ulteriori questioni potrebbero essere oggetto di dibatti legali e giudiziari. La seconda ipotesi riguarda invece il caso in cui il contenuto, pur non violando nessuna norma, viola comunque le linee guida del Social, il quale sanziona l'utente autonomamente, ad esempio rimuovendo il post o limitando la possibilità di utilizzare la piattaforma[93]. È

[91] Cit. Tribunale Frosinone, 19/01/2022, (ud. 10/01/2022, dep. 19/01/2022), n.14
[92] Cit. Cassazione penale sez. V, 21/10/2021, (ud. 21/10/2021, dep. 07/02/2022), n.4239
[93] Arianna De Conno, Libertà di espressione e social: qual è il confine? Pubblicato nel 2022
Disponibile al link:
<https://www.altalex.com/documents/news/2022/03/16/liberta-di-espressione-e-social-qual-e-il-confine>

quest'ultima l'ipotesi la più difficile da analizzare e la più controversa, in quanto bisogna chiedersi: *quando* un social limita effettivamente la libertà di espressione. Per cercare di rispondere a questa domanda si fa riferimento alla sentenza della Corte d'Appello dell'Aquila del 9.11.2021 n.1659, in cui un'utente del social network Facebook aveva più volte pubblicato fotografie, post e didascalie inneggiando al fascismo, facendo particolare riferimento a Benito Mussolini. Dopo aver ricevuto diversi avvisi da parte del gestore del servizio, gli venne sospeso l'account per un periodo di quattro mesi e l'utente per questo motivo fece causa a Facebook. Il Tribunale in primo grado accolse le domande attoree, condannando la società al pagamento di €15.000 a titolo di risarcimento del danno, giudicando erronea l'applicazione dei termini d'uso portando, quindi, ad una violazione del diritto di espressione e di manifestazione del pensiero. La motivazione avanzata dal Giudice sosteneva che non vi fosse stato alcun atto concreto volto a riaffermare o rifondare il Partito Fascista: la condotta dell'utente dovrebbe essersi concretizzata, pertanto, non vi è stata nessuna violazione della cosiddetta Legge Scelba del 1952, la quale vieta la riorganizzazione del Partito, punendo i soggetti con la reclusione dai tre ai dieci anni (art. 2). Poiché, in questo caso, l'utente aveva semplicemente espresso il proprio pensiero e le proprie ideologie per ottenere un confronto con altri utenti, il Tribunale, conferma che non ci sono state delle violazioni degli "standard della community", quindi ha voluto condannare Facebook al pagamento della somma pattuita. In appello, la Corte ha ritenuto applicabile la legge italiana e non quella irlandese, in quanto Facebook aveva richiesto di svolgere il processo in Irlanda, ai sensi del Reg. 593/2008 (legge applicabile alle obbligazioni contrattuali) all'art.6, dove si individua quale legge sia applicabile in base al Paese dove il consumatore, in questo caso utente, abbia la residenza abituale. Sempre la Corte ha individuato l'esistenza di un contratto a titolo oneroso e a prestazioni corrispettive (punto 8.4 della sentenza), per di più il contraente, in questo caso identificato come l'utente, è la parte debole e quindi necessità di una maggior tutela, a differenza dei contratti stipulati a titolo gratuito. Questo contratto viene definito a titolo oneroso in quanto, anche se non vi è un pagamento di un corrispettivo, si concede il diritto di utilizzare i propri dati personali (i quali fungono da *pagamento*) in cambio della possibilità di utilizzare il social. In conclusione, la Corte ha ritenuto che alcuni contenuti abbiano violato gli standard,

dando ragione a Facebook, come l'uso delle metafore per offendere un'iniziativa parlamentare non condivisa. La sospensione dell'account per la durata di trenta giorni è da considerarsi, quindi, legittima e non contraria al rapporto contrattuale instaurato tra i due soggetti. Ha considerato legittima la seconda sospensione, in quanto l'utente ha risposto ad un commento esprimendosi con tali parole: «[…] *Lo lascerei al suo posto esattamente come lascerei nella toponomastica di molte città via Lenin, via Marx, via Tito etc etc; per dimostrare ai posteri la stupidità umana [L'utente] è rassicurante* […]»[94]. Andando a ledere la dignità della persona definendola "stupida" e denigrandola, si ha un superamento eccessivo della libertà di manifestare il proprio pensiero. La Corte ha ritenuto illegittimi i quattro provvedimenti presi da Facebook, il quale ha rimosso: il post che conteneva la bandiera della Repubblica di Salò, in quanto, secondo le linee guida dovrebbe essere «un luogo in cui le persone si sentano libere di comunicare[95]»; una fotografia con annessa didascalia che riproduceva Monte Giano[96]; un post con una fotografia che raffigurava Adriano Visconti noto maggiore della Seconda Guerra Mondiale e, infine, l'ultimo post inerente a una foto di Benito Mussolini con descrizione "Viva Mussolini", il quale ha fatto scattare per altri trenta giorni la sospensione dell'account. Facebook riteneva che fosse una celebrazione del fascismo e di conseguenza una violazione delle linee guida, nonostante si trattasse semplicemente di esprimere, seppur in maniera non condivisibile, il proprio pensiero. Per tutte queste ragioni, la Corte d'Appello dell'Aquila ha sanzionato la compagnia al risarcimento di €3.000.[97] Questa sentenza dà vari spunti di riflessione per i gestori delle piattaforme sulla legittimità di sanzione delle condotte e provvedimenti molto severi, limitando il diritto di esprimere il proprio pensiero. Altro aspetto da tenere in considerazione è a chi assegnare il compito di regolamentare e di controllare l'Hate Speech: per l'ordinamento giuridico solo lo Stato può "limitare" le libertà dei

