

15 Febbraio 2022

Audizione della Prof.ssa Ginevra Cerrina Feroni

(Ordinario di Diritto Costituzionale Italiano e Comparato, Università di Firenze)

**Senato della Repubblica - Commissione straordinaria intolleranza, razzismo,
antisemitismo, istigazione all'odio e alla violenza**

Ringrazio molto la Presidente, il Vice Presidente e tutti gli onorevoli Senatori per l'invito a condividere alcune riflessioni in merito a temi molto significativi e rilevanti tanto sul piano istituzionale e giuridico, quanto su quello politico e dell'opinione pubblica ai quali questa illustre Commissione rivolge il suo interesse. Il lavoro che state svolgendo è davvero importante spaziando su varie direttrici in termini di approcci interdisciplinari trattati e temi compulsati.

Sommario: 1) Premessa di metodo. - 2) Definizione dell'oggetto. - 3) Odio come fattispecie penale. - 4) Un bilancio sulla libertà di espressione. - 5) L'odio in rete e il ruolo delle piattaforme.

1. Premessa di metodo.

Vorrei provare a ritagliare questo intervento su tre nodi che ritengo cruciali, lasciando poi spazio alle domande e al dibattito.

- 1) L'odio è giuridicamente definibile?
- 2) Come impatta l'ipotesi di una eventuale regolazione dell'odio nell'ambito della libertà di espressione?
- 3) Quale ruolo per le piattaforme rispetto al tema dell'*hate speech*?

Ragionerò utilizzando esclusivamente gli strumenti del giurista e, nel mio caso, del costituzionalista, per cercare di conferire una logica tanto neutra quanto può esserlo quella scientifica e, in particolare, quella giuridica. Poiché, vedete, l'istruzione del giurista impone il ricorso a canoni precisi. Potremmo dire che, affinché un discorso risulti puramente giuridico, emancipandosi da tentazioni ideologizzanti o psicologiche, necessita proprio di essere ricondotto all'interno di categorie ben riconoscibili, poiché incorniciate da parametri certi, obiettivi, inopinabili, possibilmente fattuali. Allora, preliminarmente, ciò che dovrebbe premerci è ricondurre entro tale genere di parametri i termini grandemente ampi dell'odio e dell'intolleranza, così ampi da riuscire, sovente, assai sfuggenti. E laddove tali parametri non siano obiettivamente ravvisabili, occorrerà - per così dire *sine ira et studio* - tentare di costruirli.

I temi che sta affrontando questa Commissione sono scomodi e tutt'altro che teorici: evocano fatti della vita, dolorosi e, spesso, tragici, toccano corde sensibilissime. Pertanto, ogni giudizio,

commento, valutazione di tipo tecnico - come quella che può essere chiesta ad un costituzionalista - impone davvero - e lo dico ovviamente a me stessa - una grandissima cautela. Ma attenzione. I piani devono essere distinti. Una cosa è la dimensione di riprovazione morale ed etica, altra è la traduzione della galassia della categoria “odio”, “intolleranza”, “discriminazione” in comportamenti penalmente sanzionabili.

2. Definizione dell'oggetto.

Cos'è l'odio? Ad una Fiorentina il tema dell'odio non può essere estraneo... Basti solo pensare alle fratture dovute agli odi inveterati e quasi leggendari fra ghibellini e guelfi e, all'interno di questi ultimi, fra le fazioni dei bianchi e dei neri, che valgono, a titolo esemplare, dell'animosità storica di tanta civiltà italica trasversale e trans-storica.

In termini generali, l'odio - quale movente e non come crimine in sé - è il pregiudizio, il condizionamento mentale che ispira l'azione, ovvero l'abuso verbale o fisico ai danni di un individuo o di un gruppo a causa dell'appartenenza percepita di questi ad un determinato insieme, solitamente definito da razza, genere, religione, orientamento sessuale, disabilità, classe, etnia, nazionalità, età, identità di genere, affiliazione politica, o qualsiasi altro tratto distintivo che ne viene considerata una responsabilità paradossalmente intrinseca o innata. In questi termini, l'odio (o l'intolleranza), quale possibile valido movente, è certamente quello non episodico, ma rappresentativo di una comunità che riconosce in simboli specifici attraverso i quali i suoi appartenenti si ritengono superiori agli individui appartenenti ad un'altra.

