

Papers di DIRITTO EUROPEO

www.papersdidirittoeuropeo.eu
ISSN 2038-0461

2026, n. 1

Papers di diritto europeo

www.papersdidirittoeuropeo.eu

DIRETTORE RESPONSABILE

Maria Caterina Baruffi (Ordinario di Diritto internazionale, Università di Bergamo).

COMITATO DI DIREZIONE

Francesco Bestagno (Ordinario di Diritto dell'Unione europea, Università Cattolica del Sacro Cuore di Milano; Giudice del Tribunale dell'Unione europea); **Andrea Biondi** (Avvocato generale alla Corte di giustizia; Professor of European Law e Director of the Centre of European Law, King's College London); **Fausto Pocar** (Professore emerito, Università di Milano); **Lucia Serena Rossi** (Ordinario di Diritto dell'Unione europea, "Alma Mater Studiorum" Università di Bologna).

COMITATO SCIENTIFICO

Adelina Adinolfi (Ordinario di Diritto dell'Unione europea, Università di Firenze); **Elisabetta Bani** (Ordinario di Diritto dell'economia, Università di Bergamo); **Matteo Borzaga** (Ordinario di Diritto del lavoro, Università di Trento); **Susanna Cafaro** (Ordinario di Diritto dell'Unione europea, Università del Salento); **Laura Calafà** (Ordinario di Diritto del lavoro, Università di Verona); **Javier Carrascosa González** (Catedrático de Derecho Internacional Privado, Universidad de Murcia); **Luigi Daniele** (Ordinario di Diritto dell'Unione europea, Università di Roma "Tor Vergata"); **Angela Di Stasi** (Ordinario di Diritto internazionale, Università di Salerno); **Davide Diverio** (Ordinario di Diritto dell'Unione europea, Università di Milano); **Franco Ferrari** (Professor of Law e Director of the Center for Transnational Litigation, Arbitration, and Commercial Law, New York University); **Costanza Honorati** (Ordinario di Diritto dell'Unione europea, Università di Milano-Bicocca); **Paola Mori** (Ordinario f.r. di Diritto dell'Unione europea, Università "Magna Graecia" di Catanzaro); **Matteo Ortino** (Associato di Diritto dell'economia, Università di Verona); **Carmela Panella** (Ordinario f.r. di Diritto internazionale, Università di Messina); **Lorenzo Schiano di Pepe** (Ordinario di Diritto dell'Unione europea, Università di Genova); **Alessandra Silveira** (Profesora Asociada e Diretora do Centro de Estudos em Direito da União Europeia, Universidade do Minho); **Eleanor Spaventa** (Ordinario di Diritto dell'Unione europea, Università "Bocconi" di Milano); **Stefano Troiano** (Ordinario di Diritto privato, Università di Verona); **Michele Vellano** (Ordinario di Diritto dell'Unione europea, Università di Torino).
Segretario: **Caterina Fratea** (Associato di Diritto dell'Unione europea, Università di Verona).

COMITATO DEI REVISORI

Stefano Amadeo (Ordinario di Diritto dell'Unione europea, Università di Trieste); **Bruno Barel** (Associato f.r. di Diritto dell'Unione europea, Università di Padova); **Silvia Borelli** (Associato di Diritto del lavoro, Università di Ferrara); **Laura Carpaneto** (Ordinario di Diritto dell'Unione europea, Università di Genova); **Marina Castellaneta** (Ordinario di Diritto internazionale, Università di Bari "Aldo Moro"); **Federico Casolari** (Ordinario di Diritto dell'Unione europea, **"**Alma Mater Studiorum" Università di Bologna); **Gianluca Contaldi** (Ordinario di Diritto dell'Unione europea, Università di Macerata); **Matteo De Poli** (Ordinario di Diritto dell'economia, Università di Padova); **Giacomo di Federico** (Ordinario di Diritto dell'Unione europea, "Alma Mater Studiorum" Università di Bologna); **Fabio Ferraro** (Ordinario di Diritto dell'Unione europea, Università di Napoli "Federico II"); **Daniele Gallo** (Ordinario di Diritto dell'Unione europea, LUISS Guido Carli); **Pietro Manzini** (Ordinario di Diritto dell'Unione europea, "Alma Mater Studiorum" Università di Bologna); **Simone Marinai** (Associato di Diritto dell'Unione europea, Università di Pisa); **Silvia Marino** (Ordinario di Diritto dell'Unione europea, Università dell'Insubria); **Emanuela Pistoia** (Ordinario di Diritto dell'Unione europea, Università di Teramo); **Francesca Ragno** (Ordinario di Diritto internazionale, "Alma Mater Studiorum" Università di Bologna); **Carola Ricci** (Associato di Diritto internazionale, Università di Pavia); **Giulia Rossolillo** (Ordinario di Diritto dell'Unione europea, Università di Pavia); **Vincenzo Salvatore** (Ordinario di Diritto dell'Unione europea, Università dell'Insubria); **Andrea Santini** (Ordinario di Diritto dell'Unione europea, Università Cattolica del Sacro Cuore di Milano); **Cristina Schepisi** (Ordinario di Diritto dell'Unione europea, Università di Roma "Tor Vergata"); **Martin Schmidt-Kessel** (Lehrstuhl für Deutsches und Europäisches Verbraucherrecht und Privatrecht sowie Rechtsvergleichung, Universität Bayreuth); **Chiara Enrica Tuo** (Ordinario di Diritto dell'Unione europea, Università di Genova).

COMITATO EDITORIALE

Diletta Danieli (Associato di Diritto dell'Unione europea, Università di Verona); **Teresa Maria Moschetta** (Associato di Diritto dell'Unione europea, Università di Roma Tre); **Rossana Palladino** (Associato di Diritto dell'Unione europea, Università di Salerno); **Cinzia Peraro** (Associato di Diritto dell'Unione europea, Università di Bergamo); **Federica Persano** (Ricercatore di Diritto internazionale, Università di Bergamo); **Angela Maria Romito** (Associato di Diritto dell'Unione europea, Università di Bari "Aldo Moro"); **Sandra Winkler** (Straordinario di Diritto della famiglia, Università di Rijeka).

Responsabile di redazione: **Isolde Quadranti** (Documentalista, Centro di documentazione europea, Università di Verona).

I contributi sono sottoposti ad un procedimento di revisione tra pari a doppio cieco (*double-blind peer review*).
Non sono sottoposti a referaggio esclusivamente i contributi di professori emeriti, di professori ordinari in quiescenza e di giudici di giurisdizioni superiori e internazionali.

Papers di diritto europeo
ISSN 2038-0461
Registrazione al Tribunale di Verona n. 1875 del 22 luglio 2010

Papers di diritto europeo
www.papersdidirittoeuropeo.eu

Fascicolo 2026, n. 1

INDICE

Ruggiero Cafari Panico
In attesa che la Corte di giustizia si pronunci sulle concessioni di piccole derivazioni (riflessioni in merito alle conclusioni dell'Avvocato generale Campos Sánchez-Bordona e alla loro possibile ricaduta sulle grandi derivazioni) 1

Donatella del Vescovo
La frontiera interna della cittadinanza: limiti dell'Unione europea alla naturalizzazione basata sull'investimento 21

Francesca Ferrari, Massimiliano Bina
Dalla bacheca al gatekeeper*: piattaforme digitali e moderazione* 49

Francesca Martines
Effettività del controllo giurisdizionale e resistenza istituzionale: la dinamica ripetitiva nel contenzioso sul Sahara occidentale 69

Matteo Ortino
Il Digital Services Act *e la produzione di conoscenza: le specificità rispetto ai tradizionali regimi di trasparenza* 91

Agostina Latino
Violenza algoritmica di genere, tecnologie predittive e diritto internazionale 123

Viviana Sachetti
La sostenibilità ambientale quale parametro di bilanciamento nella politica energetica dell'Unione europea 175

Francesco Spera
Alla ricerca dell'autonomia strategica attraverso atti atipici e informali: il soft law *nelle relazioni esterne dell'UE* 197

Giacomo Furlanetto
Il Digital Services Act *alla prova del caso «Mia Moglie»: note sociologico-giuridiche* 219

Dalla bacheca al *gatekeeper*: piattaforme digitali e moderazione

Francesca Ferrari* e Massimiliano Bina**

SOMMARIO: 1. Introduzione. – 2. La metamorfosi giuridica delle piattaforme digitali. – 3. La piattaforma digitale quale *gatekeeper*. Lo stato della normativa UE. – 4. La moderazione come funzione costituzionale privata. – 5. L'intelligenza artificiale: tra efficienza tecnica e limiti epistemici. – 6. Moderazione ibrida e ruolo umano. Il *Digital Services Act*. – 7. Etica, diritti fondamentali e implicazioni sistemiche. – 8. Conclusioni.