[94] Corte d'Appello dell'Aquila, Sentenza n. 1659/2021 pubbl. il 09/11/2021 RG n. 295/2020 Repert. n. 1676/2021 del 09/11/2021.
Disponibile al link: <https://giuridica.net/wp-content/uploads/2021/12/corte-appello-aquila-1659-2021-risarcimento-ban-facebook.pdf>
[95] Vedi nota 80.
Vedasi informazioni:
<https://transparency.fb.com/it-it/policies/community-standards/?source=https%3A%2F%2Fit-it.facebook.com%2Fcommunitystandards>
[96] È una montagna che fa parte dell'Appennino abruzzese, sul quale nel 1939 è stata riprodotta la scritta DUX mediante la potatura della pineta.
[97] Vedi nota 94

cittadini. Nella rete, invece, tali provvedimenti vengono, di fatto, esercitati dalle piattaforme, che, oltre a detenere il potere di decisione su quale contenuto tenere e quale eliminare, spesso sono le uniche a possedere tutti i dati necessari per fare da mediatori in ogni processo di regolazione, di conseguenza è come se si andassero a sostituire allo Stato. Ad oggi, il gruppo Meta (possessore di Facebook, Instagram, Whatsapp, Threads) detiene quasi la totalità del mercato dei dati tra tutti i social network. Proprio per questo motivo la società ha creato l'Oversight Board, il cui intento è quello di rispettare le libertà di espressione ed è garantito da un giudizio indipendente. Le decisioni che vengono prese dal Board confermano o annullano le decisioni prese relative alla rimozione dei contenuti presenti sui social. Si tiene a precisare che è un ente separato dal gruppo Meta e fornirà una valutazione indipendente e non vincolata sui singoli casi[98]. Citando l'intervista effettuata da Ginevra Cerrina Feroni, Vicepresidente del Garante per la protezione dei dati personali, ribadisce: «*Si tratta di una pericolosa forma di "privatizzazione della censura" e del correlato fenomeno di "privatizzazione della giustizia digitale su scala globale", dai contorni ancora incerti, pertanto ancor più rischioso e da seguire con la massima attenzione. Una prospettiva aberrante, nella quale lo Stato rischia di abdicare al suo ruolo, delegando integralmente alle Internet platforms la regolazione del pluralismo informativo e il bilanciamento dei diritti fondamentali*[99]». Questa problematica è difficile da risolvere: una delle ipotesi potrebbe essere la collaborazione con questi soggetti al fine di creare una regolamentazione tale da non lasciare troppo campo alle Big Tech in merito alla questione di quando è corretto censurare un'utente e quando no.

[98] Sito ufficiale del Comitato per il controllo del gruppo Meta.
Vedasi informazione al sito ufficiale:
<https://www.oversightboard.com/>
[99] Libertà di espressione, Garante privacy: "Troppo potere alle big tech, ecco come intervenire" - Intervento di Ginevra Cerrina Feroni – AgendaDigitale, 11 maggio 2021
Disponibile al link:
<https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9584640>

## Capitolo 3

## Gli effetti dell'hate speech e le possibili metodologie di risoluzione

### 3.1 Effetti dell'*hate speech* sui soggetti e sulla società

L'*hate speech*, come abbiamo già affermato nei capitoli precedenti, è un discorso d'odio diretto a un gruppo di persone e basato sulla razza, sul sesso, sull'orientamento sessuale etc. A prescindere dalle modalità in cui si manifestano i discorsi d'odio, online oppure offline, hanno comunque delle conseguenze e producono effetti negativi sia sui soggetti che sulla società nel suo complesso. Nella vittima tali comportamenti possono causare alti livelli di ansia, depressione, stress, fino ad arrecare i DCA (Disturbi del Comportamento Alimentare) o suscitare istinti suicidi[100]. Possono, perciò, portare a danni sia fisici che mentali. Un altro aspetto fondamentale da tenere in considerazione è che chi è frequentemente esposti online a contenuti a sfondo razzista, sessista, omofobo e simili potrebbe sviluppare una sorta di "desensibilizzazione" emotiva, arrivando a giustificare le discriminazioni[101]. Anche il Consiglio d'Europa, all'interno del report *Protecting Women and Girls from violance in the digital age* del dicembre del 2021, associa tutti gli elementi che compongono la violenza psicologica, come l'istigazione al suicidio, le minacce di morte, insulti, attacchi verbali, al fenomeno dell'*hate speech*[102]. Un altro effetto che può produrre è l'isolamento sociale dei soggetti, in quanto provocherebbe in loro un senso di vergona o di colpa, facendoli dunque sentire respinti dalla società. Inoltre, un altro fattore che non aiuta le vittime di comportamenti discriminatori è l'utilizzo di account falsi da parte dei loro carnefici, in quanto le persone tendono a essere più aggressive se protette dall'anonimato[103].