Ma se questo è odio - a dimostrazione della scivolosità del tema - non poche sono comunque le scriminanti che in termini antropologici e politici si è cercato di trovarvi, di ordine funzionalista oppure scettico. Eco ad esempio (*Costruire il nemico*, La Nave di Teseo, 2020) ha scritto che l'individuazione di un nemico comune serve ad una società da elemento unificante essenziale, senza il quale non riuscirebbe mai a trovare il suo momento unitario radicale. E uno dei padri della sociologia come Simmel nel suo saggio risalente *Sul conflitto (Il conflitto della civiltà moderna*, prima edizione italiana del 1925) ha ritenuto l'odio l'atteggiamento negativo inossidabile, l'inimicizia pura irriducibile per cui, laddove non esista o si estingua l'elemento che funge da oggetto del contendere, piuttosto lo si inventa o re-inventa così da mantenere vivo un pretesto, quale che sia, per rinfocolare un conflitto inestinguibile. L'odio sarebbe dunque il sentimento senza se e senza ma, ma soprattutto senza perché; gratuito, come in fondo è anche l'amore. Il filosofo scozzese Hume, nel *Libro II del Trattato sulla natura umana* di (1739), dedicato alle passioni, contempla l'odio come una delle quattro passioni principali generate dalla simpatia, in senso greco: “*provare passioni con gli altri*”. La sola razionalità dell'agire umano sarebbe la “*trasmissione della simpatia*”, il continuo interscambio tra gli uomini di emozioni e sentimenti che li influenzano profondamente, fra i quali appunto anche l'odio. Perfino in alcune pagine

sulla democrazia di Kelsen, quelle definite proprio come le più scettiche, troviamo anche l'odio quale elemento irrinunciabile della discussione nell'incoerenza assiologica dei regimi politici stanchi, saturi, esausti.

L'odio è, dunque, elemento stesso della dimensione umana.

Ciò che interessa ai nostri fini non è il sentimento in sé, ma i limiti necessari in un ordinamento per non venirne consumato interamente. Politicamente, la risposta la troviamo ricorrendo ai classici del liberalismo: dalla tolleranza di Locke (Lettera sulla tolleranza, 1689) a quella di Rawls (*Teoria della giustizia*, 1971), alla soluzione Popperiana della *società aperta* (1945, in Italia edito nel 1973) «*al maggior numero possibile di idee e ideali differenti, e magari contrastanti. Ma, pena la sua autodissoluzione, non di tutti: la società aperta è chiusa solo agli intolleranti*».

Semmai va ravvisato un mutato livello di tolleranza, così come sono cambiati irreversibilmente i toni della discussione e del pregiudizio all'aumentare dei canali di espressione e manifestazione del pensiero.

3. Odio come fattispecie penale.

Giuridicamente, invece? Quali limiti vanno posti all'intolleranza manifesta? Quali limiti da porre al discorso d'odio? Occorre, innanzitutto, prestare attenzione ad un distinguo basilare: per risultare giuridicamente rilevante il fenomeno d'odio deve poter essere denotato obiettivamente come tale e non essere meramente il frutto di una connotazione parziale o prospettica. Chiarisco: per portare a conseguenze legali il riferimento non può essere ciò che è meramente percepito, e non reale, comprovato da eventi, fenomeni riconoscibili.