1. Introduzione.

La moderazione dei contenuti online costituisce una delle sfide centrali del diritto dell'infosfera contemporanea.

L'espressione «infosfera», introdotta da Luciano Floridi per indicare l'insieme degli ambienti informazionali nei quali si svolgono le interazioni umane e artificiali[1], permette di cogliere come la regolazione dei contenuti non sia un fenomeno meramente tecnico, ma un processo che incide sulle strutture epistemiche e sociali delle società digitali[2].

A differenza dei tradizionali mezzi di comunicazione, le piattaforme digitali esercitano un potere inedito, sia in termini di ampiezza sia di intensità, nell'organizzare il discorso pubblico: determinano ciò che può essere detto, ciò che può essere visto e ciò che viene reso visibile a specifici gruppi di utenti attraverso algoritmi di raccomandazione e filtri automatici.

La moderazione – tanto più quando affidata a sistemi di intelligenza artificiale (IA) – si pone, dunque, al crocevia tra tecnologia, diritto e teoria politica, poiché influenza la definizione stessa della libertà di espressione e del pluralismo informativo.

Il problema, come osserva Floridi, non riguarda solamente l'accuratezza dei sistemi automatizzati, ma la capacità di tali sistemi di contribuire in modo legittimo alla costruzione di ambienti informazionali giusti e affidabili[3].

Questo saggio analizza la trasformazione delle piattaforme digitali da intermediari passivi a regolatori pro-attivi, il ruolo dell'intelligenza artificiale nella moderazione dei

* Professoressa associata di Diritto processuale civile, Università degli Studi dell'Insubria, cui sono da attribuirsi i paragrafi 1, 2, 4, 6, 8.

** Avvocato, Foro di Varese, cui sono da attribuirsi i paragrafi 3, 5, 7.

[1] L. FLORIDI, *The Philosophy of Information*, Oxford, 2011, pp. 6-10.

[2] L. FLORIDI, *La differenza fondamentale*, Milano, 2025, p. 51.

[3] L. FLORIDI, *The Logic of Information*, Oxford, 2019, pp. 157-165.

Papers di diritto europeo, fascicolo 2026, n. 1, pp. 49-68 ISSN 2038-0461
Accettato a seguito di referaggio: 16.02.2026 www.papersdidirittoeuropeo.eu

contenuti e le implicazioni per i diritti fondamentali, nel quadro normativo e giurisprudenziale – europeo, convenzionale e statunitense. L'approccio è discorsivo, sistematico e orientato alla comprensione delle tensioni profonde tra efficienza tecnica, responsabilità privata e garanzie costituzionali.

2. La metamorfosi giuridica delle piattaforme digitali.

La qualificazione giuridica delle piattaforme digitali nasce, nell'ordinamento europeo, dalla direttiva *e-commerce* del 2000[4], che distingue tra diverse categorie di intermediari e prevede per gli *hosting provider* un regime di responsabilità limitata, subordinato al requisito dell'assenza di un «ruolo attivo» nella gestione dei contenuti[5].

Tale nozione è stata oggetto di definizione giurisprudenziale nella sentenza *Google France* del 2010, nella quale la Corte di giustizia ha affermato che un intermediario può beneficiare dell'esenzione di responsabilità solo se si limita a un ruolo meramente tecnico e passivo[6].

La sentenza citata affronta per la prima volta la questione dell'uso di marchi come parole chiave nel servizio di *advertising online* chiamato *AdWords*[7].

La Corte stabilisce che Google, consentendo agli inserzionisti di selezionare segni distintivi altrui come *keywords*, non compie «uso del marchio» ai sensi della direttiva 89/104/CEE[8], poiché tale utilizzo non è riferibile a prodotti o servizi propri della piattaforma, ma costituisce una funzione meramente tecnica di memorizzazione e gestione di parole chiave scelte dagli utenti inserzionisti. Al contrario, l'inserzionista che seleziona un marchio registrato come *keyword* realizza un uso commerciale del segno, potenzialmente idoneo a ledere la funzione distintiva del marchio qualora l'annuncio pubblicitario non consenta all'utente medio di comprendere con chiarezza l'origine imprenditoriale dei prodotti o servizi offerti[9]. La Corte precisa, infine, che Google può beneficiare dell'esenzione di responsabilità prevista dalla direttiva sul commercio elettronico, qualora il suo ruolo rimanga neutrale, tecnico e passivo, rinviando tuttavia la valutazione concreta ai giudici nazionali.

---

[4] Direttiva 2000/31/CE del Parlamento europeo e del Consiglio, dell'8 giugno 2000, relativa a taluni aspetti giuridici dei servizi della società dell'informazione, in particolare il commercio elettronico, nel mercato interno («Direttiva sul commercio elettronico»), artt. 12-15.

[5] Corte di giustizia, sentenza del 12 luglio 2011, causa C-324/09, *L'Oréal SA e a.* c. *eBay International AG e a.*, EU:C:2011:474, punto 113.

[6] Corte di giustizia, sentenza del 23 marzo 2010, cause riunite C-236/08, C-237/08 e C-238/08, *Google France SARL e Google Inc.* c. *Louis Vuitton Malletier SA*, EU:C:2010:159, punti 112-120.

[7] A. SAVIN, *EU Internet Law*, Cheltenham, 2013, pp. 88-92.

[8] Direttiva 89/104/CEE del Consiglio, del 21 dicembre 1988, ravvicinamento delle legislazioni degli Stati membri in materia di marchi.

[9] Sul concetto di funzione distintiva del marchio: C. GALLI, *I segni distintivi*, Milano, 2019, pp. 57-62.

La costruzione giuridica adottata dalla Corte presenta tuttavia elementi problematici. L'esclusione dell'«uso del marchio» da parte di Google si fonda su una distinzione eccessivamente formalistica tra attività tecnica e attività commerciale, trascurando che l'intero modello economico di *AdWords* si basa proprio sulla monetizzazione delle parole chiave[10], incluse quelle corrispondenti a marchi celebri. La Corte, nel privilegiare l'approccio funzionale alla tutela della concorrenza e dell'innovazione digitale, finisce per sottovalutare il potenziale di sfruttamento parassitario e di confusione indiretta derivante dalla gestione algoritmica delle *keywords*[11].

Inoltre, la nozione di «ruolo attivo», rilevante ai fini dell'esenzione di responsabilità, è delineata in termini generici e suscita incertezza applicativa, poiché non tiene conto della crescente capacità delle piattaforme di influenzare la visibilità, la rilevanza e il contesto degli annunci[12]. La sentenza, pur pionieristica, appare così ancorata a una visione del *web* centrata su intermediari neutrali, ormai superata dall'evoluzione dei servizi digitali e dalle esigenze di tutela effettiva del marchio nel contesto delle economie dei dati e degli algoritmi.

La Corte ha ribadito il principio dell'esenzione di responsabilità in *L'Oréal* c. *eBay*, affermando che l'intermediario perde la neutralità quando svolge un'attività «che gli consente di conoscere o controllare i dati memorizzati»[13]. La Corte EDU, per suo conto, nel caso *Delfi AS* c. *Estonia*[14], ha ritenuto compatibile la responsabilità di un portale per i commenti d'odio pubblicati dagli utenti, valorizzando il fatto che la testata operava a fini commerciali e traesse profitto dal traffico generato dai commenti. Ha così stabilito che, in circostanze eccezionali, l'imposizione di responsabilità può essere giustificata per prevenire gravi violazioni della dignità umana.

Sebbene questa distinzione sia rimasta formalmente invariata, l'evoluzione tecnologica delle piattaforme ha reso sempre più difficile considerarle meri soggetti passivi. Da un lato, si è evidenziato[15] che, se l'attività delle piattaforme deve esaurirsi al livello prossimo a quello dell'utente, ove avvengono le relazioni *human-computer*, il loro modello di business risulta molto più pervasivo, manifestandosi come un'entità unitaria che gestisce e controlla diversi livelli dell'architettura di Internet.

