

**Commissione Straordinaria per il contrasto dei fenomeni di intolleranza, razzismo, antisemitismo e istigazione all'odio e alla violenza**

*Indagine conoscitiva sulla natura, cause e sviluppi recenti del fenomeno dei discorsi d'odio, con particolare attenzione alla evoluzione della normativa europea in materia*

Audizione del professor Pier Luigi Parcu

Centre for Media Pluralism and Media Freedom

European University Institute

Roma, 10 febbraio 2022

*Onorevole Presidente, onorevoli Senatrici e Senatori,*

grazie per aver invitato me e il Centro che dirigo a partecipare ai lavori di questa Commissione. È un onore salutare la presidente Liliana Segre e contribuire alla vostra indagine conoscitiva, portando i risultati della ricerca e delle elaborazioni che, come Centro che si occupa del pluralismo e della libertà di informazione nell'Unione Europea, abbiamo sviluppato in circa dieci anni di attività.

Il nostro osservatorio non è focalizzato esclusivamente sul fenomeno dei discorsi d'odio, ma, indagando in maniera olistica su tutte le dimensioni del pluralismo e della libertà di informazione, e avendo come ambito tutta l'Unione Europea e (per alcuni aspetti) i Paesi candidati, ci permette di poter dare un contributo su tre punti evidenziati nel vostro programma di lavoro, ossia

- a) Inquadrare la situazione italiana nel contesto dell'Unione Europea, presentando i risultati del Media Pluralism Monitor, lo strumento con il quale ogni anno ci sforziamo di misurare lo stato del pluralismo dell'informazione dove è contenuto uno specifico indicatore in merito ai rischi dell'esposizione a contenuti illegali o dannosi. **(MPM2021)**.
- b) Collocare il fenomeno dell'*hate speech* nell'ambito dell'evoluzione tecnologica, economica e sociale del panorama informativo, inquadrando il tema dei discorsi d'odio e della loro propagazione nella rete in quello, più generale, dei rischi e delle sfide che la digitalizzazione pone al nostro sistema sociale, economico e democratico, che è necessario tener presente e affrontare senza negare i vantaggi e le opportunità che la stessa innovazione digitale ha aperto.
- c) Ricordare i rimedi che, in virtù della crescente attenzione e preoccupazione verso il fenomeno, si sono messi e si stanno mettendo in atto. Anche in questo caso, pur con le proprie specificità, le problematiche legate al contrasto dei discorsi d'odio sono riconducibili al più vasto dibattito sulla regolamentazione e sulla responsabilità della diffusione dei contenuti in rete, nel rispetto del principio della libertà di espressione.

A tale riguardo, sintetizzeremo le tappe fondamentali dell'evoluzione del quadro regolatorio europeo, cantiere in piena evoluzione, nel quale, in relazione ai principi fondamentali costituenti dell'Unione, l'obiettivo della lotta all'hate speech e alla disinformazione deve coniugarsi con quello della tutela della libertà di espressione).

In conclusione di questa breve analisi, cercheremo di mostrare potenzialità e rischi delle misure attualmente in discussione; di ipotizzare alcune proposte specifiche che possono incidere sui veicoli e sui meccanismi di propagazione dell'odio (e della disinformazione) in rete; e di sostenere la necessità un approccio complessivo e multidimensionale al fenomeno.

#### **a) Alcune evidenze dai risultati del Media Pluralism Monitor (2021).**

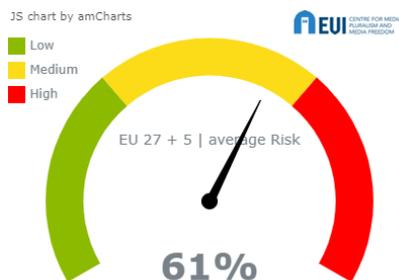
Vorrei partire dai dati che emergono dalla ricerca che conduciamo al Centre for Media Pluralism and Media Freedom all'Istituto Universitario Europeo di Firenze.

Il Media Pluralism Monitor (MPM), principale progetto del nostro Centro, è uno strumento di indagine e monitoraggio basato su un vasto questionario sullo stato dei media. Introdotto dal 2014, è adesso alla sua quinta applicazione a tutti i Paesi dell'Unione Europa e ai 5 Paesi candidati all'accesso. Da due anni i risultati del MPM sono utilizzati come fonte di riferimento nella valutazione di uno dei quattro pilastri del rapporto annuale della Commissione Europea sullo Stato di Diritto nei paesi dell'Unione.

La nostra indagine verte su 200 variabili, raggruppate in quattro aree: Protezione fondamentale, Pluralismo di mercato, Indipendenza politica e Inclusione sociale. Per ogni Paese monitorato, c'è un team di esperti locali (sono oltre 100 i ricercatori nel network), che rispondono utilizzando fonti primarie e dati secondari; le loro risposte sono riviste e controllate dai ricercatori del Centro, e in alcuni casi – le variabili per le quali c'è un maggiore margine di soggettività nella risposta – riviste e ulteriormente validate da un altro gruppo di esperti nazionali. L'intera ricerca si basa dunque su un approccio olistico e porta a una valutazione del rischio: un algoritmo di calcolo "pesa" le diverse variabili e indicatori e porta a definire se, per ciascun sub-indicatore, indicatore e area, c'è un rischio basso, medio o alto per il pluralismo dell'informazione.