---

[100] Brendesha M. Tynes and Kimberly J. Mitchell, Black Youth Beyond the Digital Divide: Age and Gender Differences in Internet Use, Communication Patterns, and Victimization Experiences, Journal of Black Psychology, 2014, Vol. 40(3), pag. 292.

[101] Wiktor Soral, Michał Bilewicz, Mikołaj Winiewski, Exposure to hate speech increases prejudice throughdesensitization, Aggressive Behavior, Volume 44, Issue 2, pagg. 136-146

[102] Adriane van der Wilk, Protecting Women and Girls from violance in the digital age, Consiglio d'Europa, dicembre 2020, pagg. 1-69.
Consuntabile al seguente link:
< https://rm.coe.int/the-relevance-of-the-ic-and-the-budapest-convention-on-cybercrime-in-a/1680a5eba3>

[103] Mainack Monda, Leandro Araújo Silva, Fabrício Benevenuto, A Measurement Study of Hate Speech in Social Media, HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media, 2017, pag.90

| Categories | Example of hate targets |
|---|---|
| Race | nigga, nigger, black people, white people |
| Behavior | insecure people, slow people, sensitive people |
| Physical | obese people , short people, beautiful people |
| Sexual orientation | gay people, straight people |
| Class | ghetto people, rich people |
| Gender | pregnant people, cunt, sexist people |
| Ethnicity | chinese people, indian people, paki |
| Disability | retard, bipolar people |
| Religion | religious people, jewish people |
| Other | drunk people, shallow people |

*Figura 9: parole più utilizzate per offendere in base alle diverse categorie*

In che modo gli effetti dell'*hate speech* danneggiano anche la società? Uno dei principali effetti riguarda l'aumento della violenza, sia fisica che verbale, e della discriminazione verso i gruppi considerati più vulnerabili. Tale fenomeno avviene attraverso intimidazioni, aggressioni o qualsiasi altra forma di abuso. Un recente studio condotto da VOX, l'*Osservatorio Italiano sui Diritti*, afferma che su 629.151 tweet, 583.063 erano discorsi d'odio, in particolare il 44,56% era rivolto alle donne (di cui l'89,9% sono dei tweet negativi) e il 31,84% a disabili (di cui l'99,9% sono dei tweet negativi)[104].

*L'hate speech* può portare anche a una divisone sociale: ne è un esempio la discriminazione verso le minoranze religiose che, oltre che a un deterioramento del rispetto sociale, può minare anche la coesione all'interno della comunità. Con tali azioni si incentiva l'odio e si arreca un danno psicologico alle vittime, le quali si sentono sminuite nel loro essere[105]. Tutti i discorsi d'odio rafforzano i pregiudizi preesistenti, alimentando la xenofobia, razzismo e l'omofobia e creando un ambiente in cui la tolleranza e l'accettazione sono messe in discussione.

[104] Mappa dell'intolleranza 7.0, VOX, 2022, pag. 4
[105] Valeria Fabretti, Gli effetti simbolici dello Hate Speech: riflessioni sul caso dell'odio rivolto a minoranze religiose, contenuto nel libro I discorsi dell'oltre: fascino e pericoli della polarizzazione a cura di Massimo Leone, Fondazione Bruno Kessler, 2023, pagg.23-32

| | Tweet totali | Tweet negativi rilevati | Tweet positivi | Tweet negativi geolocalizzati |
|---|---|---|---|---|
| Migranti | 53.962 (8,58%)* | 42.762 (79,2%)** | 11.200 (20,8%)** | 16.925 |
| Donne | 280.332 (44,56%)* | 251.950 (89,9%)** | 28.382 (10,1%)** | 90.924 |
| Islamici | 855 (0,13%)* | 854 (99,9%)** | 1 (0,1%)** | 284 |
| Disabili | 200.339 (31,84%)* | 197.957 (98,8%)** | 2.382 (1,2%)** | 68.632 |
| Ebrei | 39.236 (6,24%)* | 38.329 (97,7%)** | 907 (2,3%)** | 13.573 |
| Omosessuali | 54.427 (8,65%)* | 51.215 (94,1%)** | 3.212 (5,9%)** | 19.745 |
| **TOTALI** | **629.151** | **583.067 (93%)*** | **46.084 (7%)*** | **210.083** |

*Figura 10: tabella dei tweet rilevati da VOX contenti discriminazioni*

### 3.2 Metodi di risoluzione per contrastare i discorsi d'odio

*Education can counter hatred because it can contribute to inculcate children and youth with the values of respect for diversity, peaceful coexistence, and dialogue.*