---

[10] B. EDELMAN, B. LOCKWOOD, *Measuring Bias in 'Organic' Web Search*, in *Journal of Industrial Economics*, 2011, pp. 1-27.

[11] S. BRADSHAW, P.N. HOWARD, *The Global Organization of Social Media Disinformation Campaigns*, in *Journal of International Affairs*, 2018, pp. 23-32.

[12] G. SARTOR, *Providers Liability: From the eCommerce Directive to the Digital Services Act*, in *Computer Law & Security Review*, 2022, pp. 45-60.

[13] Sentenza *L'Oréal SA e a.* c. *eBay International AG e a.*, cit., punti 124-131, punto 116.

[14] Corte europea dei diritti dell'uomo (Grande Camera), sentenza del 16 giugno 2015, ricorso n. 64569/09, *Delfi AS* c. *Estonia*, punti 110-122.

[15] G. BUTTARELLI, *La regolazione delle piattaforme digitali: il ruolo delle istituzioni pubbliche*, in *Giornale di diritto amministrativo*, 2023, pp. 116-127, spec. p. 119.

Dall'altro, gli algoritmi di raccomandazione, i sistemi di profilazione e i meccanismi di *ranking* dei contenuti costituiscono elementi centrali del funzionamento dei servizi, e rappresentano una forma di *curation* che conferisce alle piattaforme un potere informazionale senza precedenti[16].

In dottrina si è progressivamente consolidata l'idea che la disciplina della libertà di espressione *online* sia ormai il prodotto di un «private ordering», ossia un ordinamento autonomo creato e gestito dalle piattaforme digitali. Tale concetto, elaborato originariamente da Lessig, si fonda sulla constatazione che il codice – inteso come architettura tecnica delle piattaforme – operi come una forma di regolazione privata almeno altrettanto incisiva della legge statale[17].

Su questa linea si sviluppano gli studi più recenti sull'influenza crescente dei *terms of service* e delle *policy* di moderazione, che non rappresentano semplici contratti tra privati, ma veri e propri strumenti normativi capaci di definire i confini della partecipazione pubblica nello spazio digitale.

Secondo la celebre ricostruzione di Kate Klonick, le piattaforme globali operano come veri e propri «governatori privati dell'espressione», dotati di poteri regolatori che includono l'adozione di norme sostanziali (*community standards*), la predisposizione di procedure decisionali interne e la creazione di organi di revisione con funzioni quasi-giurisdizionali[18]. L'analisi di Klonick mostra come la moderazione dei contenuti non sia un'attività residuale o meramente tecnica, ma costituisca una funzione strutturale dell'ecosistema comunicativo digitale, necessaria per preservare la stabilità della piattaforma, mantenere l'*engagement* e garantire la sostenibilità economica dei servizi.

Ulteriori contributi teorici hanno approfondito il carattere para-normativo del potere delle piattaforme. Balkin ha proposto di considerare i grandi intermediari come *information fiduciaries*, soggetti che, gestendo infrastrutture centrali dell'espressione, assumono responsabilità analoghe a quelle delle istituzioni pubbliche[19].

Gillespie, dal canto suo, descrive i processi di *content moderation* come pratiche editoriali che influenzano direttamente la visibilità e la percezione dei contenuti *online*[20].

Cohen ha messo in luce come tali poteri regolativi siano parte integrante dell'economia dell'informazione contemporanea, contribuendo alla costruzione di forme di governamentalità privata e di sorveglianza strutturale[21].

---

[16] N. HELBERGER, K. KLEINEN-VON KÖNIGSLÖW, R. VAN DER NOLL, *Regulating the New Information Intermediaries as Gatekeepers of Information Diversity*, in *Info*, 2015, no. 6, pp. 50-71.

[17] L. LESSIG, *Code and Other Laws of Cyberspace*, New York, 1999, cap. 1-3.

[18] K. KLONICK, *The New Governors: The People, Rules, and Processes Governing Online Speech*, in *Harvard Law Review*, 2018, pp. 1598-1670.

[19] J. BALKIN, *Information Fiduciaries and the First Amendment*, in *UC Davis Law Review*, 2016, pp. 1183 -1234.

[20] T. GILLESPIE, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven, 2018.

[21] J.E. COHEN, *Between Truth and Power: The Legal Constructions of Informational Capitalism*, Oxford, 2019.

In questa prospettiva, la moderazione non può più essere concepita come semplice attività accessoria: essa diviene una funzione normativa primaria, in cui le piattaforme agiscono come veri imprenditori normativi, produttori autonomi di regole che modellano l'esercizio della libertà di espressione.

Ne deriva un modello ibrido di *governance*, caratterizzato da un pluralismo normativo nel quale la regolazione pubblica e quella privata si intrecciano profondamente, ridefinendo il ruolo dello Stato e degli attori privati nella tutela del discorso pubblico *online*[22].

3. La piattaforma digitale quale *gatekeeper*. Lo stato della normativa UE.

Nell'autunno del 2022 le istituzioni dell'UE hanno promulgato due regolamenti finalizzati a disciplinare la *platform economy*. Gli strumenti normativi hanno lo scopo precipuo di attuare la strategia dell'UE in materia che, al fine di limitare i possibili effetti collaterali e distorsivi del mercato, prevede di adottare una prospettiva *ex ante* attribuendo maggiori oneri e responsabilità ai soggetti che detengono maggior potere[23].

Da un lato, il *Digital Services Act* (DSA)[24] riguarda la responsabilità degli intermediari *online* per i contenuti di terze parti, la sicurezza degli utenti *online* e gli obblighi asimmetrici che impongono dei doveri di diligenza per i diversi fornitori di servizi della società dell'informazione[25]. La disciplina conferma l'esenzione da responsabilità dei prestatori di servizi di *mere conduit*, di *caching* e di *hosting* per contenuti illegali già emergente dalla citata direttiva *e-commerce* ed impone nuovi obblighi di attivazione per la rimozione degli stessi. Sotto altro aspetto, il DSA definisce la *very large platform* alla quale attribuisce degli obblighi supplementari di valutazione del rischio (art. 26), di adottare misure per la sua mitigazione (art. 27), di disporre un *audit* annuale indipendente (art. 28), di trasparenza supplementare per la pubblicità *online* (art. 30), di accesso ai dati da parte delle autorità nazionali ai fini di un controllo sulla sua attività (art. 31).

L'imposizione ad un intermediario di un obbligo generalizzato di monitoraggio dei contenuti degli utenti, tuttavia, è stata dichiarata incompatibile con il diritto dell'Unione

---

[22] N.P. SUZOR, *Lawless: The Secret Rules that Govern Our Digital Lives*, Cambridge, 2019.

[23] Proposta di regolamento del Parlamento europeo e del Consiglio relativo a mercati equi e contendibili nel settore digitale (legge sui mercati digitali), COM(2020) 842 final del 15 dicembre 2020.

[24] Regolamento (UE) 2022/2065 del Parlamento e del Consiglio, del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali).

[25] M. STUCCHI, *DSA: le nuove regole per le piattaforme online*, in *Il Diritto industriale*, 2024, pp. 270-277; S. DEL GATTO, *Il Digital Services Act: un'introduzione*, in *Giornale di diritto amministrativo*, 2023, pp. 724-729; G. FINOCCHIARO, *Responsabilità delle piattaforme e tutela dei consumatori*, in *Giornale di diritto amministrativo*, 2023, pp. 730-736; E. LONGO, *Libertà di informazione e lotta alla disinformazione nel Digital Services Act*, in *Giornale di diritto amministrativo*, 2023, pp. 737-745; G. SGUEO, *L'architettura istituzionale del Digital Services Act*, in *Giornale di diritto amministrativo*, 2023, pp. 746-752.

con la sentenza *SABAM* c. *Netlog*[26], che ha precisato come un simile obbligo violerebbe non solo la direttiva *e-commerce*, ma anche i diritti fondamentali degli utenti e della piattaforma, in particolare la libertà d'impresa (art. 16 Carta dei diritti fondamentali dell'UE) e la libertà di espressione (art. 11 Carta dei diritti fondamentali dell'UE). In particolare, la Corte ha osservato che un sistema di filtraggio generalizzato non sarebbe in grado di distinguere adeguatamente tra contenuti illeciti e contenuti legittimi, con il rischio di un'ampia rimozione errata.