L'area dell'Inclusione sociale, valuta il rischio relativo ai livelli di accesso ai media delle minoranze, delle comunità locali e regionali, delle donne, delle persone con disabilità; valuta inoltre il livello di alfabetizzazione mediatica e la **protezione dalla disinformazione e dai discorsi d'odio**. Le variabili che compongono quest'ultimo indicatore sono costruite in modo da tener conto della complessità del fenomeno: da un lato, la difficoltà di misurare la diffusione dei contenuti dannosi e la vulnerabilità ad essi; dall'altro, la valutazione delle politiche (eventualmente) adottate per contrastarli, e l'efficacia degli sforzi messi in atto per proteggere le categorie più esposte ai discorsi d'odio (donne, minoranze religiose ed etniche, persone disabili). Nelle risposte al questionario i ricercatori sono chiamati a accertare non solo se nel loro Paese sono state introdotte misure specifiche di contrasto ai discorsi d'odio,

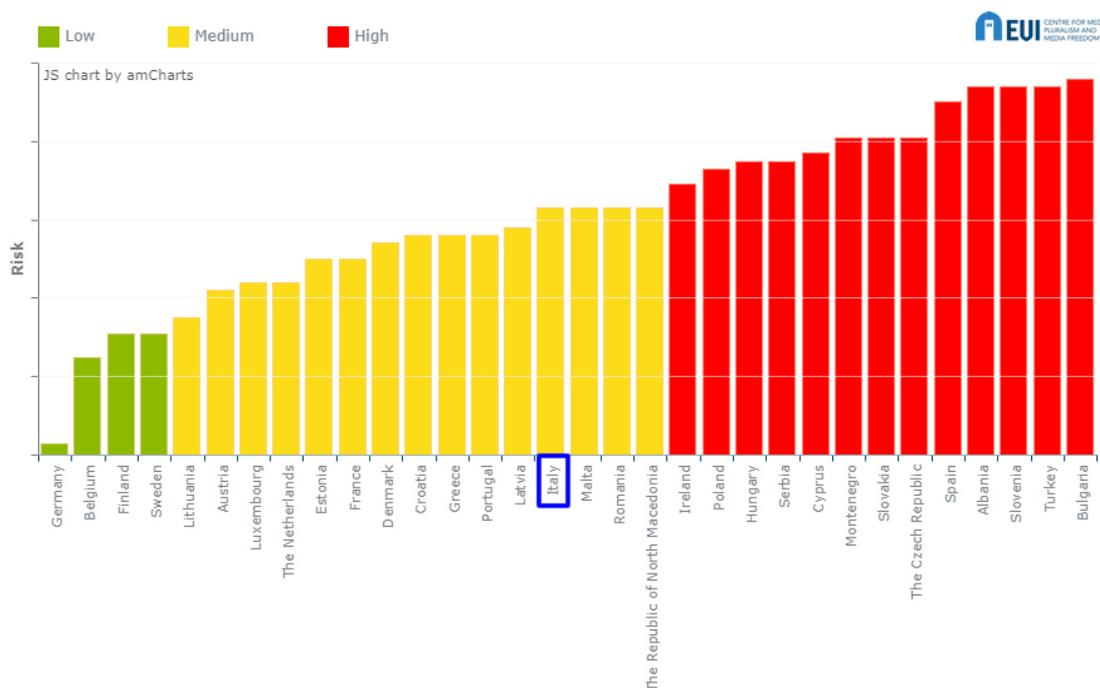
e se sono efficaci, ma anche se questo viene fatto nel rispetto dei diritti e delle libertà fondamentali, *in primis* la libertà di espressione.



Nella media dei Paesi monitorati (i 27 paesi dell’Unione Europea più i cinque candidati), l’indicatore sulla Protezione contro i contenuti illegali e dannosi presenta un rischio medio-alto. Si tratta di un livello preoccupante, poiché segnala che nella maggior parte dei paesi non esiste un quadro politico per combattere la diffusione della disinformazione, oppure quest’ultimo presenta aspetti critici in termini di rispetto di altre libertà fondamentali.

Andando a guardare la posizione del nostro Paese, l’Italia presenta un rischio medio-alto, al 63% in un intervallo che va da 0 (nessun rischio) a 100 (massimo rischio), si tratta di un risultato anche lievemente superiore alla media europea, che è del 61%.

### Media Pluralism Monitor 2021 – I rischi nella Protezione contro i contenuti illegali e dannosi (Unione Europea + Paesi candidati)