*Alice Wairimu Nderitu*[106]

Per far fronte all'aumento dei discorsi d'odio, specialmente online, i vari governi europei devono affrontare questa minaccia adottando delle misure e azioni efficaci per combatterli. Nella Raccomandazione CM/Rec(2022)16 il Comitato dei Ministri, ha emanato delle linee guida che raccomandano ai 46 Stati membri nell'art. 2 "Quadro Giuridico" di andare a distinguere i casi più gravi di *hate speech,* che devono essere vietati dal Codice penale, il quale dovrebbe essere utilizzato come *extrema ratio*. Nel punto 11 del medesimo articolo si stabilisce che gli Stati membri dovrebbero specificare nel loro diritto penale nazionale quali siano le forme di incitamento all'odio soggette, come nella lettera B «*incitamento pubblico all'odio, alla violenza o alla discriminazione; contro minacce razziste, xenofobe, sessiste dirette contro le persone LGBT*»[107]. Devono, inoltre, distinguere i casi in cui si utilizzano delle espressioni offensive meno gravi, in questo caso bisognerebbe applicare il diritto civile e amministrativo.

Per combattere l'incitamento all'odio i governi europei dovrebbero stabilire delle disposizioni chiare e attuabili per contrastare l'*hate speech*. Anche le Nazioni Unite si

[106] United Nations Special Adviser on the Prevention of Genocide
[107] Recommandation CM/Rec (2022)16, du Comité des Ministres aux États members sur la lutte contre le discours de haine, 2022.
Disponibile al seguente link:
<https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a67951>

sono poste l'obbiettivo di cercare di eliminare i discorsi d'odio, promuovendo società sempre più pacifiche e inclusive mediante l'utilizzo di campagne di sensibilizzazione, come la *Summer School on Misinformation, Disinformation and Hate Speech,* organizzata dall'Istituto interregionale delle Nazioni Unite per la ricerca sul crimine e la giustizia (UNICRI, United Nations Interregional Crime and Justice Research Institute) e dalla Società Italiana per l'Organizzazione Internazionale (SIOI) tenutasi dal 3 al 7 luglio 2023. L'intento di questa iniziativa è quella di dare una panoramica generale al quadro giuridico internazionale e agli standard sui diritti umani utilizzati per contrastare i discorsi d'odio e la disinformazione, oltre a far comprendere il ruolo che hanno le piattaforme social nella diffusione di tali messaggi e come si possa utilizzare l'AI per creare disinformazione[108]. Sempre la SIOI e l'UNICRI hanno lanciato la prima edizione della *Autumn School on Hate Speech* che si terrà a ottobre del 2023 con gli stessi scopi di quella precedente[109]. Oppure la campagna di comunicazione *Lascia l'odio senza parole* già descritta nei capitoli precedenti. Questi sono obbiettivi totalmente in linea con l'Agenda 2030.

In risposta alla crescente presenza di manifestazione d'odio sia online che offline, il Segretario generale dell'Assemblea delle Nazioni Unite Antonio Guteress ha lanciato il 18 giugno 2019 una strategia e un piano d'azione sul loro contrasto. Questa strategia vuole sottolineare la necessità di combattere l'odio nella sua totalità, pur rispettando la libertà di espressione e di opinione, collaborando con organizzazioni, media, aziende tecnologiche e le piattaforme[110]. A seguito delle crescenti preoccupazioni globali si è voluta istituire il 18 giugno 2022 la Giornata internazionale contro i discorsi d'odio.

Per contrastare in maniera efficace bisogna agire in primo luogo da un punto di vista educativo, in quanto è uno degli strumenti più efficaci di prevenzione, rafforzando le varie

[108]UNICRI, SIOI, Summer School on Misinformation, Disinformation and Hate Speech, 2023.
Disponibile al seguente link:
< https://www.onuitalia.it/summer-school-on-misinformation-disinformation-and-hate-speech-3-7-luglio-2023-formato-ibrido-roma-italia-e-online-scadenza-per-la-presentazione-delle-domande-20-giugno-2023/>
[109] UNICRI, SIOI, Autumn School on Hate Speech, 2023.
Disponibile al seguente link:
< https://unicri.it/Training/School-Hate-Speech-2022 >
[110] Antonio Guteress, United Nations Srategy and Place of Action on Hate Speech, 2019.
Disponibile al seguente link:
< https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf>

politiche e programmi educativi già presenti nelle scuole con misure più specifiche. Si può ricorrere alla Global Citizenship Education (GCED), promossa dall'UNESCO che si basa sull'educazione alla pace e ai diritti umani. Il punto cardine della GCED è quello di infondere in tutti gli studenti «i valori, gli atteggiamenti e i comportamenti che supportano la cittadinanza globale responsabile: creatività, innovazione e impegno per la pace, i diritti umani e lo sviluppo sostenibile»[111] oltre a cercare di insegnare l'utilizzo corretto del mondo digitale[112]. Ulteriori metodi di risoluzione sono l'utilizzo della datificazione e del machine learning, come descritto nel capitolo 1 di questa tesi.