La legislazione positiva - e quella di cui ciclicamente si prospetta l'introduzione - fa riferimento, in particolare, all'istigazione o incitamento all'odio, duplicando, sovente, i piani di valutazione. Odio è istigazione all'odio... Con definizioni che diventano, evidentemente, tautologiche, così mostrando tutti i limiti espressivi e le intrinseche contraddizioni. Ne sono prova i tentativi (meritori) effettuati nell'ambito, ad esempio, del Consiglio d'Europa con la raccomandazione del Comitato dei Ministri del 30 ottobre 1997 sull'*hate speech* e del 21 marzo 2016, n. 15, contro il razzismo e l'intolleranza. Non è un caso che, in letteratura, si esprime l'idea che l'"odio" è più che altro un «contenitore definitorio che include forme espressive anche piuttosto diverse tra loro» (Spigno, *Discorsi d'odio*, 2018). E neppure è un caso che a livello europeo non si sia arrivati ad una definizione legislativa di odio, preferendosi rinviare agli Stati il compito di adottare le misure necessarie al fine di criminalizzare l'istigazione pubblica alla violenza e all'odio (vedi Decisione quadro del Consiglio del 28 novembre 2008) o, comunque, rinviando il compito a strumenti di *soft law* (vedi Codice di condotta per contrastare l'illecito incitamento all'odio on line adottato nel 2016 con lo scopo di contrastare il proliferare dell'incitamento all'odio razzista e xenofobo adottato con l'aiuto di Facebook, Microsoft, You Tube, Twitter). Strumenti,

questi, che pongono problemi non indifferenti in punto di inquadramento di fonti del diritto, specie su temi di questo calibro.

I Maestri del diritto penale insegnano che dobbiamo sempre muoverci nell'ambito dei principi di tassatività-determinatezza della fattispecie (Palazzo, *Il principio di determinatezza in diritto penale*, 1979), oltre che di materialità e di offensività del reato (Mantovani, *Diritto penale -parte generale*, 3°ed.,1992). Il rispetto del principio di determinatezza del diritto penale (art. 25, secondo comma Cost.) è stato elevato a principio supremo dell'ordinamento giuridico (Corte cost. 115/2018). Quando si invoca la determinatezza si sta, infatti, invocando uno dei capisaldi dello Stato di diritto, ovvero il principio per cui il legislatore, quando redige fattispecie penali, deve attenersi ad un canone di chiarezza lessicale e limpidezza strutturale operando una scrupolosa selezione terminologica, che riduca gli spazi di invasione creativa da parte del giudice. Emerge così plasticamente l'anima garantista del principio di determinatezza e la sua intima connessione con il principio di separazione dei poteri, fondante dello Stato di diritto.

Il problema si pone poi, del resto in concreto, ragionando proprio sul 604 bis codice penale che punisce - come sappiamo - propaganda e istigazione. La prima fattispecie è di difficile perimetrazione e la seconda è pure problematica nella sua valenza criminosa, in punto di effettiva probabilità della realizzazione della violenza. L'istigazione, insomma, rende la violenza di per sé possibile, ma quanto probabile?

Non aiuta l'esperienza comparata. La legge francese, precisa fino alla pedanteria nel disegno del sistema sanzionatorio e correttivo, è decisamente ambigua nella definizione concreta delle fattispecie (mi riferisco alla disciplina recata dalla *Loi contre le contenus haineux sur internet*, c.d. *Loi Avia*, adottata da ultimo con *Loi n° 2020-766* del 24 giugno 2020, dopo la sentenza del *Conseil Constitutionnel* n° 2020-801 del 18 giugno 2020). Quanto alla legge britannica sui reati d'istigazione all'odio, da rilevare è invece l'indifferenza per le intenzioni di chi li compie (*Section 66, Sentencing Act 2020*). I casi di crimine d'odio sono, infatti, definiti sulla base della percezione della vittima. Ma se l'odio diventa una questione di mera percezione e non di intenti riscontrabili ed è considerato irrilevante perfino il contesto, come difendersi da un uso strumentale, pretestuoso, o emulativo del diritto e d'altro canto dall'abuso del potere censorio?

Elementi concreti arrivano invece dalla giurisprudenza statunitense che è riuscita, da tempo, ad ancorare divieti di contenuti d'odio e di discriminazione solo laddove da esso si generino, in considerazione del contesto, un diretto, specifico e imminente pericolo (è il c.d. *emergency test*, il riferimento d'obbligo è al caso *Brandenburg v. Ohio* (1969).