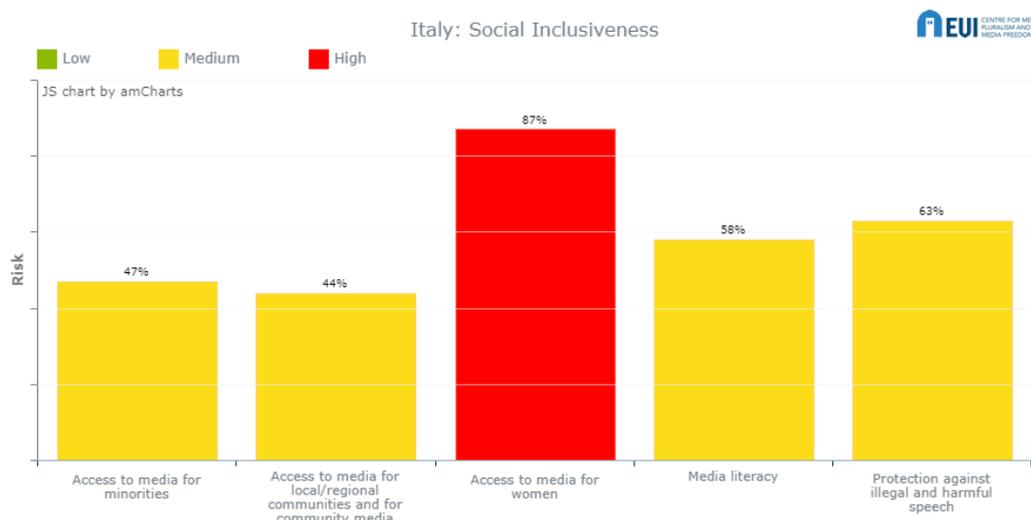


Fonte: MPM2021

Il risultato italiano è dovuto al fatto che, in assenza di una evoluzione del quadro normativo, sono state messe in atto iniziative per indurre le piattaforme online a prendere impegni in materia, ma manca la possibilità di verificare e valutare gli effetti della loro messa in pratica. Il livello preoccupante di rischio per questo indicatore va anche valutato insieme agli altri

indicatori nell'area dell'inclusione sociale, in particolare l'alfabetizzazione mediatica. Una bassa alfabetizzazione mediatica in Italia, con un rischio al 58%, rispetto a una media europea del 49%, contribuisce a una maggiore vulnerabilità agli attacchi e ai discorsi d'odio.

### Media Pluralism Monitor 2021 – Italia – I rischi nell'area dell'inclusione sociale



Fonte: MPM2021

In parallelo, i dati pubblici a disposizione, da fonti ufficiali o da osservatori indipendenti, mostrano che i crimini d'odio sono cresciuti.

Dal report MPM per l'Italia, anno 2021:

I crimini d'odio riportati dalla polizia sono raddoppiati dal 2015 al 2019 (anno al quale risalgono gli ultimi dati disponibili). Nel 2019, sono stati denunciati 1.119 crimini d'odio, 805 dei quali relativi a razzismo e xenofobia; 107 per pregiudizi relativi agli orientamenti sessuali e all'identità di genere; 207 per pregiudizi contro altri gruppi (come le persone con disabilità). Secondo l'Osservatorio Vox sui diritti fondamentali, nel 2020 l'hate speech online ha modificato le sue caratteristiche: meno attacchi in quantità, ma un uso più "professionale" dei social media da parte degli haters, spesso organizzati in gruppi con azioni coordinate. Quasi la metà degli attacchi sono stati specificamente diretti contro le donne.

Ma è cresciuta anche la consapevolezza sul problema e ci sono numerose iniziative dal basso, della società civile.

Per quanto riguarda il quadro europeo, la maggior parte dei Paesi, come il nostro, fa affidamento su un apparato legale costruito per contrastare i crimini d'odio offline, non efficace nell'ambiente digitale, in quanto di difficile e lenta attivazione, e inadatto a tracciare e reprimere la disseminazione dell'odio online.

Alcuni Paesi hanno adattato il loro quadro normativo: tra questi troviamo Germania, Francia, Spagna e Bulgaria. La legge tedesca, adottata nel 2017 (NetzDG), prevede dettagliate norme che pongono a carico delle piattaforme la responsabilità della rimozione tempestiva di contenuti illegali (entro 24 ore dalla segnalazione, in caso di contenuti "manifestamente illegali"), e obblighi di trasparenza del loro operato. A cinque anni dalla sua introduzione ci sono i primi bilanci sulla sua efficacia. Nel 2021, uno studio commissionato dal Ministero della

Giustizia (condotto dal prof. Niels Eifert) ha concluso che “gli obiettivi prefissati sono stati in gran parte raggiunti”. Recentemente la legge è stata ulteriormente rafforzata.

In Francia è stata adottata nel luglio 2020 la cosiddetta “legge Avia”, ispirata dalla legge tedesca. La legge Avia intende rafforzare il contributo degli intermediari digitali alla lotta contro i “contenuti odiosi online” (razzismo, antisemitismo, discriminazione di genere, omofobia, tra gli altri) nonché contro il terrorismo e la pedopornografia, chiedendo ai provider e alle piattaforme di rimuovere il contenuto odioso quando segnalato, entro 24 ore. A seguito della legge Avia, nell'ottobre 2020 è stato creato un Osservatorio sull'odio online.

In Francia, l'adozione di questa legge ha suscitato un vivace dibattito, nel quale i punti critici evidenziati sono: il ruolo decisionale delle piattaforme private, il rischio di autocensura o di affidamento al meccanismo selettivo degli algoritmi, la mancanza di controllo giudiziario dell'intero processo.

Anche in altri paesi, come in Spagna e Bulgaria, i tentativi di sviluppare un quadro giuridico con cui combattere la disinformazione online hanno sollevato preoccupazioni riguardo al loro impatto sulla libertà di espressione.

Un approccio diverso è quello di Paesi come la Finlandia e la Lituania, che hanno introdotto piani d'azione per combattere l'*hate speech* senza modificare il proprio quadro giuridico. Le azioni coinvolgono polizia, media, scuola e formazione (media literacy), con lo sviluppo di strumenti più efficaci e il rafforzamento delle abilità tecnologiche a tutti i livelli citati.

#### **In sintesi:**

Il MPM segnala un livello di rischio medio nell'Unione Europea, che è dovuto nella maggior parte dei casi al fatto che in molti Paesi il quadro giuridico non è stato adattato all'evoluzione dei discorsi d'odio online, e in alcuni Paesi al fatto che le nuove leggi pongono rischi per la libertà di espressione e lo stato di diritto.