[111] Global citizenship education, UNESCO.
Tradotto dall'inglese: « the values, attitudes and behaviours that support responsible global citizenship: creativity, innovation, and commitment to peace, human rights and sustainable development».
Disponibile al seguente link: <https://en.unesco.org/themes/gced>
[112] The preventive role of education, United Nations, Hate Speech, 2021.
Consuntabile al seguente link:
<https://www.un.org/en/hate-speech/impact-and-prevention/preventive-role-of-education>

## Conclusioni

La datificazione e l'*hate speech* sono due fenomeni distinti, ma intrinsecamente collegati nell'era digitale. La datificazione si riferisce alla crescente raccolta, elaborazione e utilizzo dei dati in vari aspetti della nostra vita quotidiana, mentre l'*hate speech* è una manifestazione negativa delle comunicazioni online, caratterizzata da discorsi di odio, discriminazione e violenza verbale rivolti a gruppi o individui sulla base della razza, religione, orientamento sessuale e altri aspetti.

La mia tesi ha esaminato come la datificazione abbia un ruolo centrale per contrastare la diffusione dei discorsi d'odio su Internet. Ho analizzato come la raccolta sistematica dei dati personali e dei comportamenti online abbia alimentato la polarizzazione e la radicalizzazione di gruppi online, creando un ambiente fertile per l'*hate speech*[113]. L'uso di algoritmi di raccomandazione[114], che mostrano contenuti simili a quelli che l'utente ha visualizzato in precedenza, aiutando a prevedere le scelte e offrire dei consigli mirati, aiutano a creare le cosiddette *echo chamber* (camere dell'eco), in cui le persone vengono esposte a opinioni simili alle loro. Questo può alimentare l'odio e la radicalizzazione, facilitando la diffusione dell'*hate speech*. Ho approfondito come gli algoritmi di *Machine Learning* possono essere addestrati per rilevare automaticamente l'*hate speech* e limitarne la diffusione: come si è potuto notare nella figura 8, quando un utente ha provato a caricare un'immagine a sfondo razzista sulla piattaforma social Instagram, la datificazione ha identificato il contenuto e passato l'informazione agli algoritmi che l'hanno rimosso ancora prima della pubblicazione. Inoltre, l'analisi dei dati può aiutare a identificare i modelli di comportamento che portano all'*hate speech*, consentendo agli operatori delle piattaforme e alle autorità di intervenire in modo mirato.

Altro tema che ho sviscerato in questo elaborato riguarda le normative vigenti che concernono i discorsi d'odio e di come l'Unione Europea stia cercando di armonizzare gli Stati membri mediante la creazione di leggi sempre più comunitarie e di come l'Italia stia cercando di adeguarsi. Il processo di datificazione però, come si è analizzato, può

[113] Valeria Fabretti, Gli effetti simbolici dello Hate Speech: riflessioni sul caso dell'odio rivolto a minoranze religiose, contenuto nel libro I discorsi dell'oltre: fascino e pericoli della polarizzazione a cura di Massimo Leone, Fondazione Bruno Kessler, 2023, pag. 23.
[114] Francesco Ricci, Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook, Spinger, 2010, pag.1.

comportare anche dei rischi, ad esempio un'eccessiva violazione della privacy, delle discriminazioni dovuti a dei *bias*, come nel caso della scelta del personale da parte di algoritmi. Per contrastare questi rischi, è fondamentale adottare una serie di misure: è necessario promuovere la trasparenza nell'uso dei dati e garantire il consenso informato da parte degli utenti al trattamento dei dati personali, a partire dal modo in cui vengono raccolti fino al modo in cui vengono elaborati e utilizzati. Inoltre, è importante sviluppare algoritmi più etici ed equi al fine di evitare la diffusione dell'*hate speech*.

Altro aspetto da tenere in considerazione riguarda quanto sia fondamentale la formazione e la sensibilizzazione dei cittadini allo scopo di educarli ai pericoli portati dai discorsi d'odio, incoraggiando comportamenti online e offline più responsabili. In questo caso sia L'Unione Europea sia l'Italia si stanno muovendo nella giusta direzione con campagne sempre più attente a questa tematica. È essenziale una stretta collaborazione tra le piattaforme online e le autorità governative per identificare e rimuovere i contenuti dannosi in maniere tempestiva. Bisogna sottolineare che il contrasto all'*hate speech* non dovrebbe comportare una limitazione della libertà d'espressione, principio che viene tutelato anche dalla Costituzione oltre che dalle normative europee. Le soluzioni devono essere rispettose dei diritti fondamentali, garantendo al contempo un ambiente online più sicuro e inclusivo per tutti.

La datificazione offre numerose opportunità, ma presenta anche dei rischi significativi in termine di diffusione di discorsi d'odio. Affrontare questi rischi richiede un impegno costante e congiunto da parte della società, delle aziende Tech e dai governi stessi al fine di rimuovere e punire severamente qualsiasi contenuto ritenuto discriminatorio o violento.