È per questo che mi permetto di rivolgere a questa eminente Commissione il suggerimento di formulare in termini esatti la fattispecie che intendesse, se del caso, normare. Che sia una fattispecie chiara! inopinabile! Il miglior approccio per una corretta costruzione (o ricostruzione) della fattispecie

penale può avvenire solo attraverso un'operazione od una serie di operazioni ermeneutiche preventive. L'interpretazione cioè deve avvenire preliminarmente, in questa sede legislativa, e non in una successiva fase applicativa. Che si limiti la discrezionalità degli interpreti e, soprattutto, dei giudici nella determinazione *a posteriori* dei casi di reato.

Partendo dalla ricognizione dei nostri stessi preconcetti all'approccio ad una materia tanto complessa. Una complessità che può trovare proprio in questa Commissione il luogo più alto di sintesi e, quindi, di oggettivazione. Ogni precomprensione inerente ai temi della discriminazione, dell'intolleranza, dell'istigazione alla violenza, ogni vissuto, codici etico-comportamentali, contatti personali con l'odio diventa ricchezza nella determinazione, anzi, pre-determinazione della fattispecie criminosa contenente, coerente, completa.

L'odio stesso, in fondo, è un concetto ermeneutico in sé, da sviscerare ed esaminare guardandolo da ogni linea prospettica possibile, anche quella opposta, anche quella del nemico, dell'odiatore stesso. Così da comprendere, dall'altro lato, chi siamo noi o come siamo visti. L'odio come prisma e non come linea di demarcazione a sua volta discriminante, non come nuovo ed ulteriore oggetto di una meta-discriminazione. Giganteggia ancora nell'immaginario di tutti noi la tetra mole del Ministero dell'Amore che Orwell eresse nel suo *1984*. Quell'Autorità terribile che sceglieva chi e come odiare e puniva gli "psicoreati" e di fronte alla quale si stagiava il Ministero complementare della Verità.

Personalmente ho sempre ritenuto che la criminalizzazione dei reati di opinione, già abbondantemente previsti dall'ordinamento, non sia la strada più efficace da seguire. Pare, anzi, da monitorare la sempre maggiore (e preoccupante) pressione volta a conferire al diritto penale un ruolo che non gli è proprio: non *extrema ratio*, ma funzione pedagogica-etico-morale che dovrebbe, ragionevolmente, essere affidata a strumenti diversi, principalmente educativi, formativi, culturali, a partire dal linguaggio, nelle famiglie e nelle scuole fin dall'infanzia.

4) Un bilancio sulla libertà di espressione.

Un bilancio sulla libertà di espressione è passaggio ineludibile per contestualizzare il tema dell'odio.

Siamo lontani, certo, dall'assolutezza del modello culturale di libertà di espressione come quello del Primo Emendamento statunitense nel quale "*Il Congresso non potrà fare alcuna legge [...] che limiti la libertà di parola o di stampa*". Una previsione incentrata sul divieto per il legislatore federale di intervenire limitando la libertà di espressione e, dunque, costruita alla stregua di un limite nei confronti dei poteri pubblici. In Europa il modello è ovviamente diverso, anche perché il discorso d'odio trova la sua dimensione nel necessario bilanciamento tra libertà di espressione e dignità dell'individuo (divieto di discriminazioni art. 14 CEDU, art. 21 Carta diritti). La libertà di espressione non è assoluta

(art. 10 CEDU, art. 11 Carta diritti), ma incontra limitazioni previste *ex lege* e proporzionali rispetto al raggiungimento degli obiettivi.