È generalizzata la mancanza di dati sul fenomeno, e anche sul monitoraggio dell'efficacia delle contro-misure eventualmente messe in atto per contrastarlo.

In Italia, un Paese a rischio medio-alto, il risultato è legato all'assenza di misure specifiche, ai limiti dell'auto-regolazione delle piattaforme, e ad altri fattori di vulnerabilità, *in primis* la bassa alfabetizzazione mediatica e la maggiore esposizione delle minoranze ai discorsi d'odio.

Nel quadro europeo, sono da segnalare come esempi di strade alternative per combattere i discorsi d'odio quello Germania (dove si è scelta la via legislativa); e quelli della Finlandia e Lituania (dove si è scelto di intervenire con piani d'azione specifici, senza cambiare il quadro legislativo).

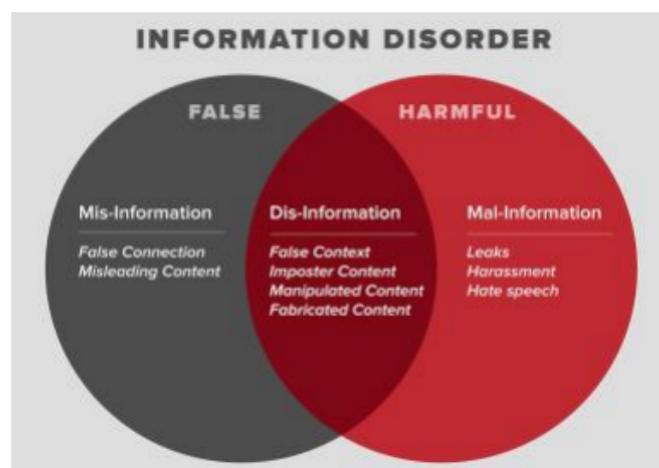
#### **b) L'informazione nell'ecosistema digitale**

Nell'era digitale, la diffusione dell'informazione ha raggiunto livelli senza precedenti, tanto da fare parlare di “era dell'informazione” (M. Castells, *The rise on the network society. The information age: economy, society and culture*. Blackwell, 1996), caratterizzata da facilità e rapidità di accesso alle notizie e alle informazioni, riduzione dei costi di produzione e distribuzione, riduzione delle barriere geografiche, fisiche, economiche.

Ma le stesse caratteristiche economiche e tecnologiche dell'era dell'informazione portano, insieme a indubbi ed indiscussi vantaggi, una serie di rischi, sui quali l'attenzione e le

preoccupazioni sono via via cresciute. Data spartiacque, in questo senso, è il 2016, con la Brexit prima e le elezioni presidenziali americane subito dopo: in entrambi i casi, è emerso il tema dell’impatto della diffusione di informazioni inesatte o false su scelte fondamentali per la vita democratica, fino alla tragica morte della deputata inglese Joe Cox, vittima prima di attacchi online e poi di un attentato omicida durante la campagna elettorale sulla Brexit.

Un rapporto del 2017 del Consiglio d’Europa inquadra l’hate speech nel più generale contesto del “disordine informativo”, all’estremo – il più pericoloso - della mal-informazione (Council of Europe 2017, *Information disorder: Toward an interdisciplinary framework for research and policy making*).



Fonte: Council of Europe report DGI(2017)09

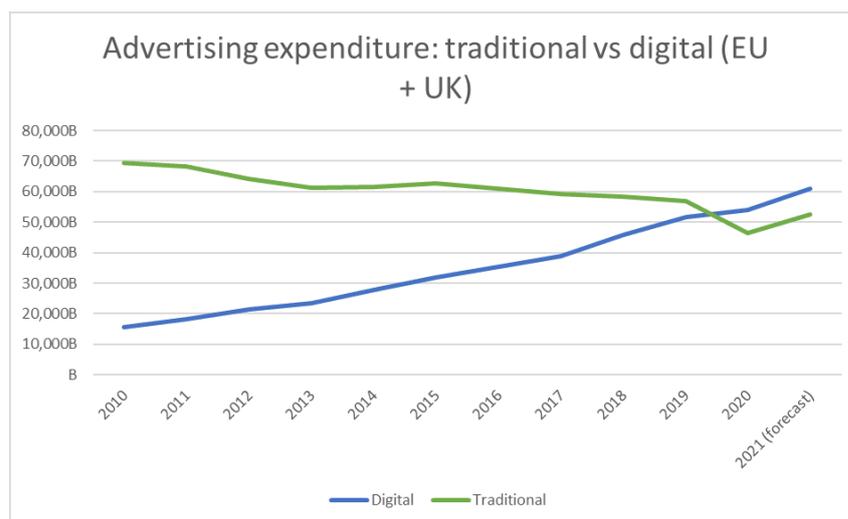
La mis-informazione (la distribuzione o condivisione non intenzionale di notizie false), la dis-informazione (la distribuzione o condivisione intenzionale di notizie false), la mal-informazione (l’uso di notizie vere o false con la deliberata intenzione di recare danno – categoria in cui sono compresi i discorsi d’odio), non sono fenomeni nati con internet, né sono fenomeni limitati all’ambiente digitale. Non è demonizzando internet che si possono affrontare questi pericoli, ma cercando di comprendere quali caratteristiche strutturali dell’ecosistema digitale li possono nutrire, amplificare, disseminare.