**Bibliografia**


1. M. Turing, Mind a quarterly review of psychology and philosophy, I. Computing Machinery and Intelligence, Vol. Lix. No. 236, ottobre 1950, pp. 433-460
2. John McDermott, R1: a rule-based configurer of computer systems, aprile 1980, Carnegie-Mellon University.
3. Henry Shevlin, Karina Vold, Matthew Crosby & Marta Halina, The limits of machine intelligence, Embo reports, Volume 20 Issue 10, dell'ottobre 2019, pag. 1-5.
4. George Dvorsky, How Much Longer Before Our First AI Catastrophe? Pubblicato nell'aprile 2013
5. Daron Acemoglu e Pascual Restrepo, The wrong kind of AI? Artificial Intelligence and the future of labor demand, Working Paper 25682, National Bureau of Economic Research, marzo 2019, pp.1-10.
6. Claudio Sarra, Il mondo-dato, 2019, Cleup SC, pag. 17.
7. Garante per la protezione dei dati personali, Cosa intendiamo per dati personali?
8. U.S. Department of Labor, Guidance on the Protection of Personal Identifiable Information.
9. ENISA (European Union Agency for Cybersicurity), Pseudonymisation techniques and best practiesnovembre 2019, pag.9
10. Claudio Sarra, Datificazione e ingegneria del simbolico, in Saggi a margine dei seminari virtuali di Journal of Ethics and Legal Tecnologies, Primiceri Editore, Padova 2022, pag. 155.
11. Mohamed Y. Eltabakh, Mayuresh Kunjir, Ahmed Elmagarmid, Mohammad Shahmeer Ahmad, Cross Modal Data Discovery over Structured and Unstructured Data Lakes, Proceedings of the VLDB Endowment Volume 16 Issue 11, agosto 2023, pagg.1-17.
12. Kenneth Cukier and Viktor Mayer-Schoenberger, The Rise of Big Data, MAY/JUNE 2013, Published by: Council on Foreign Relations, p.35
13. Seref SAGIROGLU and Duygu SINANC, Big Data: A Review, May 2013, Published by: Gazi University Department of Computer Engineering, Faculty of Engineering, Ankara, Turkey,

14. C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P.C. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw-Hill Companies, 2011, pag.26.
15. Ceylan Onay and Elif Öztürk, A review of credit scoring research in the age of Big Data, Journal of Financial Regulation and Compliance Vol. 26 No. 3, 2018, pag.382.
16. Camera dei deputati, La piramide dell'odio in Italia, Commissione "Jo Cox" su fenomeni di odio, intolleranza, xenofobia e razzismo, luglio 20216.
17. VOX (Osservatorio italiano sui diritti), Mappa dell'intolleranza 7.0, 2022
18. G.W. Allport, The nature of Prejudice, 1954, pag.49
19. G. Ziccardi, L'odio online, Milano: Raffaello Cortina, 2016, cit. pag. 21.
20. Peter Watson, Psychology and race, New Brunswick, NJ: Aldine Transaction, 2007, p.46
21. Edoardo Bazzaco, Ana García Juanatey, Jon Lejardi, Anna Palacios y Laia Tarragona, ¿Es odio? Manual práctico para reconocer y actuar frente a discursos y delitos de odio, pubblicato a Barcellona, novembre 2017, pag. 7-42.
22. Oscad (Osservatorio per la Sicurezza Contro gli Atti Discriminatori), Commissione straordinaria per il contrasto dei fenomeni di intolleranza, razzismo, antisemitismo e istigazione all'odio e alla violenza, Audizione del Prefetto Vittorio Rizzi, 22 giugno 2021, pag.1
23. Amnesty International Sezione Italiana, Hate Speech conoscerlo e contrastarlo guida breve per combattere i discorsi d'odio online, 2019, pag.19
24. European Commission,7th evaluation of the Code of Conduct, novembre 2022, pag.1
25. Vishal Gupta, A Survey of Text Mining Techniques and Applications, Lecturer Computer Science & Engineering, University Institute of Engineering & Technology, Journal of emerging technologies in web intelligence, VOL. 1, NO. 1, AUGUST 2009
26. Guansong Pang, Jitendra Singh Malik, Anton van den Hengel, Deep Learning for Hate Speech Detection: A Large-scale Empirical Evaluation, Singapore Management University, Research Collection School Of Computing and Information Systems, 2022, pagg. 1-4