Dall'altro, la disciplina complementare al DSA, costituita dal *Digital Markets Act* (DMA)[27], nel riconoscere ad alcuni dei principali protagonisti dei mercati digitali il ruolo di *gatekeeper*[28], si pone nel solco tracciato dall'evoluzione sopra descritta[29], al fine di prevenire eventuali problemi di equità, contendibilità e trasparenza che scaturiscono dalla particolare configurazione del settore, ciò indipendentemente dagli effetti reali, potenziali o presunti sulla concorrenza in un dato mercato derivanti dal comportamento di un dato *gatekeeper*[30].

In particolare, la normativa impone alcuni obblighi alle imprese che offrono servizi denominati «di piattaforma di base»[31] e che si trovino in una posizione tale da controllare l'accesso al mercato di riferimento, al fine di garantire l'equità e la contendibilità dei mercati nel settore digitale e a prescindere dall'esistenza di un pregiudizio alla concorrenza.

A questi strumenti normativi specifici della *platform economy* si affianca il *General Data Protection Regulation* (GDPR)[32] che tutela qualsiasi informazione riguardante una persona fisica identificata o identificabile, stabilendo che il trattamento dei dati personali da parte di terzi (ovvero la loro raccolta, la registrazione, l'organizzazione, la strutturazione, l'archiviazione, l'uso e la distruzione) può avvenire unicamente per motivi

---

[26] Corte di giustizia, sentenza del 16 febbraio 2012, causa C-360/10, *SABAM* c. *Netlog,* EU: C:2012:85, punti 45-52.

[27] Regolamento (UE) 2022/1925 del Parlamento europeo e del Consiglio, del 14 settembre 2022, relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali).

[28] Il 6 settembre 2023, la Commissione, ai sensi dell'art. 3 DMA, ha designato sei società come *gatekeeper*: Alphabet, Amazon, Apple, ByteDance, Meta e Microsoft.

[29] S. BERNASCONI, *Il ruolo del diritto internazionale privato e processuale nell'attuazione del «pacchetto sui mercati e servizi digitali» (DMA&DSA)*, in *Rivista di diritto internazionale privato e processuale*, 2024, pp. 1184-1258; G. BUTTARELLI, *La regolazione delle piattaforme digitali: il ruolo delle istituzioni pubbliche*, cit., p. 116; M. COLANGELO, *La regolazione ex ante delle piattaforme digitali: analisi e spunti di riflessione sul Regolamento sui mercati digitali (Regolamento (UE) 2022/1925 del 14 settembre 2022)*, in *Le Nuove leggi civili commentate*, 2023, pp. 415-440.

[30] Cfr. considerando 11 DMA.

[31] L'art. 2, n. 2, DMA include i servizi di intermediazione *online*, i motori di ricerca *online*, i servizi di *social network online*, i servizi di piattaforma per la condivisione di video, i servizi di comunicazione elettronica interpersonale indipendenti dal numero, i sistemi operativi, i *browser web*, i servizi di *cloud computing* e i servizi pubblicitari *online*.

[32] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).

specificamente individuati (ovvero il consenso dell'interessato, esecuzione di un contratto, l'adempimento di un obbligo legale, la protezione degli interessi vitali dell'interessato o di un'altra persona, l'esecuzione di un compito di interesse pubblico, il soddisfacimento di interessi legittimi). La tutela apprestata comprende il diritto di accesso del titolare dei dati personali, di ottenere la loro rettifica o la cancellazione e di pretendere la portabilità, nonché il diritto di opporsi al loro trattamento. Inoltre, per quello che ci interessa, stabilisce obblighi rigorosi per le piattaforme *online*, come la nomina di un responsabile della protezione dei dati e l'implementazione di misure di sicurezza adeguate e il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato (art. 22 GDPR).

I sopra menzionati regolamenti rappresentano la sintesi tra la necessità di tutelare i diritti riconosciuti dalla Carta dei diritti fondamentali dell'Unione europea e quella di agevolare l'innovazione tecnologica, contrastando la disinformazione e la diffusione di contenuti illegali. Si tratta di un ritorno ad una regolamentazione pubblica che ha lo scopo di garantire il funzionamento del libero mercato e la tutela dei consumatori che intende implementare attraverso il riconoscimento di alcuni diritti procedimentali in capo agli utenti[33]. Non si tratta di una disciplina che intende acquisire il controllo pubblico sulle infrastrutture di rete, o di impedire l'accentramento in capo a pochi soggetti dei diversi livelli di controllo dell'architettura della rete, ma di regolamentare le condotte delle piattaforme nel mercato e rispetto agli utenti. In altre parole, le istituzioni europee tentano di sottoporre l'attività delle piattaforme digitali alla *rule of law* di origine pubblicistica e, quindi, all'idea che il rispetto di procedure previste *ex ante* sia in grado di limitare comportamenti abusivi ed arbitrari[34].

4. La moderazione come funzione costituzionale privata.

Le piattaforme definiscono ciò che può circolare e ciò che deve essere rimosso attraverso regole interne – le *community guidelines* – che spesso presentano un contenuto più ampio rispetto alle previsioni normative.

Tale fenomeno è evidente soprattutto negli Stati Uniti, dove la *Section* 230 del *Communications Decency Act* conferisce agli intermediari un'ampia immunità non solo rispetto ai contenuti degli utenti, ma anche rispetto alle decisioni assunte in buona fede per limitarli o rimuoverli.

Sostanzialmente la sezione in questione prevede che nessun fornitore e nessun utilizzatore di servizi Internet possa essere considerato responsabile come editore o autore di una qualsiasi informazione fornita da terze persone. Secondo molti, su questa regola si

[33] G. BUTTARELLI, *La regolazione delle piattaforme digitali: il ruolo delle istituzioni pubbliche*, cit., pp. 119-120.

[34] G. BUTTARELLI, *La regolazione delle piattaforme digitali: il ruolo delle istituzioni pubbliche*, cit., p. 122.

è costruita la fortuna dei *social network* e, dunque, la stessa ha dato forma ad Internet nel bene e nel male; è stata inserita nel 1996 e dunque agli albori dell'ascesa globale della rete e sancendo che le piattaforme non sono responsabili di ciò che viene pubblicato da altri su di loro conferisce alle società che le gestiscono ampia discrezione nel modo in cui moderano i post e gli altri contenuti[35].

Nel 1997, nel caso *Zeran v. America Online* la Corte d'appello per il Quarto circuito identifica l'ampiezza dell'immunità conferita dalla sezione 230, stabilendo che quest'ultima garantisce il *provider* anche quando egli sia a conoscenza della natura illecita (mediante segnalazione) dei contenuti presenti sulla sua piattaforma[36]. Questa interpretazione amplia l'immunità del *provider* al di là dei limiti già estesi stabiliti nel precedente *Compuserve*, stabilendo che la sezione 230 non si limita ad escludere una qualificazione dei *provider* come editori, ma preclude che possano essere caratterizzati come dei distributori di contenuti (al pari di un edicolante) poiché ciò comprometterebbe

[35] Molti si sono chiesti quale sia l'origine di questa regola e la spiegazione la si rinviene in J. KOSSEFF, *The Twenty-Six Words That Created the Internet*, New York, 2019, ove l'autore racconta che la stessa ha origine nel contesto di una controversia risalente al 1954: un libraio, che aveva venduto un libretto erotico a due agenti sotto copertura, venne denunciato perché ritenuto colpevole di diffusione in materiale osceno. Nel corso del processo a suo carico il libraio si difese sostenendo di non aver mai letto quel libro e di non sapere di che cosa si trattasse così come non può sapere cosa trattano centinaia di libri nel suo negozio. Il libraio venne assolto e il principio stabilito: chi distribuisce contenuti in una responsabilità sui contenuti stessi, altrimenti si violerebbe il primo emendamento della Costituzione americana oltre ad arrecare danno all'indotto dell'editoria. La tendenza a garantire l'irresponsabilità del *provider* per contenuti caricati da terzi è ribadita anche dal *Digital Millennium Copyright Act* del 1996 Sul tema cfr. ampiamente R. IMPERADORI, *La responsabilità dell'Internet Service Provider per violazione del diritto d'autore: un'analisi comparata,* in *Trento Law and Technology group*, *Student paper n. 21*, 2014, pp. 1-227, reperibile *online*.