Semplificando, possiamo dividere queste caratteristiche in due grandi categorie:

- 1) **Caratteristiche economiche.** Il modello di business che caratterizza l’ambiente digitale è quello delle piattaforme. Nel modello economico dei media tradizionali, i media erano al centro di un mercato a due lati, nel quale da un lato vendevano (oppure offrivano gratis) contenuti da loro prodotti al pubblico, attirando l’attenzione dei lettori-utenti; dall’altro vendevano tale attenzione (l’audience, i lettori) agli inserzionisti pubblicitari. Nell’ambiente digitale, le piattaforme sono al centro di un mercato multi-dimensionale, nel quale raccolgono e distribuiscono contenuti generati sia dai media tradizionali che dagli utenti stessi; attraggono così l’attenzione degli utenti; e la vendono agli inserzionisti pubblicitari con una incrementata efficienza, dovuta al fatto di poter fornire informazioni e messaggi personalizzati su misura

dell'utente stesso, utilizzando la risorsa centrale del nuovo ecosistema dei media: i dati personali dei suoi utenti. Grazie a questo modello, le piattaforme sono in grado di sfruttare effetti di rete diretti e indiretti, ed economie di scala. Anche nell'era precedente, quella che ha caratterizzato l'evoluzione del sistema dei media nel Novecento, economie di scala e tendenza al gigantismo erano diffuse, tant'è che tutta la legislazione a tutela della concorrenza e del pluralismo esterno dell'informazione si è costruita attorno agli obiettivi di garantire la trasparenza della proprietà dei mezzi di informazione, limitare la loro concentrazione, e contrastare i pericoli di conflitti tra gli interessi economici e l'indipendenza editoriale e politica. Ma nell'era digitale il tema del pluralismo esterno dell'informazione si pone in termini del tutto nuovi. Da un lato, i media tradizionali sono stati spiazzati e messi in crisi dai concorrenti digitali sul mercato "dell'attenzione", che ne hanno fatto crollare le entrate. Dal 2020 nei Paesi dell'Unione europea le entrate pubblicitarie online hanno sorpassato la somma delle entrate di tutti i media: televisione, radio giornali, cinema, affissioni.

### Entrate pubblicitarie in EU – tradizionali vs digitali



Fonte: elaborazione CMPF su dati WARC

Dall'altro lato, i nuovi protagonisti dell'ecosistema informativo sono soggetti che – come da loro stessi ampiamente rivendicato – non si considerano “media”, in quanto non producono contenuti originali, ma raccolgono e diffondono contenuti prodotti da altri, professionalmente o no; così facendo, condizionano/controlano la distribuzione dell'informazione in mercati altamente concentrati (winners-take-all). La “innovazione distruttrice” del digitale certamente non riguarda solo l'industria dei media e le sfide poste al quadro della legislazione antimonopolistica sono diffuse a tutti i settori e in tutti i Paesi occidentali; ma nel settore dei media gli elementi di preoccupazione e allarme sono particolari e maggiori, dato che in ballo non ci sono solo le libertà economiche e la sopravvivenza di un mercato competitivo, ma la configurazione del dibattito nella sfera pubblica e la tutela di principi come la libertà di espressione, il pluralismo, il diritto all'informazione, il diritto alla privacy.

2) **Caratteristiche tecnologiche.** Nell'ambiente digitale è cambiato il modo in cui si accede all'informazione e la si consuma. L'informazione può essere prodotta da innumerevoli soggetti e da questi distribuita in tempo reale: questo ha aperto spazi senza precedenti di pluralismo e diversità, soprattutto in contesti totalitari (si pensi alla cosiddetta primavera araba), ma non solo (si pensi alla possibilità di far conoscere velocemente avvenimenti che si verificano in qualsiasi luogo, anche remoto e non coperto dal presidio dei media); ma allo stesso tempo può esporre alla diffusione di informazioni non verificate da soggetti professionali e responsabili sia dal punto di vista giuridico sia reputazionale dei contenuti che pubblicano; tali informazioni possono provenire ed essere disseminate dagli utenti stessi, oppure da soggetti e organizzazioni, a volte non riconoscibili o anche non soggetti alla giurisdizione del Paese in cui operano (strategie di disinformazione). Non solo. Le modalità tecniche della riproduzione e diffusione dei contenuti portano alla loro disaggregazione, le notizie "escono" dal contenitore in cui sono state prodotte, per cui spesso non si riesce a definire o ricordare la fonte da cui l'informazione proviene. E soprattutto si arriva all'informazione attraverso "porte d'accesso" algoritmiche, che selezionano i contenuti proposti sulla base di criteri orientati alla massimizzazione della diffusione (sia in quantità che in velocità) e alla personalizzazione dell'offerta, resa possibile dalla disponibilità di dati sulle preferenze e le scelte dell'utente anche attraverso diverse piattaforme. Queste modalità d'accesso possono favorire la creazione di "bolle", o la polarizzazione della sfera pubblica: non entreremo qui nell'analisi – anche controversa – di tali fenomeni, ma, per quello che interessa in questa sede, è opportuno sottolinearne un aspetto: la maggiore difficoltà nel riconoscere la fonte, il ruolo degli aggregatori automatici, e la generazione e diffusione di contenuti in virtù di aggregatori e "diffusori" anonimi (bots) sono motivi di allarme specifico per il fenomeno dei discorsi d'odio online.