27. Codice di condotta per lottare contro le forme illegali di incitamento all'odio online, 30 giugno 2016, Commissione Europea, ottrobre 2019.
28. Art. 3 Cost
29. Governo italiano Presidenza del Consiglio dei ministri, Campagna di comunicazione "Lascia l'odio senza parole, maggio 2023.
30. Consiglio d'Europa, Raccomandazione CM/2010(5), parte I, lett. B, para. 6
31. Art. 10 CEDU
32. DECRETO-LEGGE 26 aprile 1993, n. 122, convertito poi in legge il 25 giugno 1993 n.205
33. Camera dei deputati, Dichiarazione dei diritti in Internet, 2015.
34. AGCOM, Delibera n.159/19/CONS, Regolamento recante disposizioni in materia di rispetto della dignità umana e del principio di non discriminazione e di contrasto all'hate speech.
35. Convenzione Internazionale sull'eliminazione di ogni forma di discriminazione razziale, New York 1965.
36. Tribunale Roma, N. R.G. 64894/2019, Sezione diritti della persona e immigrazione civile,23/02/2020.
37. Assemblea Generale delle Nazioni Unite, Dichiarazione Universale dei Diritti Umani, dicembre 1948.
38. Art. 7 UDHR.
39. Art. 29 UDHR.
40. Rif. Sentenza n. 11 del 1968.
    I Diritti Fondamentali nella Giurisprudenza della Corte costituzionale, pag.5.
41. Rif. Sentenza n. 84 del 17 aprile 1969, pag.1378
42. Art. 30 della Dichiarazione dei Diritti Umani.
43. United Nations, Promotion and protection of the right to freedom of opinion and expression, del 2019 pag.5
44. Alessandro Di Rosa Hate speech e discriminazione, Mucchi Editore, maggio 2020, pag. 28.
45. Convenzione Europea dei Diritti dell'Uomo.
46. Consiglio d'Europa, Convention on Cybercrime, European Treatry Series – No. 185, Budapest, novembre 2001.

47. STE no.189.
48. Un'Europa più inclusiva e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio, Bruxelles 9.12.2021.
49. DECISIONE QUADRO 2008/913/GAI DEL CONSIGLIO sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale
50. Regolamento UE 2022/2065
51. Commissione Europea, Digital Services Act package.
52. A. SPENA, La parola (-) odio. Sovraesposizione, criminalizzazione e interpretazione dello Hate Speech, cit., p. 577
*53.* Definizione di Ciberspazio, Treccani.
54. Alex Ingrassia, Il ruolo dell'ISP nel ciberspazio: cittadino, controllore o tutore dell'ordine? pagg. 5-6.
55. NetzDG (Netzwerkdurchsetzungsgesetz)
56. Definizione ripresa nella sentenza numero 11679 del 2023 della Corte di Cassazione a pagina 5.
57. Articolo 4 GDPR
*58.* Art. 22 GDPR «L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».
59. Paolo Emanuele Rozo Sordini, LA LIBERTÀ DI ESPRESSIONE NELL'ERA DIGITALE: DISCIPLINA INTERNAZIONALE E PROBLEMATICHE, Working Paper n.52, cit. pag. 5-6, ottobre 2013.
60. United States Securities and Exhange Commission, Annual report Meta Platform Inc. for fiscal year ended, 31 dicembre 2022. Rif. pag. 40.
61. Cit. Tribunale Frosinone, 19/01/2022, (ud. 10/01/2022, dep. 19/01/2022), n.14
62. Cit. Cassazione penale sez. V, 21/10/2021, (ud. 21/10/2021, dep. 07/02/2022), n.4239
63. Arianna De Conno, Libertà di espressione e social: qual è il confine? Pubblicato nel 2022
64. Corte d'Appello dell'Aquila, Sentenza n. 1659/2021 pubbl. il 09/11/2021 RG n. 295/2020 Repert. n. 1676/2021 del 09/11/2021.

65. Libertà di espressione, Garante privacy: "Troppo potere alle big tech, ecco come intervenire" - Intervento di Ginevra Cerrina Feroni – AgendaDigitale, 11 maggio 2021.
66. Brendesha M. Tynes and Kimberly J. Mitchell, Black Youth Beyond the Digital Divide: Age and Gender Differences in Internet Use, Communication Patterns, and Victimization Experiences, Journal of Black Psychology, 2014, Vol. 40(3), pag. 292.
67. Wiktor Soral, Michał Bilewicz, Mikołaj Winiewski, Exposure to hate speech increases prejudice throughdesensitization, Aggressive Behavior, Volume 44, Issue 2, pagg. 136-146.
68. Adriane van der Wilk, Protecting Women and Girls from violance in the digital age, Consiglio d'Europa, dicembre 2020, pagg. 1-69.
69. Mainack Monda, Leandro Araújo Silva, Fabrício Benevenuto, A Measurement Study of Hate Speech in Social Media, HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media, 2017, pag.90.
70. Mappa dell'intolleranza 7.0, VOX, 2022, pag. 4.
71. Recommandation CM/Rec (2022)16, du Comité des Ministres aux États members sur la lutte contre le discours de haine, 2022.
72. Valeria Fabretti, Gli effetti simbolici dello Hate Speech: riflessioni sul caso dell'odio rivolto a minoranze religiose, contenuto nel libro I discorsi dell'oltre: fascino e pericoli della polarizzazione a cura di Massimo Leone, Fondazione Bruno Kessler, 2023, pagg.23-32.
73. Francesco Ricci, Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook, Spinger, 2010, pag.1.