Eppure “prendere sul serio” la libertà di espressione del pensiero non è cosa da poco neppure in Europa. Nel nostro ordinamento, e ripartendo dai fondamenti, le va, infatti, riconosciuto il rango di prima tra le libertà, appartenente al “nucleo duro” dei diritti inviolabili dell’uomo e, quale principio supremo dell’ordinamento, sottratto alla revisione costituzionale. Nei sistemi liberal-democratici la libertà di espressione è collocata alle basi dell’ordito costituzionale e su di essa affonda la rappresentanza politica, frutto della libera contesa delle idee: vale a dire la distribuzione del potere, le scelte di indirizzo politico, il controllo da parte dell’opinione pubblica. Ciò a differenza essenziale di quello che accade nelle forme di Stato autocratiche. In tale prospettiva si parla correttamente di *Stato totale* (secondo la formula di Forsthoff, *Der totale Staat*, 1933), dove invece prevale la conformazione incondizionata del tutto all’ideologia dominante, sia essa d’impronta laica, sia essa di matrice religiosa. La stretta connessione tra libertà di espressione e liberaldemocrazia è stata ampiamente argomentata dagli scrittori politici e dalla dottrina giuridica. Libertà di pensiero è dunque, alla prova di resistenza, libertà di dissenso e di critica, diritto alla differenziazione rispetto a tutto ciò che è ancorato alla trama dei valori costituiti o, comunque, riferibile agli assetti dominanti del potere inteso in senso lato (Pacillo, *I delitti contro le confessioni religiose dopo la legge 24 febbraio 2006*, n. 85, 2007). E così reca con sé un altro elemento essenziale della liberaldemocrazia: vale a dire la garanzia delle minoranze e del dissenso, tanto che - si dice - un campanello d’allarme deve, invece, suonare quando il dissenso tace (Chiodi, *Tacito dissenso*, Torino, 1990).

Eppure nonostante la nostra Costituzione contenga come unico limite espresso alla libertà di espressione il buon costume, la Corte costituzionale ha enucleato dal testo costituzionale una serie di situazioni giuridiche meritevoli di tutela idonee a sorreggere limitazioni alla libertà di manifestazione del pensiero. In particolare, nella prospettiva di una tutela sistemica dei diritti, la libertà di manifestazione del pensiero può andare incontro ad una limitazione nel caso in cui determini una lesione immediata e definitiva di valori costituzionalmente garantiti. Nel perseguire tale tutela sistemica dei diritti la Corte costituzionale ha individuato quali beni costituzionalmente protetti idonei a sorreggere limitazioni della libertà di espressione: il diritto all’onore, al decoro, alla rispettabilità, alla riservatezza, alla reputazione (sentt. 122 del 1970, 38 del 1973, 86 del 1974), la tutela dei minori (sentt. nn. 9 e 25 del 1965 e 16 del 1981), il sentimento religioso (sent. n. 188 del 1975), la difesa della patria (sentt. nn. 16 del 1973 e 31 del 1982), l’ordine pubblico (sentt. nn. 120 del 1957, 19 del 1962, 25 del 1965, 87 del 1966, 199 del 1972, 15 del 1973, 210 del 1976, 138 del 1985, 112 del 1993), la sicurezza pubblica (sentt. 1 del 1956, 65 del 1970), l’esigenza di prevenzione dei reati (sentt. n. 1 del 1956, 120 e 121 del 1957, 38 del 1961), la tranquillità pubblica (sentt. nn. 33, 120, 121 del 1957), la quiete pubblica (sent. n. 38 del 1961), l’esigenza di impedire la ricostituzione del partito fascista (sent. 1 del

1957 e 74 del 1958), il metodo democratico (sent. 87 del 1966), la tutela dello Stato con riguardo alla tutela dell'esistenza, dell'integrità, dell'unità, della indipendenza, della pace e della difesa militare e civile dello Stato (sent. n. 25 del 1965), la giustizia (sentt. nn. 25 del 1965, 18 del 1966, 1 e 18 del 1981, 196 del 1987), il prestigio dell'ordine giudiziario (sent. n. 100 del 1981), il prestigio del Governo, dell'ordine giudiziario e delle forze armate (sent. n. 20 del 1974), l'ordine economico (sent. n. 87 del 1966) e l'economia pubblica (sentt. nn. 123 del 1976 e 73 del 1983).