**In sintesi:**

L'ecosistema digitale dell'informazione è centrato sull'uso dei dati personali degli utenti, che possono essere consumatori, creatori dei contenuti o svolgere entrambi i ruoli. In tale sistema, rischi per il pluralismo dell'informazione emergono dal lato dell'offerta e della domanda.

Dal lato dell'offerta, poiché i media, in competizione con le piattaforme, perdono una parte rilevante delle loro fonti di finanziamento, e questo riduce l'offerta informativa qualificata; e poiché loro stessi sempre più offrono un'informazione targettizzata e personalizzata, riducendo così l'esposizione a fonti diverse.

Dal lato della domanda, per la creazione di un ambiente iper-personalizzato, con il rischio di "camere dell'eco" e "bolle" che possono ridurre la scelta effettiva dei consumatori. In tale contesto va inquadrato il fenomeno dei discorsi d'odio: aumentata capacità di disseminazione, esposizione e vulnerabilità, e inefficacia degli strumenti di contrasto progettati per un contesto off-line.

### c) L'evoluzione del quadro regolamentare internazionale ed europeo

Un effetto comune delle caratteristiche economiche e tecnologiche appena riassunte è che, insieme al sistema dei media tradizionali, esse hanno impattato tutto il sistema giuridico e regolatorio costruito in un secolo a tutela del pluralismo dell'informazione, del contrasto ai poteri dei grandi conglomerati, della trasparenza della comunicazione politica, fino alla tutela dei diritti individuali fondamentali. Tratto comune è la non-responsabilità delle piattaforme per i contenuti che veicolano, sancita dalla direttiva europea sul Commercio Elettronico 2000/31 (e da una analoga normativa negli USA). Da allora, nei venti e più anni successivi si è sviluppato un dibattito crescente sulla necessità di intervenire sulla questione della responsabilità delle piattaforme. In quest'ottica, le misure per contrastare l'*hate speech* affrontano dilemmi comuni a tutte le politiche di contrasto alla disinformazione, in particolare nella necessità e difficoltà di bilanciare il contrasto alla disinformazione e all'*hate speech* con la tutela della libertà di espressione.

#### Livello internazionale

La **Convenzione internazionale sull'eliminazione di ogni forma di discriminazione razziale**, adottata nel 1965, pone alcune disposizioni generali, e specificamente prevede che “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”...;

La **Convenzione internazionale sui diritti civili e politici, adottata nel 1976**, prevede (articolo 20-2) che sia vietata dalla legge “qualsiasi difesa dell'odio nazionale, razziale o religioso che costituisca incitamento alla discriminazione, all'ostilità o alla violenza”.

Entrambi questi trattati sono stati progettati in un'era pre-Internet e, quindi, l'odio online non è stato preso in considerazione. Tuttavia, come sostiene Alkiviadou (2018), le disposizioni possono essere utilizzate per regolamentare l'incitamento all'odio che si trova su Internet in quanto non sono le caratteristiche e gli effetti del fenomeno a essere cambiati, ma, piuttosto, il veicolo che utilizza per la diffusione. Lo svantaggio di queste disposizioni è che, anche a causa di ragioni storico-sociali al momento della stesura, queste disposizioni affrontano solo i discorsi razzisti e religiosamente discriminatori.

#### Consiglio d'Europa

Nel 1997, la Raccomandazione del Comitato dei Ministri del Consiglio d'Europa (97) 20 definiva l'*hate speech* come “ricomprensivo tutte le forme espressive che diffondono, incitano, promuovono o giustificano l'odio razziale, la xenofobia, l'antisemitismo o altre forme di odio basate sull'odio sull'intolleranza, tra cui: l'intolleranza espressa attraverso il nazionalismo aggressivo e l'etnocentrismo, la discriminazione e l'ostilità nei confronti di minoranze, migranti e persone di origine straniera”. I principi elencati in quella prima

raccomandazione del '97 erano dunque da applicare all'hate speech così definito, ed **“in particolare ai discorsi d'odio disseminati attraverso i media”**.

Il **Protocollo addizionale alla Convenzione del Consiglio d'Europa sulla criminalità informatica**, riguardante la criminalizzazione degli atti di razzismo e xenofobia commessi a mezzo di sistemi informatici (adottato nel 2003, entrato in vigore a livello internazionale nel 2006). Questo documento è specificamente finalizzato al contrasto delle attività di odio online. Si occupa di razzismo e xenofobia nel quadro di minacce online, insulti, materiali e negazione del genocidio

Il Consiglio d'Europa sta lavorando a uno schema di Raccomandazione agli Stati membri, per **“un approccio globale per affrontare l'incitamento all'odio, anche nel contesto online, nel quadro del rispetto dei diritti umani”**

Inoltre, sempre nel quadro CoE, la raccomandazione 2018(2), pur non affrontando specificamente il tema del discorso d'odio, si occupa del ruolo e la responsabilità degli intermediari digitali.

### Unione Europea

**Decisione quadro sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale** (2008). Questo è forse il documento più importante a livello di Unione Europea, in quanto può essere utilizzato per criminalizzare l'incitamento all'odio. Non definisce direttamente l'incitamento all'odio. Tuttavia, vieta diverse forme di espressione e atti che rientrano nel quadro dei "Reati in materia di razzismo e xenofobia". Sebbene questo documento non affronti la questione dell'attività online, non la esclude.