**Sitografia**


1. Giorgio Grossi professore dell'Università degli Studi di Milano Bicocca, La nostra esistenza "datificata": ecco la nuova era del digitale, Agenda Digitale, 2023.
   < https://www.agendadigitale.eu/cultura-digitale/la-principale-conseguenza-iperevolutiva-della-rivoluzione-digitale-lesistenza-datificata/>
2. L'hate speech e la violenza verbale.
   <https://www.dirittodellinformatica.it/ict/web/lhate-speech-e-la-violenza-verbale-online.html/>
3. CESIE, sCAN: Sfide e risultati del 4°esercizio di monitoraggio in base al Codice di Condotta sottoscritto da Commissione Europea e piattaforme, IT.
   <https://cesie.org/studi/scan-monitoring-exercise/>
4. Definizione di Troll:
   <https://www.urbandictionary.com/define.php?term=troll>
5. Cineca, Text mining.
   <https://www.cineca.it/text-mining>
6. AGCOM, Progetto IMSyPP "Innovative Monitoring Systems and Prevention Policies of Online Hate Speech", 2020.
   <https://www.agcom.it/956>
7. Dichiarazione dei diritti in Internet, 2015.
   <https://www.camera.it/application/xmanager/projects/leg17/commissione_internet/dichiarazione_dei_diritti_internet_pubblicata.pdf>
8. Legge 29 maggio 2017, n. 71.
   <https://www.gazzettaufficiale.it/eli/id/2017/06/3/17G00085/sg>
9. Articolo 604-bis.
   <https://www.gazzettaufficiale.it/atto/serie_generale/caricaArticolo?art.versione=1&art.idGruppo=60&art.flagTipoArticolo=1&art.codiceRedazionale=030U1398&art.idArticolo=604&art.idSottoArticolo=2&art.idSottoArticolo1=10&art.dataPubblicazioneGazzetta=1930-10-26&art.progressivo=0#:~:text=E'%20vietata%20ogni%20organizzazione%2C%20associazione,%2C%20etnici%2C%20nazionali%20o%20religiosi.>

10. Dichiarazione Universale dei Diritti Umani. <https://www.ohchr.org/sites/default/files/UDHR/Documents/UDHR_Translations/itn.pdf>
11. Promotion and protection of the right to freedom of opinion and expression, United Nations, General Assembly del 2019, pag.5 <https://digitallibrary.un.org/record/3833657>
12. Carta dei Diritti Fondamentali dell'Unione Europea. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT>
13. Nuova Zelanda, strage in diretta video in due moschee a Christchurch: 49 i morti, 2019. <https://www.ansa.it/sito/notizie/mondo/oceania/2019/03/15/nuova-zelanda-strage-in-due-moschee-a-christchurch-49-i-morti_e6b34ef9-9876-4644-8f1c-8d1801681b6b.html>
14. Transparency Center Facebook, Organizzazioni e persone pericolose. <https://transparency.fb.com/it-it/policies/community-standards/dangerous-individuals-organizations/>
15. Standard della community di Facebook. <https://transparency.fb.com/it-it/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2F>
16. COMPAS Recidivism Risk Score Data and Analysis, Propublica. <https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis>
17. Alessando Mariani, Data Science Ethics: scovare le discriminazioni nei dati, Politecnico di Milano. <https://www.frontiere.polimi.it/data-science-ethics-scovare-discriminazioni-dati/>
18. Commissione Europea, Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione, Bruxelles, 21.4.2021, pag.29.

<https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0006.02/DOC_1&format=PDF>

19. Standard della community di Facebook.

<https://transparency.fb.com/it-it/policies/community-standards/?source=https%3A%2F%2Fit-it.facebook.com%2Fcommunitystandards>

20. Transparency Center di Facebook, Bullismo e Intimidazioni.
<https://transparency.fb.com/it-it/policies/community-standards/bullying-harassment/>
21. Meta, Safety Center.
<https://about.meta.com/actions/safety>
22. Transparency Center di Facebook, Incitamento all'odio.
<https://transparency.fb.com/it-it/policies/community-standards/hate-speech/>
23. Sito ufficiale del Comitato per il controllo del gruppo Meta.
<https://www.oversightboard.com/>
24. UNICRI, SIOI, Summer School on Misinformation, Disinformation and Hate Speech, 2023.
< https://www.onuitalia.it/summer-school-on-misinformation-disinformation-and-hate-speech-3-7-luglio-2023-formato-ibrido-roma-italia-e-online-scadenza-per-la-presentazione-delle-domande-20-giugno-2023/>
25. UNICRI, SIOI, Autumn School on Hate Speech, 2023.
< https://unicri.it/Training/School-Hate-Speech-2022 >
26. Antonio Guteress, United Nations Srategy and Place of Action on Hate Speech, 2019.
< https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf>
27. Global citizenship education, UNESCO.
<https://en.unesco.org/themes/gced>
28. The preventive role of education, United Nations, Hate Speech, 2021.
<https://www.un.org/en/hate-speech/impact-and-prevention/preventive-role-of-education>