[36] Corte d'appello federale degli Stati Uniti d'America per il Quarto Circuito, sentenza del 2 ottobre 1997, *Zeran* v. *America Online*, Inc., 129 F. 3d 327 (4th Cir. 1997). Il ricorrente era stato vittima di minacce di morte e molestie telefoniche per un anno dopo che i suoi recapiti erano stati pubblicati da un utente anonimo su un forum interattivo gestito dal *provider* America Online, in collegamento alla vendita di oggetti satirici sull'attentato terroristico di Oklahoma City. Dunque, il contenuto del post era illecito. Il richiedente aveva ripetutamente chiesto l'eliminazione dal forum dei suoi dati personali, ma l'America Online vi aveva proceduto solamente a distanza di parecchio tempo non sufficiente ad impedire, nel frattempo, la comparsa di altri post che riportassero i suoi recapiti. Nel processo, il ricorrente aveva asserito la negligenza della America Online nella gestione del post, argomentando che la *Section* 230 non si applicasse a beneficio dell'America Online, poiché il ruolo di quest'ultima nella vicenda si configurava alla stregua non di un editore o di un produttore (per riprendere la terminologia impiegata nell'*Act*) bensì di un «distributore», equiparabile ad un'edicola, e pertanto responsabile se al corrente dell'illiceità del contenuto in questione. Ad avviso del ricorrente, America Online era consapevole dell'illiceità proprio in conseguenza delle ripetute richieste di cancellazione da lui avanzate. La Corte d'Appello ha stabilito invece che l'immunità era comunque applicabile ad America Online, perché ritenere il *provider* colpevole per aver semplicemente distribuito un dato illecito lo avrebbe equiparato al produttore del contenuto. Una dichiarazione di colpevolezza avrebbe indotto i *providers* ad una censura eccessiva delle comunicazioni, per timore di essere chiamati a rispondere ed inoltre, l'assenza di immunità avrebbe comportato, per i *providers*, la necessità di eseguire una verifica circa la sussistenza di estremi per una eventuale diffamazione, con una decisione che avrebbe dovuto essere pressoché istantanea, ciò che è praticamente impossibile data la natura delle comunicazioni su internet. Si è in tal modo superata l'incertezza ingenerata da due casi in cui si erano posti dubbi circa la responsabilità del *provider*: *Cubby, Inc.* v. *CompuServe, Inc.*, 776 F. Supp. 135 (S.D.N.Y. 1991) e *Stratton Oakmont, Inc.* v. *Prodigy Services Co.*, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

irreparabilmente il funzionamento dell'ecosistema digitale. La pronuncia, in particolare, precisa che imporre obblighi di controllo preventivo trasformerebbe gli intermediari in censori generalizzati, con un effetto dissuasivo sulla libertà di espressione.

La vasta estensione dell'immunità garantita dalla sezione 230 viene ulteriormente ribadita in successive decisioni statunitensi nelle quali è affermato il principio per il quale essa si estende anche ai casi in cui il terzo, autore delle asserite diffamazioni, è vincolato al *provider* per mezzo di un contratto di lavoro autonomo[37] o nei quali il terzo abbia commesso un evidente illecito penale[38].

Questo paradigma, tuttavia, si fonda su un modello di piattaforma ormai superato: quello della bacheca virtuale neutrale. L'attività di moderazione contemporanea si basa, invece, su un intreccio complesso tra disponibilità del contenuto, visibilità algoritmica e raccomandazione personalizzata. Non si tratta più, come nel modello originario, di decidere semplicemente se un contenuto debba rimanere online o essere rimosso: la moderazione opera ormai attraverso un continuum di interventi che includono la classificazione algoritmica, l'ordinamento nei *feed*, la prioritizzazione o la riduzione di visibilità dei contenuti e, più in generale, la loro *curation* dinamica. Le piattaforme determinano non solo ciò che è accessibile, ma anche e soprattutto quanto, quando e a chi un contenuto viene effettivamente mostrato, incidendo così sulla sua rilevanza sociale e sulla percezione pubblica del discorso.

La moderazione diviene dunque una modalità di ingegneria dell'attenzione, in cui ogni contenuto è sottoposto a un processo continuo di valutazione e riallocazione

[37] *Blumenthal* v. *Drudge and America Online, Inc.*, 992 F. Supp. 44 (D.C.C. 1998). Il giornalista (*columnist*) internet Matthew Drudge aveva esposto affermazioni asseritamente diffamatorie nei confronti di Sidney Blumenthal, noto ex-giornalista e dipendente della Casa Bianca. Drudge riceveva circa 3.000 dollari mensili come corrispettivo dei suoi servizi; nel contratto di lavoro, l'America Online si era riservata il diritto di rimuovere qualsiasi contenuto in violazione dei *Terms of Servic*e della America Online stessa, ma la responsabilità ed il controllo editoriale rimanevano interamente del giornalista. Blumenthal ha convenuto in giudizio sia Drudge sia la America Online, quest'ultima in qualità di creatrice del post che ha scatenato la controversia, data la riserva presente nel contratto. L'immunità conferita dalla *Section* 230 sarebbe dunque venuta meno, secondo Blumenthal, poiché l'informazione non era stata «fornita da un altro fornitore di contenuto informativo». La Corte ha tuttavia respinto la tesi di Blumenthal, sostenendo che Drudge non era né un agente né un dipendente dell'America Online, bensì un lavoratore autonomo; pertanto, l'informazione proveniva effettivamente da un fornitore di contenuti diverso e America Online poteva godere dell'immunità anche in questo caso.

[38] In *Doe* v. *America Online, Inc.*, No. 97-2587, 1998 WL 712764 (Fla. Ct. App. Oct. 14, 1998), la madre di un utente minorenne delle *chat-rooms* gestite da America Online aveva denunciato il *provider*, asserendo che un utente avesse ripetutamente richiesto di poter visionare un filmato nel quale l'utente avrebbe compiuto atti sessuali con il figlio. La madre ha sostenuto che America Online era stata negligente nel consentire tali richieste. I *Terms of Service* dell'America Online contenevano un espresso divieto della distribuzione di materiale «illegale, nocivo, osceno o comunque indesiderabile» per mezzo dei suoi canali telematici, ed il provider si era riservato il diritto di eliminare tale contenuto. Nonostante fossero giunte molte e ripetute richieste di bloccare l'attività dell'utente, America Online non aveva preso alcun provvedimento nei suoi confronti. La madre aveva da ciò dedotto che la condotta dell'utente, ma anche quella di America Online, costituissero una violazione delle leggi dello Stato della Florida che sanzionavano la partecipazione nella distribuzione della pedopornografia. La Corte d'Appello dello Stato della Florida ha fatto riferimento alla sentenza *Zeran* v. *America Online* per stabilire che un *provider* può godere dell'immunità sia in quanto editore sia in quanto distributore.

algoritmica. In questo scenario, l'intermediazione non è mai neutrale: è il risultato di decisioni tecniche e normative integrate, attraverso le quali la piattaforma esercita un ruolo attivo nella strutturazione della sfera pubblica digitale. Infatti, la transizione dalla moderazione umana alla moderazione algoritmica costituisce uno degli snodi centrali nell'evoluzione delle piattaforme digitali. Tale passaggio non avviene in un momento puntuale, ma si sviluppa gradualmente in relazione a trasformazioni tecnologiche, economiche e sociali che hanno ridefinito la scala e la natura stessa della comunicazione online. Le prime forme di moderazione, tipiche degli anni novanta e dei primi anni duemila, erano basate quasi interamente sul lavoro umano: amministratori, *community managers* e moderatori volontari controllavano contenuti numericamente limitati in ambienti digitali che ancora riflettevano la struttura lineare delle bacheche elettroniche[39]. A questo modello si ispirava la stessa *Section* 230 del *Communications Decency Act* del 1996, concepita per un Internet popolato da *forum* e primi *provider*, e non da ecosistemi globali fondati sulla personalizzazione algoritmica.

