SINTESI

Studio per la commissione JURI

Parlamento europeo

L'impatto degli algoritmi per filtrare o moderare i contenuti online
Gli "Upload filter"[1]

SINTESI

Lo studio analizza il filtraggio automatico dei contenuti online e guarda ai filtri automatici come un aspetto della moderazione dei materiali generati dagli utenti. Presenta le tecnologie di filtraggio attualmente utilizzate per i diversi tipi di media, quali testi, immagini o video. Lo studio esamina inoltre le principali criticità nell'ambito dell'attuale quadro giuridico e presenta proposte di regolamentazione nel contesto di una futura **legge dell'UE sui servizi digitali**.

Contesto

Lo studio analizza il filtraggio automatico dei contenuti online e guarda ai filtri automatici come un aspetto della moderazione dei materiali generati dagli utenti. Presenta le tecnologie di filtraggio attualmente utilizzate per i diversi tipi di media, quali testi, immagini o video. Lo studio esamina inoltre le principali criticità nell'ambito dell'attuale quadro giuridico e presenta proposte di regolamentazione nel contesto di una futura **legge dell'UE sui servizi digitali**.

La **direttiva sul commercio elettronico**, giunta al suo ventesimo anno di vita, ha svolto un ruolo decisivo nello sviluppo dell'economia digitale e dell'ambiente dell'informazione online, ma continua a essere applicata in un contesto tecnologico, economico e sociale completamente mutato: le società di Internet sono diventate attori globali dotati di ingenti risorse finanziarie e tecnologiche e l'accesso all'informazione e all'interazione sociale si è spostato online.

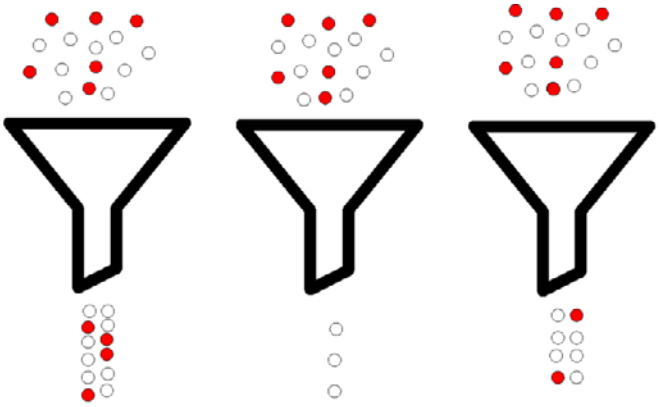
La nuova **legge sui servizi digitali** dovrebbe occuparsi del contesto tecnologico, economico e sociale. Una questione fondamentale riguarda la regolamentazione dei servizi digitali, in particolare quelli che rientrano nella categoria delle piattaforme online, vale a dire i servizi digitali il cui scopo è facilitare l'interazione digitale tra utenti (imprese o singoli). In particolare, le piattaforme per i contenuti generati dagli utenti consentono a questi ultimi di esprimersi, creare, trasmettere o accedere a informazioni e creazioni culturali e di avere interazioni sociali, ma offrono anche opportunità di comportamenti dannosi: inciviltà e aggressività negli scambi individuali, disinformazione nella sfera pubblica, settarismo e polarizzazione della politica, nonché illegalità, sfruttamento e manipolazione.

[1] Studio completo in inglese : https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU(2020)657101_EN.pdf

Dipartimento tematico "Diritti dei cittadini e affari costituzionali"
Direzione generale delle Politiche interne
PE 657.101 – Settembre 2020

Per prevenire comportamenti illeciti e dannosi online è necessaria la *moderazione*, vale a dire la governance attiva delle piattaforme intesa a garantire, nella misura in cui è ragionevolmente possibile, che gli utenti interagiscano in maniera produttiva, pro-sociale e lecita. La moderazione deve facilitare la cooperazione e prevenire gli abusi: in sua assenza le comunità online tendono a diventare disfunzionali e a rimanere vittime di spammer, ecc.

Lo studio analizza un aspetto fondamentale della moderazione attuale, vale a dire il filtraggio automatizzato inteso a classificare, e quindi a declassare o escludere, i materiali generati dagli utenti. I filtri automatici sono necessari per monitorare l'enorme quantità di materiale caricato online e individuare contenuti (potenzialmente) illeciti e abusivi. Tuttavia, il loro utilizzo presenta rischi, in quanto può portare all'esclusione di contenuti validi e può incidere sulla libertà di espressione, sull'accesso all'informazione e sul dialogo democratico.



La direttiva sul commercio elettronico adotta un approccio "non interventista" alla moderazione, in particolare per quanto riguarda il filtraggio: protegge i fornitori dalla responsabilità per i contenuti illeciti generati dagli utenti, mentre vieta agli Stati di imporre ai fornitori obblighi generali di monitoraggio. Negli ultimi anni questo approccio è stato messo in discussione, in quanto sono state sviluppate tecnologie per il filtraggio.

Scopo

L'obiettivo dello studio è fornire un'analisi approfondita delle questioni tecnologiche e giuridiche relative alla regolamentazione del filtraggio online. Presenta le tecnologie di filtraggio attualmente utilizzate per i diversi tipi di media, quali testi, immagini o video, esaminandone i punti di forza e le vulnerabilità:

- la ricerca di metadati, l'hashing e il rilevamento delle impronte digitali sono utilizzati per identificare in modo affidabile le copie di opere digitali note;
- il meccanismo di "lista nera" è impiegato per reperire espressioni non desiderate;
- le tecniche avanzate per l'elaborazione del linguaggio naturale sono impiegate per analizzare il significato e il contesto;
- numerose tecniche spesso basate sull'IA sono utilizzate per identificare immagini indesiderate oppure combinazioni di testo e immagini.