I beni giuridici idonei a limitare la libertà di manifestazione del pensiero sono, dunque, innumerevoli. È possibile affermare che la Corte costituzionale tramite la propria giurisprudenza abbia ampliato di molto le maglie delle limitazioni della libertà di espressione. Il che ingenera più di qualche perplessità sulla ricostruzione operata a livello giurisprudenziale, soprattutto alla luce dell'affievolimento della libertà di espressione che una simile giurisprudenza appare suscettibile di determinare, come già scriveva nel 1965 un giurista del calibro di Crisafulli (*In tema di limiti alla cronaca giudiziaria*, in *Giur. cost.*, 1965). Appare, dunque, quantomai opportuno continuare ad operare un severo scrutinio su tali restrizioni alla libertà di espressione, poiché la libertà - per dirla con Calamandrei - non può essere mai data per scontata.

5. L'odio in rete e il ruolo delle piattaforme.

Riprendiamo allora tutto questo discorso e portiamolo direttamente alla questione universale del presente, cioè della rete, dove il tema dell'odio esplode. Nella situazione di anarchia regolatoria su ciò che è odio, fanno il loro ingresso le piattaforme con uno spettro di azioni che poggia sull'evidente asimmetria contrattuale e che va dalla rimozione successiva, in pratica sanzionatoria, dei contenuti fino alla misura di prevenzione della chiusura degli *account* dei recidivi o di *haters* algoritmicamente presunti.

La libertà costituzione di espressione in rete è oggi definita dalle piattaforme che valutano i confini eticamente e giuridicamente accettabile, decidendo cosa si può dire e cosa non si può dire nell'agorà digitale. Il tema è molto serio e deve partire da un esame di realtà. Il coinvolgimento delle piattaforme nella regolazione della libertà di espressione è divenuta un'esigenza da cui sembra difficile prescindere. Se per alcuni versi esse rappresentano i canali più diretti ed efficaci, per altri, si dimostrano inadeguate a concepire una strategia in linea con i parametri costituzionali delle moderne democrazie. È la conseguenza dello stesso *design* istituzionale della Rete, concepito come spazio fluido transnazionale che rende, al momento, questi attori forse gli unici soggetti in grado di regolare efficacemente i contenuti digitali e di dare forma alla realtà dell'espressione online. Così facendo però la libertà di espressione *online*, finisce per dipendere dalle piattaforme, con ruolo alternativo allo Stato, essendo ormai appurato che Facebook, Youtube, Twitter, ecc...non sono (più) un semplice servizio di intermediazione di informazioni, ma sono diventati essi stessi strutture di *governance* del sistema. Con

la conseguenza che ogni politica di regolazione di Internet portata avanti dalle piattaforme comporta necessariamente una riduzione dell'area dei mezzi di reazione in mano allo Stato.

L'approccio delle piattaforme al tema dell'*hate speech* evidenzia ad oggi non poche fragilità nel processo di moderazione della libertà di espressione, tra cui: a) l'assenza di chiari e conoscibili parametri di valutazione negli *standard di community*; b) l'assenza di un effettivo *legal reasoning* e di *procedural safeguards* che porta ad una sommarietà e trascuratezza del procedimento; c) l'approssimazione delle esemplificazioni degli *standard di community*; d) l'utilizzo dell'algoritmo.

Sinteticamente, sul primo profilo, va quantomeno segnalato che gli *standard di community* sono predisposti dai *content policy teams* e riguardano le politiche di moderazione dei contenuti alla luce dei criteri generali di condotta che sono in essi contenuti. Tali regole stabilite dalle piattaforme rivestono una funzione di vera e propria norma che incide sui diritti fondamentali personalissimi, anche se nella loro stesura non è coinvolto nessun attore pubblico. Il loro contenuto è spesso vago e lo standard di protezione della libertà di espressione altamente discrezionale. Facebook, ad esempio, rimette la sua politica di moderazione in massima parte a *guidelines* interne il cui contenuto e il cui processo di formazione non è sempre lineare. C'è inoltre una mancanza di trasparenza e di *accountability* sulle procedure di rimozione dei post, fattori questi che portano ad una lesione non solo della singola manifestazione del pensiero, ma anche degli stessi precipitati giuridici alla base del principio di uguaglianza.