A partire dalla metà degli anni duemila, l'esplosione delle piattaforme partecipative – da YouTube (2005) a Facebook e Twitter – ha determinato un incremento esponenziale del volume dei contenuti generati dagli utenti, rendendo la moderazione umana non solo difficoltosa, ma strutturalmente insufficiente. La dottrina ha sottolineato come la gestione manuale fosse incapace di garantire rapidità, uniformità e capacità di intervento su contenuti multilingue e multimodali[40]. In questa fase emergono i primi sistemi automatici di classificazione, inizialmente limitati a funzioni specifiche come la rilevazione di spam o la prevenzione della pornografia minorile mediante tecnologie come *PhotoDNA* di Microsoft, sviluppata nel 2009.

Il periodo compreso tra il 2010 e il 2016 segna una svolta decisiva: l'affermazione del *machine learning* e del *deep learning* introduce forme di moderazione ibride, nelle quali gli algoritmi assumono il ruolo di filtri preliminari, incaricati di segnalare o rimuovere contenuti prima dell'intervento umano. Studiosi come Klonick e Roberts evidenziano come la scala raggiunta dalle piattaforme renda l'automazione «non più opzionale, ma necessaria»[41], trasformando la moderazione da attività artigianale a un processo industrializzato sostenuto da grandi investimenti tecnologici.

Dopo il 2016, scandali come l'ingerenza informativa nei processi elettorali e il caso *Cambridge Analytica* accelerano ulteriormente l'evoluzione dei sistemi di controllo dei contenuti. Le piattaforme sviluppano strumenti di classificazione predittiva, sistemi di

[39] S.L. ROBERTS, *Behind the Screen: Content Moderation in the Shadows of Social Media*, New Haven-London, 2019, pp. 42-57.

[40] K. KLONICK, *The New Governors: The People, Rules, and Processes Governing Online Speech*, cit., pp. 1610-1615.

[41] K. KLONICK, *The New Governors: The People, Rules, and Processes Governing Online Speech*, cit., p. 1623; J. GRIMMELMANN, *The Virtues of Moderation*, in *Yale Journal of Law & Technology*, 2015, pp. 42-109, reperibile *online*.

rimozione automatica dei contenuti terroristici e tecniche avanzate di *graph analysis*. In questa fase, la maggior parte dei contenuti rimossi non passa più attraverso moderatori umani, ma viene eliminata direttamente da filtri algoritmici basati su enormi *dataset* di addestramento[42].

La trasformazione più profonda, tuttavia, riguarda l'estensione della moderazione oltre la dimensione binaria permesso/non permesso.

La moderazione contemporanea opera come una forma di *governance* algoritmica della visibilità: non si limita a decidere se un contenuto debba restare online, ma determina quanto, quando e a chi esso debba essere mostrato.

Il *ranking*, la riduzione di visibilità, lo *shadow banning*, la raccomandazione personalizzata e i sistemi predittivi di rischio rappresentano strumenti attraverso i quali le piattaforme esercitano un potere editoriale senza precedenti. Come osservano Gillespie, Pasquale e Zuboff, la centralità del *ranking* algoritmico trasforma la moderazione in un processo continuo di «ingegneria dell'attenzione», capace di plasmare la percezione pubblica del dibattito online molto più della semplice rimozione dei contenuti[43].

In termini cronologici, il passaggio da una moderazione prevalentemente umana a una fondata sull'intelligenza artificiale può dunque essere collocato tra il 2010 e il 2016, con una definitiva affermazione negli anni successivi. Sul piano concettuale, invece, esso corrisponde al superamento del modello della piattaforma come spazio neutrale e statico, sostituito da un ambiente dinamico, continuamente riorganizzato da algoritmi di classificazione, raccomandazione e previsione del comportamento degli utenti.

5. L'intelligenza artificiale: tra efficienza tecnica e limiti epistemici.

L'enorme mole di contenuti caricati ogni secondo rende impossibile una moderazione interamente umana e le piattaforme utilizzano sistemi di IA addestrati su vasti *dataset* per rilevare contenuti violenti, pornografici, d'odio o terroristici[44].

Tali sistemi permettono una reazione rapida e scalabile, ma presentano limiti strutturali che la dottrina ha ampiamente evidenziato: incapacità di comprendere il contesto, riproduzione di *bias* culturali, difficoltà nell'identificare ironia, satira o contenuti politici sensibili. L'utente tende a fare affidamento sull'*output* di una tecnologia che viene definita «intelligente», in ragione di un prestito concettuale da altre discipline[45],

[42] E. NEWTON, *AI Content Moderation after 2016: A New Paradigm*, in *Journal of Online Governance*, 2020.

[43] S. ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, 2019; T. GILLESPIE, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, cit.; F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, 2015.

[44] J. GRIMMELMANN, *Note*, *Regulation by Software*, in *The Yale Law Journal*, 2005, pp. 1721-1758, reperibile *online*.

[45] L. FLORIDI, *La differenza fondamentale*, cit., p. 90 ss.

ed è rafforzato dall'efficienza con cui è in grado di svolgere certi compiti, tanto che presuppone erroneamente la capacità della nuova tecnologia di scartare *fake news* e di fornire risposte veritiere; da qui, il rischio di una diffusione incontrollata di informazioni false e stereotipi negativi[46].

Con il c.d. *Artificial Intelligence Act* (*AI Act*)[47] sono state introdotte delle regole armonizzate in materia di IA con lo scopo dichiarato di promuoverne lo sviluppo, garantendo un elevato livello di protezione degli interessi pubblici e dei diritti fondamentali del diritto dell'Unione[48] attraverso la regolamentazione dei sistemi di intelligenza artificiale con un approccio basato sul rischio[49] che proibisce l'utilizzo di certi sistemi («rischi inaccettabili») e interviene su altri, o imponendo obblighi e standard di condotta (per sistemi «ad alto rischio» o per «rischi specifici») o adottando codici di condotta («rischi minimi o limitati»).

Per quello che a noi interessa, i sistemi di IA progettati per generare contenuti e interagire con singoli impongono ai fornitori ed ai *deployer* degli obblighi informativi e di trasparenza in quanto potenzialmente in grado di arrecare danno agli utenti ed a costituire una potenziale minaccia ai processi democratici tramite la diffusione di informazioni false o manipolate. Allo stesso tempo, questo costituisce un esempio di rischio sistemico che il *gatekeeper* è tenuto ad individuare ed attenuare ai sensi del DMA.

In ogni caso, il regolamento incoraggia l'elaborazione di codici di condotta da parte dei fornitori, dei *deployer* e delle organizzazioni che li rappresentano così da stimolare l'adozione dei requisiti per i sistemi di IA ad alto rischio su base volontaria e contribuire a diffondere un'IA etica ed affidabile[50]. L'*AI Act*, in altre parole, svolge un ruolo proattivo, riconoscendo la necessità di regolare anche sistemi di IA a basso rischio, ma ne affida il compito alla autonomia privata.

6. Moderazione ibrida e ruolo umano. Il *Digital Services Act*.

[46] M. CAPPARELLI, *Disinformazione online, intelligenza artificiale e ruolo dell'autoregolamentazione*, in *Giurisprudenza Italiana*, 2024, pp. 480-483, spec. p. 481.

[47] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale). Per una panoramica, cfr. R. PETRUSO, G, SMORTO, *Il Regolamento europeo sull'intelligenza artificiale: una prima lettura*, in *La Nuova giurisprudenza civile commentata*, 2024, pp. 989-1004.

[48] Cfr. considerando 8 regolamento 2024/1689.

[49] Cfr. considerando 26 regolamento 2024/1689. G. SMORTO, *Distribuzione del rischio e tutela dei diritti nel Regolamento europeo sull'intelligenza artificiale. Una riflessione critica*, in *Il Foro Italiano*, 2024, V, cc. 208-220; G. FINOCCHIARO, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Il Diritto dell'informazione e dell'informatica*, 2022, pp. 303-322; A. SOLOW-NIEDERMANN, *Administering Artificial Intelligence*, in *Southern California Law Review*, 2020, pp. 635-696.

[50] Cfr. considerando 165 regolamento 2024/1689.