Lo studio esamina l'accuratezza dei sistemi di filtraggio:

- essi si basano su metodi probabilistici e, pertanto, non è possibile evitare tutti gli errori;
- l'identificazione di una risposta che può essere considerata corretta pone problemi, dato che la "verità di fondo" è fornita da valutazioni umane.

Vista la fallibilità e la soggettività dei filtri automatici, il loro funzionamento dovrebbe essere controllabile. Lo studio analizza i metodi per fornire trasparenza e possibilità di ricorso:

- per garantire la trasparenza, è necessario che le persone interessate e la società in generale siano informati del processo di filtraggio;

- occorrono meccanismi di appello e ricorso per consentire agli utenti di presentare contestazioni e rimediare agli errori.

Lo studio prende altresì in esame l'accessibilità e i costi delle tecnologie di filtraggio. Esse sono utilizzate principalmente dai grandi operatori, che spesso sviluppano potenti sistemi interni, mentre alcune soluzioni sono accessibili anche alle imprese più piccole.

Lo studio infine analizza la regolamentazione del filtraggio. Introduce in primo luogo alcune premesse che sarebbe opportuno prendere in considerazione:

- il filtraggio automatico non dovrebbe essere scoraggiato in quanto è una componente essenziale di un'efficiente moderazione online;
- il filtraggio è fallibile, anche se sviluppato e attuato in buona fede per contrastare i contenuti abusivi;
- il filtraggio pro-sociale ammissibile non si limita ai contenuti illeciti; può legittimamente riguardare qualsiasi tipo di contenuto sconveniente nel contesto di una determinata comunità online;
- sebbene il filtraggio giustificato non si limiti ai materiali illeciti, non dovrebbe basarsi su scelte incostanti o arbitrarie dei proprietari e dei moderatori delle piattaforme;
- non esistono norme applicabili meccanicamente in grado di determinare con certezza cosa può essere accettabile o inaccettabile dal punto di vista giuridico in una piattaforma.
- per valutare il comportamento dei fornitori è possibile avvalersi di criteri di dovuta diligenza/ragionevolezza;
- una regolamentazione inadeguata può indurre a un filtraggio eccessivo o insufficiente che potrebbe non consentire di accedere a materiale valido oppure farebbe rimanere online contenuti dannosi;
- vi è incertezza a riguardo di quali obblighi di monitoraggio dei contenuti siano vietati dal diritto dell'UE.

Lo studio, sulla base di tali premesse, illustra diverse opzioni strategiche:

- sarebbe opportuno prendere in considerazione un aggiornamento dei principi generali sull'immunità stabiliti dalla direttiva sul commercio elettronico. Si potrebbe inoltre precisare nuovamente in che misura il diritto dell'UE debba proteggere i fornitori dall'imposizione di obblighi giuridici relativi al monitoraggio, alla rimozione o al blocco dei contenuti.
- Andrebbe inoltre chiarito che l'impegno alla moderazione e, in particolare, all'applicazione del filtraggio di contenuti illegali o abusivi non dovrebbe pregiudicare le immunità o gli altri vantaggi concessi ai fornitori.
- Dovrebbero essere introdotti mezzi di ricorso procedurali per le rimozioni online, in modo che coloro che hanno caricato contenuti esclusi dai filtri siano informati delle decisioni, ricevano spiegazioni e possano contestare tali decisioni ottenendo risposte da umani.
- Le autorità pubbliche dovrebbero affrontare il problema del filtraggio online attraverso regolamenti specifici e decisioni su casi controversi. Potrebbero essere coordinate dagli organismi esistenti dell'UE o da un'autorità di nuova istituzione.
- È opportuno fornire sostegno alle piccole e medie imprese, che dovrebbero essere ritenute responsabili solo in caso di mancata adozione di misure (compreso il filtraggio) a loro accessibili, da un punto di vista sia tecnologico che economico.

- Dovrebbe essere sviluppato un approccio a livello dell'UE in materia di regolamentazione del filtraggio, che tenga conto al contempo delle diversità nazionali nella valutazione della liceità dei contenuti online.
- È necessario un ampio dibattito sulla moderazione e, in particolare, sul filtraggio che coinvolga non solo le autorità politiche e amministrative, ma anche la società civile e il mondo accademico.

Autori esterni: Prof. Giovanni Sartor, Istituto universitario europeo di Firenze.
Coautori: Prof. Giovanni Sartor e Dott. Andrea Loreggia, sotto la supervisione del Prof. Sartor.

Amministratore della ricerca responsabile: Udo BUX Assistente redazionale: Monika LAZARUK
Contatto: poldep-citizens@europarl.europa.eu

Il documento è disponibile su Internet all'indirizzo: http://www.europarl.europa.eu/committees/it/supporting-analyses-search.html

PE 657.101

Stampa ISBN 978-92-846-7187-8| doi: 10.2861/89350| QA-02-20-825-IT-C
PDF ISBN 978-92-846-7185-4| doi: 10.2861/87955| QA-02-20-825-IT-N