Sul secondo profilo, ovvero l'assenza di un effettivo *legal reasoning* e di *procedural safeguards*, va evidenziato che la fase di moderazione dei potenziali *hate speech* inizia con una segnalazione, che può provenire o dagli utenti o dalle risultanze algoritmiche. I vari *team* di selezionatori (*content policy teams*) o l'algoritmo sono incaricati di valutare in pochi secondi l'illiceità di un contenuto, a prescindere dal tipo di illecito. E qui si pone il primo problema, perché se per violazioni più macroscopiche, come la pedopornografia e il terrorismo, la decisione è più immediata ed intuitiva, per altre fattispecie, come sono appunto i fragili contorni dell'*hate speech*, si richiederebbe competenza giuridica e un maggior tempo di riflessione. Per far parte dei *content policy team* di Facebook, che dovrebbero essere composti da presunti "esperti" di temi quali il terrorismo, l'incitamento all'odio e la pedopornografia, non è richiesta alcuna specifica qualificazione giuridica o anche solo specializzazione. Quando il (presunto) illecito è rimosso, non sempre la compiuta rimozione del contenuto e la sua motivazione viene comunicata all'utente. Anche questo rappresenta un profilo problematico. Allo stato è data registrare un'opaca procedura di contestazione dei post rimossi: nella notifica che comunica la rimozione del contenuto è presente l'opzione di ricorrere chiedendo una seconda revisione, senza però la possibilità per l'utente di fornire una accurata argomentazione in difesa del post in esame. Un secondo revisore deciderà *sic et simpliciter* entro 24h se confermare la rimozione. Manca quindi un vero e proprio contraddittorio.

Un terzo profilo problematico riguarda l'individuazione concreta delle condotte d'odio. Ogni piattaforma ha proceduto a sé, talvolta descrivendo e altre volte esemplificando le espressioni potenzialmente riconducibili alla fattispecie dell'*hate speech*. Prendiamo ad esempio il caso di Twitter. Twitter rimane generalmente molto vago al proposito, considerando proibiti: «i contenuti che augurano, promuovono o auspicano morte, lesioni corporali gravi e durature o serie malattie a un'intera categoria protetta e/o a persone che potrebbero farne parte»; i «contenuti umilianti o insultanti», oppure discorsi «disumanizzanti». Vengono inseriti una dozzina di “casi di scuola”, esemplificativi ma non esaustivi: «ti uccido», «ti dovrebbe venire il cancro», «tutti i [membri di un gruppo religioso] sono terroristi», ecc., si tratta di espressioni che certamente eccedono sotto ogni profilo il limite della libertà di parola. Per Twitter vi sarebbe dunque un duplice criterio di valutazione: uno che riguarda il destinatario dell'espressione (gruppi protetti o gruppi che hanno subito genocidi) e uno, alternativo o cumulativo, che riguarda il contenuto: violenza o augurio, speranza o invocazione di un serio danno a una persona o a un gruppo di persone. Un'analisi più approfondita evidenzia, tuttavia, la presenza di fattispecie la cui offensività è ambigua o difficilmente valutabile e, comunque, non sembra contenere i presupposti necessari per integrare un vero e proprio discorso d'odio. È il caso degli standard di community di Facebook, sebbene la sua esemplificazione sia molto più ricca e dettagliata rispetto a Twitter. Facebook divide le tipologie dei contenuti vietati in tre livelli (non è chiaro se a questi livelli corrisponda anche un *climax* di offensività o siano distinti solo per la natura delle offese; né viene graduata la risposta a seconda della gravità dell'odio espresso dalla dichiarazione). Tuttavia, alcuni degli esempi lasciano perplessi per il loro carattere di per sé innocuo. Non sarebbe infatti possibile neppure appellare qualcuno, che appartiene ad una comunità minoritaria, con i seguenti aggettivi: «bugiardo», «ignorante», «folle», «incapace», «inutile». Ma vi è di più: secondo il livello 3 (“Contenuti rivolti a persone o a gruppi di persone”) non si possono esprimere giudizi sull' «esclusione esplicita» di persone o gruppi di persone: in particolare utilizzando i verbi «espellere» o «non consentire». Dovendosi presumere che sarebbe, ad esempio, contraria alle regole dell'azienda una affermazione del tipo: “Tutti i clandestini devono essere espulsi”. Non si può, inoltre, parlare di «negazione della possibilità di accedere a servizi sociali». Dunque, non si tratta solo di insulti, ma anche di argomentazioni che esprimono una propria visione su temi di rilevanza politica-istituzionale. Ma in una conversazione su tali temi come si può distinguere un crimine d'odio da una posizione, per quanto contestabile, di chi intendesse ad esempio escludere *tout court* gli immigrati dai servizi sociali? E come valutare la posizione di chi sostenesse che solo gli stranieri regolari siano da ammettere al godimento di prestazioni socio-sanitarie e non invece gli irregolari? Che ruolo deve avere poi il contesto in cui la frase è espressa (ad esempio in un clima scherzoso tra amici)? E come viene valutato?