La moderazione algoritmica non elimina la necessità di intervento umano. I moderatori valutano contenuti complessi, analizzano ricorsi, contestualizzano situazioni culturali delicate e gestiscono casi limite. Da un lato, le ricerche di Sarah Roberts hanno evidenziato come i lavoratori della moderazione siano esposti quotidianamente a contenuti traumatici, spesso senza adeguate tutele[51]. Dall'altro, Floridi osserva che gli agenti artificiali non si limitano a interpretare dati, ma concorrono a strutturare l'ambiente informazionale in cui la conoscenza viene prodotta e condivisa[52]. Ciò significa che la moderazione automatizzata non è solo un'attività tecnica, ma un intervento sul «tessuto epistemico» della società: la rimozione errata – e spesso invisibile – di contenuti legittimi produce effetti sistemici sulla qualità del discorso pubblico.

Il *Digital Services Act*[53] interviene in modo diretto sul tema della moderazione dei contenuti, riconoscendo l'insufficienza di un modello fondato esclusivamente su processi decisionali automatizzati[54] analogamente a quanto previsto dall'art. 22 GDPR, che attribuisce al titolare dei dati personali 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. L'impianto normativo degli artt. 17-22 DSA si articola attorno a un principio cardine: la moderazione dei contenuti, soprattutto quando incide sulla libertà di espressione e sulla posizione giuridica degli utenti, non può essere delegata interamente all'intelligenza artificiale, ma deve prevedere garanzie procedurali e forme di supervisione umana. Tale principio rappresenta una scelta di distinzione netta rispetto alla tradizione statunitense della *Section* 230, in cui il ruolo umano nella moderazione non è giuridicamente imposto.

L'art. 17 DSA disciplina le procedure di segnalazione e di azione (*notice and action*), imponendo ai prestatori di *hosting* di comunicare agli utenti «una spiegazione chiara e motivata della decisione» adottata in relazione ai contenuti segnalati[55]. Questo obbligo di motivazione è incompatibile con decisioni integralmente automatizzate, in quanto il prestatore deve essere in grado di fornire una giustificazione comprensibile e

[51] S.T. ROBERTS, *Behind the Screen: Content Moderation in the Shadows of Social Media*, cit.

[52] L. FLORIDI, *La differenza fondamentale*, cit., pp. 150-160.

[53] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio, del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali).

[54] M. CAPPARELLI, *Disinformazione online, intelligenza artificiale e ruolo dell'autoregolamentazione*, cit., p. 480; F. CASOLARI, *Il "Digital Services Act" e la costituzionalizzazione dello spazio digitale europeo*, in *Giurisprudenza Italiana*, 2024, pp. 462-465; B. GRAZZINI, *"Fake news" e disinformazione*, in *Giurisprudenza Italiana*, 2024, pp. 491-498; U. RUFFOLO, *Piattaforme e content moderation negoziale*, in *Giurisprudenza Italiana*, 2024, pp. 442-452.[55] Art. 17, par. 3, DSA: «[i]l prestatore di servizi di *hosting* (…) fornisce all'utente una spiegazione chiara e motivata della decisione adottata».

[55] Art. 17, par. 3, DSA: «[i]l prestatore di servizi di *hosting* (…) fornisce all'utente una spiegazione chiara e motivata della decisione adottata».

contestuale, che presuppone un controllo umano almeno a valle della decisione algoritmica.

Il cuore delle garanzie risiede però nell'art. 20 DSA, che istituisce un meccanismo interno di gestione dei reclami. La norma stabilisce che gli utenti destinatari di una decisione di moderazione abbiano diritto a contestarla tramite «un sistema di gestione dei reclami facilmente accessibile e dotato di personale umano adeguatamente qualificato»[56]. Tale sistema deve consentire una revisione effettiva anche delle decisioni prese mediante strumenti automatizzati (art. 20, par. 4). L'obbligo di presenza umana è qui esplicito e non derogabile: il legislatore europeo esclude consapevolmente la possibilità che i sistemi di reclamo vengano gestiti da algoritmi, riconoscendo che l'interpretazione del contesto comunicativo richiede capacità discrezionali non surrogabili dall'IA.

Un ulteriore tassello è fornito dall'art. 22 DSA, che disciplina l'uso dei sistemi di moderazione automatizzati. La norma dispone che i prestatori che utilizzano strumenti automatici debbano adottare «garanzie adeguate, in particolare la supervisione e il controllo umano»[57].

Il DSA, dunque, non si limita a chiedere trasparenza sul funzionamento dei filtri, ma impone un'architettura procedurale ibrida, nella quale l'automazione è sempre integrata da verifiche umane, soprattutto quando sono in gioco diritti fondamentali.

Infine, per i prestatori di dimensioni molto grandi (VLOP), il DSA introduce un regime di responsabilità rafforzato. L'art. 34 DSA richiede che tali soggetti effettuino valutazione del rischio sistemico, tra cui i rischi connessi all'uso dei sistemi di raccomandazione e moderazione automatica. L'art. 35 DSA impone poi misure di mitigazione proporzionate, che possono includere la revisione umana delle decisioni automatizzate, mentre l'art. 36 DSA prescrive obblighi di *auditing* indipendente sui sistemi di moderazione, con riferimento espresso ai rischi derivanti da *bias* algoritmici e da errori sistemici. L'insieme di queste disposizioni conferma che l'automazione non è solo uno strumento tecnico, ma anche un potenziale fattore di rischio giuridico e sociale, che richiede interventi umani costanti.

La dottrina ha ampiamente sottolineato come questa architettura normativa introduca un modello di moderazione «human-in-the-loop», che intende bilanciare efficienza algoritmica e garanzie procedurali. Secondo Angelopoulos, il DSA afferma una «nuova costituzionalità digitale della moderazione», nella quale la presenza umana diviene elemento essenziale di legittimazione delle decisioni restrittive della libertà di espressione[58]. Keller osserva che il DSA rappresenta «la prima conferma legislativa

[56] Art. 20, par. 2, DSA: «[i] prestatori di piattaforme online istituiscono un sistema interno di gestione dei reclami facilmente accessibile e dotato di personale umano adeguatamente qualificato».

[57] Art. 22, par. 3, lett. *b*, DSA: «[i] prestatori adottano garanzie adeguate, in particolare la supervisione e il controllo umano dei sistemi automatizzati».

[58] A.N. ANGELOPOULOS, *The Digital Services Act and the Constitutionalization of Online Speech Governance*, in *European Law Review*, 2023.

europea del fatto che la moderazione algoritmica, da sola, non soddisfa gli standard minimi del giusto procedimento»[59]. Altri autori evidenziano che la presenza di personale umano nei processi di revisione mira non solo a correggere gli errori degli algoritmi, ma anche a introdurre un grado di responsabilità personale nel processo di moderazione[60].

Nel complesso, il DSA si presenta come una chiara presa di posizione normativa contro l'automazione totale della moderazione: la tecnologia può accelerare il processo, ma la valutazione finale deve coinvolgere un decisore umano, capace di comprendere il contesto, applicare principi proporzionati e garantire il rispetto dei diritti fondamentali.

Un ulteriore profilo decisivo riguarda il rapporto tra il DSA e il quadro dei diritti fondamentali dell'Unione, che trova il suo fulcro nell'art. 52 della Carta dei diritti fondamentali dell'UE, il quale stabilisce che ogni limitazione a un diritto fondamentale deve rispettare il principio di proporzionalità, essere necessaria e rispondere effettivamente a un obiettivo di interesse generale riconosciuto dall'Unione.

La moderazione dei contenuti – soprattutto quando affidata a sistemi automatizzati – costituisce una restrizione della libertà di espressione e deve pertanto essere sorretta da un adeguato impianto procedurale. In questo senso, gli obblighi di supervisione umana e di motivazione introdotti dagli artt. 17-22 DSA rappresentano strumenti essenziali per garantire che la limitazione del diritto non sia né arbitraria né tecnicamente opaca.