**Commissione Europea, Codice di condotta sul contrasto del discorso d'odio illegale online**, firmato nel 2016 da Facebook, Microsoft, YouTube e Twitter (successivamente si sono aggiunti come firmatari: Instagram, Snapchat, Dailymotion, Jeuxvideo.com, TikTok, LinkedIn). Il Codice di condotta sottolinea il ruolo delle società firmatarie nel contrasto al discorso d'odio illegale, e al tempo stesso l'importanza di tutelare e promuovere la libertà di espressione. I principali impegni inseriti nel codice di condotta sono i seguenti:

**“Al ricevimento di una valida notifica di rimozione, le Società IT devono riesaminare tali richieste rispetto alle loro regole e linee guida comunitarie e, ove necessario, alle leggi nazionali di recepimento della Decisione Quadro 2008/913 /GAI, con termini dedicati per la revisione delle richieste”**.

**Monitoraggio:** l'ultima valutazione mostra che in media le aziende stanno valutando l'81% dei contenuti segnalati entro 24 ore e il 62,5% dei contenuti ritenuti illegali incitamento all'odio è rimosso.

Lo stesso dilemma caratterizza il successivo **Code of Practice contro la disinformazione online**, firmato tra la Commissione europea e le principali piattaforme il 26 aprile 2018 (Nella sua comunicazione, la CE riconosce che la disinformazione "lede la libertà di espressione, un diritto fondamentale sancito dalla Carta dei diritti fondamentali dell'Unione europea", ma anche che "il contenuto legale, anche quando ritenuto dannoso, è generalmente protetto da

libertà di espressione e deve essere affrontato in modo diverso rispetto ai contenuti illegali, caso nel quale la rimozione del contenuto stesso può essere giustificata”).

Sviluppi in corso: **Digital Services Act (DSA)** (presentato dalla Commissione europea nel dicembre 2020 come una proposta legislativa che mira a "contribuire alla sicurezza online e alla protezione dei diritti fondamentali e a stabilire una struttura di governance solida e duratura per l'efficace vigilanza dei fornitori di servizi di intermediazione." Si tratta di un regolamento orizzontale, tenendo conto della natura transfrontaliera dei servizi digitali e per evitare la frammentazione nel mercato interno. Il compito principale, evidenziato nella proposta, è garantire trasparenza e responsabilità, in particolare per le piattaforme online molto grandi, in merito alle loro pratiche di moderazione dei contenuti, alla pubblicità e ai processi algoritmici. Stabilisce inoltre obblighi per le grandi piattaforme di valutare i rischi posti dai loro sistemi e richiede loro di "sviluppare strumenti di gestione del rischio appropriati per proteggere l'integrità dei loro servizi dall'uso di tecniche manipolative". Il DSA intende distinguere tra contenuto illegale e contenuto che è dannoso, ma non illegale (come si è visto, questa è la categoria in cui a volte ricade il discorso d'odio). Questi ultimi non dovrebbero essere definiti nella legge sui servizi digitali e non dovrebbero essere soggetti a obblighi di rimozione.

Tuttavia sul tema la **Risoluzione del Parlamento europeo** 20 ottobre 2020 (2020/2022 (INI)), su "Digital service act and fundamental rights issues posed" invita a porre in atto una pluralità di strumenti per proteggere le persone dall'esposizione a contenuti dannosi ma non illegali (tra i quali le garanzie del procedimento giudiziario, la trasparenza delle piattaforme, il ruolo del servizio pubblico e il potenziamento dell'alfabetizzazione mediatica).

È in corso di definizione anche la nuova normativa europea sulla libertà dei media (**Media Freedom Act**), annunciata dalla presidente Ursula von der Leyen nel discorso all'Unione del 2021, sulla quale è aperta dal 10 gennaio la procedura di consultazione pubblica.

#### **In sintesi:**

A livello internazionale è riconosciuta la necessità di intervenire, e adeguare il quadro istituzionale e normativo alle sfide poste nel nuovo contesto digitale.

Una "via europea" si è introdotta con i Codici di condotta su *hate speech* e disinformazione, inizialmente basata sulla volontaria adesione delle piattaforme e sui loro impegni nel contrasto a questi due fenomeni.

Nei documenti europei è sempre riaffermata la necessità di attuare gli interventi contro *hate speech* (e disinformazione) all'interno del rispetto dei principi e diritti fondamentali, tra i quali la libertà di espressione.

Sono emersi i limiti dell'autoregolamentazione: l'ulteriore potere che così si attribuisce a soggetti già dominanti sul mercato – quello di rimuovere/accettare contenuti -; la difficoltà di limitare e controllare un uso discrezionale di tale funzione/potere; la scarsa trasparenza nel resoconto l'attività svolta e rendere pubblici criteri e dati.

## Conclusioni

Il quadro fin qui delineato fa emergere la complessità del fenomeno, ma allo stesso tempo l'urgenza di un intervento, chiamando le istituzioni, la ricerca, la società civile a mettere in campo strumenti e le competenze.

Un importante contributo potrà arrivare dall'evoluzione del quadro normativo dell'Unione europea, con il Digital Services Act (che impone nuovi obblighi alle grandi piattaforme) e con il Digital Markets Act (che ha lo scopo di garantire mercati digitali equi e aperti). Anche la nuova legge sui media, l'annunciato European Media Freedom Act, fa parte di questo quadro. Il "disordine informativo" – di cui i discorsi d'odio sono la manifestazione estrema e più pericolosa – si può in primo luogo combattere con una informazione libera, indipendente, plurale, affidabile e responsabile: una "buona" informazione che aiuti a scacciare quella "cattiva". Naturalmente anche i media stessi hanno avuto e hanno un ruolo nel diffondere i discorsi d'odio; e in tempi recenti questo fenomeno è persino cresciuto, con una interazione e una sorta di circolo vizioso tra social media e media tradizionali. Ma questi ultimi restano comunque responsabili di quel che pubblicano e diffondono.