Un ultimo aspetto consiste nella imperfezione delle decisioni algoritmiche. Sono milioni ogni mese i contenuti illeciti rimossi grazie all'uso degli algoritmi. L'uso della selezione algoritmica è

diventata una variabile essenziale ad esempio per Twitter e Google, i quali rimettono allo strumento buona parte delle segnalazioni. Il tentativo di accompagnare le decisioni che riguardano interessi generali o diritti fondamentali attraverso un metodo matematico de-umanizzato è finalizzato, tra i vari aspetti, a due obiettivi primari: ridurre al massimo il margine di errore e garantire una totale imparzialità e trasparenza sulla limitazione della libertà di espressione. Gli aspetti problematici dell'algoritmo riguardano proprio la sua mancanza di accuratezza con riferimento al contesto e gli intrinseci *bias* che lo caratterizzano. Questo deriva da un fatto ormai noto, che cioè alla base dell'algoritmo c'è sempre una conoscenza condizionata da un'attività umana di programmazione. Il potere di orientare l'algoritmo sta nelle mani di chi detiene il codice dell'algoritmo. Chi possiede il codice lo utilizza nei confronti di altri soggetti e non c'è nessuna simmetria in questa operazione. Ma, soprattutto, gli algoritmi sono incapaci di interpretare il contesto: fino a che punto gli algoritmi (anche quelli di apprendimento automatico) possono leggere le nostre vere intenzioni, lo "stile" con cui noi condividiamo un commento o una foto? Quanto possano leggere i costrutti psicologici di ciascun utente e inserirli nello specifico contesto di riferimento, formato anch'esso da altrettanti costrutti psicologici dei diversi utenti?

Su tutti questi temi, l'approvazione recente del *Digital Service Act* da parte del Parlamento europeo rappresenta certamente un significativo passo in avanti quanto a misure per contrastare il diffondersi di contenuti illegali online. Non è irrilevante che si siano pensate garanzie efficaci per gli utenti, compresa la possibilità di contestare le decisioni di moderazione dei contenuti delle piattaforme e introdotte misure di trasparenza per le piattaforme medesime anche sugli algoritmi utilizzati. Colpisce però, proprio alla luce delle riflessioni sui problemi definitivi proposti, che in tutto il testo non vi sia mai il riferimento all'odio come definizione di fattispecie, ma solo generici richiami nella Relazione al testo (pag. 4 e 6 peraltro con riferimento al sopra citato Codice di condotta sull'Hate Speech online della Commissione Europea) e nei Considerando (n. 12, n. 57, n. 69). Il che per un testo che ha l'ambizione di diventare la pietra miliare europea sul contrasto all'*hate speech* in rete non è lacuna da poco. Anche di questo profilo, in una prospettiva regolatoria italiana, credo dovrebbe tenersi conto.

Grazie per la vostra attenzione.