Quanto alla responsabilità delle piattaforme nel contrasto ai discorsi d'odio online, si è già detto che i nuovi obblighi per le grandi piattaforme posti nel Digital Services Acts, in relazione ai discorsi d'odio, sarebbero limitati ai casi in cui il discorso d'odio è un crimine. Finora, i Codici di condotta volontariamente firmati dalle stesse piattaforme (sia contro la disinformazione che contro i discorsi d'odio) hanno fatte emergere i dilemmi e i limiti di cui si è detto, riguardo alla trasparenza del processo, alla sua effettività, al difficile bilanciamento con il rispetto di altri diritti umani fondamentali, primo tra tutti la libertà di espressione.

Tutti dilemmi e rischi sui quali il dibattito nel mondo politico, della ricerca e della società civile è aperto e potrà portare – si spera – a soluzioni nel medio e lungo periodo.

Vi sono però degli aspetti specifici, legati agli strumenti e alle modalità di diffusione del discorso d'odio nell'ambiente digitale, che potrebbero essere oggetto di interventi più immediati e di diretta applicazione. In particolare, si dovrebbero studiare gli strumenti per limitare o impedire alcune pratiche che si sono rivelate particolarmente adatte alla diffusione di disinformazione e *hate speech*. Tra queste, l'uso di bots - sistemi automatizzati di produzione e diffusione di contenuti -, che spesso sono specificamente progettati per dirigere, manipolare e orientare il discorso pubblico in rete, amplificando l'esposizione a contenuti non credibili, dannosi, offensivi; e l'ampio ricorso all'anonimato.

Per quanto riguarda l'uso dei bot: nel Code of Practice contro la disinformazione online, firmato da tutte le principali piattaforme con la Commissione europea, si fa menzione del problema. In particolare, le parti si impegnano a "mettere in atto politiche chiare sull'identità e l'uso improprio dei bot sui loro servizi", e a contrastare "usi inammissibili" degli stessi. Si concorda anche, in quel Codice, sulla importanza di stabilire "un chiaro sistema di segnalazione e regole per i bot, per assicurare che la loro attività non sia confusa con quella di interazioni umane". Sui criteri e l'attuazione di questi impegni non si ha però un monitoraggio comprensivo. Mentre ci sono evidenze di una intensificazione del ruolo dei bot nella disseminazione di disinformazione in occasione dell'emergenza COVID-19: secondo una

ricerca della Carnegie Mellon University che ha studiato l'attività dei bot nei primi mesi della pandemia, analizzando oltre 200 milioni di tweet, quasi la metà degli account che disseminavano messaggi sul COVID-19 erano account automatizzati: e si dovevano a bot il 62% dei 1000 retweet più influenti, e oltre l'80% dei 50 più influenti. Nel settembre dello scorso anno, Twitter ha introdotto una funzione che permette ai bot di "auto-identificarsi" come tali, su base volontaria (questo perché possono esserci bot "buoni", per esempio sulla diffusione di informazioni sui vaccini, o su emergenze).

Si potrebbe, invece che affidarsi all'autoregolamentazione in questo campo, prevedere un obbligo a carico delle piattaforme: l'obbligo di marcare, rendere riconoscibili, i bot. **In sostanza, garantire all'utente il diritto di sapere se sta interagendo con un essere umano o un software, e dunque se è una persona che pubblica il contenuto, lo diffonde, lo rilancia, lo commenta; oppure è un sistema automatizzato.**

La diffusione dell'hate speech è inoltre quasi sempre dovuta ad account anonimi. Quello della tutela dell'anonimato in rete è un tema delicato, che però con le dovute cautele andrebbe affrontato, per delimitare efficacemente i casi in cui la tutela fondamentale alla privacy della propria persona e corrispondenza possa essere limitata per tutelare altri diritti fondamentali.

Infine, una raccomandazione che emerge dal nostro lavoro in tema di pluralismo dell'informazione, è quella di ricercare ed utilizzare una pluralità di strumenti per affrontare le sfide e i rischi digitali. Ancor più tale approccio ci sembra valido per il contrasto ai discorsi d'odio, per il quale sarebbe sbagliato focalizzare l'attenzione solo sulla repressione, senza intervenire su tutte le politiche e le azioni positive di prevenzione, prima tra tutte quelle dirette a far crescere **l'alfabetizzazione mediatica**. Gli esempi europei, sia in positivo sia in negativo, ci chiamano inoltre a vigilare non solo sull'astratta validità di nuove leggi o interventi regolatori, ma sulla loro **applicabilità** e sulla loro pratica attuazione; elemento fondamentale è la **trasparenza** del processo, da pretendere per l'azione dei soggetti pubblici e a maggior ragione per quelli privati, ai quali si rischia di delegare funzioni delicatissime per la vita collettiva. Un monitoraggio costante, tempestivo e pubblico dei criteri e degli effetti di tali interventi si rende necessario, con l'attivo coinvolgimento dei soggetti della società civile.

Grazie per l'attenzione