La giurisprudenza dell'Unione ha più volte affermato che l'uso di sistemi automatizzati o di trattamenti massivi di dati può costituire una violazione dei diritti fondamentali in assenza di garanzie effettive e trasparenti. Nelle sentenze *Digital Rights Ireland* e *Schrems*, la Corte di giustizia ha evidenziato che la sorveglianza sistematica o il trattamento algoritmico non supervisionato sono incompatibili con il nucleo essenziale dei diritti alla *privacy*, alla protezione dei dati e alla libertà di comunicazione quando l'interessato non può comprendere né contestare l'operato del sistema[61]. Questi principi, originariamente sviluppati nel contesto della sorveglianza statale e del trasferimento di dati verso paesi terzi, sono oggi pienamente applicabili anche alla *governance* algoritmica delle piattaforme: la mancanza di trasparenza e di controllo umano è di per sé un rischio giuridico, poiché ostacola l'effettività dei rimedi e la verificabilità del rispetto del principio di proporzionalità.

---

[59] D. KELLER, *Automated Moderation and the DSA's Human Oversight Mandate*, in *Journal of European Internet Law*, 2023.

[60] J. SMITH, *Human Oversight in Content Moderation after the DSA*, in *Common Market Law Review*, 2023, pp. 1123-1154.

[61] Corte di giustizia (Grande Sezione), sentenza dell'8 aprile 2014, cause riunite C-293/12 e C-594/12, *Digital Rights Ireland Ltd* e *Kärntner Landesregierung*, punti 37-69, sulla necessità che ogni trattamento massivo o automatizzato sia limitato allo stretto necessario e soggetto a «garanzie efficaci e verificabili» e Corte di giustizia (Grande Sezione), sentenza del 6 ottobre 2015, causa C-362/14, *Schrems c. Data Protection Commissioner*, EU:C:2015:650, punti 94-104, sulla necessità di rimedi effettivi e di controllo umano nei sistemi automatizzati di trasferimento e trattamento dati.

Il DSA recepisce – e in parte codifica – tale orientamento giurisprudenziale, costruendo un regime nel quale l'automazione non può mai sostituire integralmente l'intervento umano, specialmente quando la decisione implica un bilanciamento tra libertà di espressione, tutela della dignità umana e interessi pubblici rilevanti. È proprio alla luce dell'art. 52 della Carta che la supervisione umana prevista dal DSA assume rango costituzionale: essa è la garanzia minima affinché la moderazione algoritmica non si traduca in una restrizione arbitraria, opaca o sproporzionata dei diritti fondamentali degli utenti[62].

7. Etica, diritti fondamentali e implicazioni sistemiche.

La moderazione, umana e algoritmica, incide su una pluralità di diritti fondamentali e, in particolare, sulla libera manifestazione del pensiero che ai sensi dell'art. 21 Cost. tutela non solo il soggetto che lo manifesta, ma anche chi lo ascolta[63]. Quest'ultimo aspetto, si è sottolineato, riguarda il diritto di fruire di tutte le comunicazioni e di non vederle censurate. La libertà di espressione, pertanto, abbia essa una fonte umana o sia frutto della risposta dell'IA ad un *prompt*, è frequentemente compressa a causa di rimozioni preventive o erronee frutto dell'attività di *content moderation*[64]. Si pensi, ad esempio, alle conseguenze della sentenza *Glawischnig-Piesczek*[65] nella parte in cui ha riconosciuto che un intermediario può essere obbligato a rimuovere non solo un contenuto specificamente individuato come illecito, ma anche contenuti «equivalenti», cioè caratterizzati da un contenuto diffamatorio sostanzialmente analogo. Nella parte centrale della motivazione, infatti, la Corte afferma che tale obbligo può estendersi a contenuti «che, pur non essendo identici, presentano differenze minime e non richiedono una valutazione autonoma». Ciò implica che la piattaforma deve operare un giudizio semantico, con un margine di discrezionalità che solleva interrogativi importanti sul potenziale rischio di sovra-rimozione.

Sotto altro aspetto, il diritto alla protezione dei dati personali è influenzato dall'uso di sistemi di profilazione necessari alla personalizzazione dei contenuti. La non discriminazione è messa in pericolo dalla riproduzione di *bias* negli algoritmi.

---

[62] Art. 52, par. 1, Carta dei diritti fondamentali dell'UE stabilisce che ogni limitazione dei diritti fondamentali deve essere «necessaria e proporzionata» e «prevista dalla legge».

[63] U. RUFFOLO, *AI generativa, libertà di manifestazione del pensiero e diritto d'autore*, in U. RUFFOLO, C. AMIDEI, *Diritto dell'intelligenza artificiale*, vol. II, Roma, 2024, pp. 139-192.

[64] U. RUFFOLO, *Piattaforme e* content moderation*: "censura privata" o* soft law *governabile dall'autonomia negoziale (contrattuale, autodisciplinare, coregolamentare)? La efficacia "orizzontale" dei precetti costituzionali quali l'art. 21 ed il limite dell'ordine pubblico (e della "meritevolezza" dell'interesse contrattuale)*, in U. RUFFOLO, C. PINELLI, *I diritti delle piattaforme*, Torino, 2023, pp. 43-58.

[65] Corte di giustizia, sentenza del 3 ottobre 2019, causa C-18/18, *Glawischnig-Piesczek* c. *Facebook*, EU:C:2019:821, punto 46.

Come sottolinea Floridi, la questione fondamentale è quella di garantire che l'infosfera rimanga un ambiente «giusto» nel quale tutti possano partecipare al discorso pubblico in condizioni paritarie[66].

La moderazione è un'attività che ha in sé le virtù necessarie per questo fine – e, quindi, è un'attività «etica» in un senso profondo – che può realizzarsi non solo attraverso la creazione di norme condivise tra i membri della comunità, che costituisce la tecnica per lo più adottata per questo fine specifico, ma anche attraverso l'implementazione di altre modalità (esclusione, trasferimento dei costi, organizzazione dei contenuti) per mezzo delle quali è possibile influire sulle dinamiche della comunità stessa[67].

In questo ambito, il moderatore può direttamente influenzare le norme che emergono naturalmente dalle interazioni dei membri della comunità attraverso la loro formulazione e, indirettamente, può incoraggiare gli utenti a rispettare le norme della comunità, oppure intervenire quando le stesse non vengono rispettate[68]. In questi ultimi due casi, si può ricorrere a diverse tecniche di moderazione al fine di indirizzare i membri della comunità verso comportamenti virtuosi. Ad esempio, l'organizzazione dei contenuti non solo permette agli utenti di individuare quelli più interessanti, ma svolge pure una funzione educativa che gli aiuta a riconoscere le caratteristiche che consentono ad un contenuto di essere messo al primo posto. Allo stesso modo, l'esclusione di un membro particolarmente autorevole dalla comunità, anche se giustificata, potrebbe ridurre la fiducia nel moderatore. Insomma, ogni decisione presa nell'attività di moderazione ha un impatto sulla comunità e questo effetto deve essere preso in considerazione perché contribuisce a formare una identità condivisa che rafforzi il senso di appartenenza dei partecipanti e il loro impegno per il bene della comunità[69].

Valutazioni latamente etiche sottendono, altresì, alle deliberazioni attraverso le quali si decide di attuare la moderazione dei contenuti che includono la scelta di optare per una moderazione umana o automatizzata, trasparente o segreta, *ex ante* o *ex post*, centralizzata o diffusa[70], da adattare alle diverse comunità[71].

8. Conclusioni.

La moderazione dei contenuti rappresenta oggi uno dei terreni centrali in cui si misura la qualità della governance dell'infosfera contemporanea. Le piattaforme digitali, ormai configurabili come infrastrutture essenziali del discorso pubblico, esercitano un

[66] L. FLORIDI, *La differenza fondamentale*, cit., p. 211.

[67] J. GRIMMELMANN, *The Virtues of Moderation*, cit., p. 61 ss.; l'autore inquadra l'attività di *norm-setting* come una delle tecniche attraverso le quali si attua la moderazione dei contenuti.

[68] J. GRIMMELMANN, *The Virtues of Moderation*, cit., p. 62.

[69] J. GRIMMELMANN, *The Virtues of Moderation*, cit., p. 63.

[70] J. GRIMMELMANN, *The Virtues of Moderation*, cit., p. 63 ss.

[71] J. GRIMMELMANN, *The Virtues of Moderation*, cit., p. 70 ss.

potere regolativo che incide direttamente sull'effettività della libertà di espressione, sulla pluralità informativa e sulle dinamiche democratiche. L'intelligenza artificiale, pur offrendo capacità di gestione e reazione impossibili da replicare con strumenti umani, non elimina i nodi critici che emergono quando decisioni rilevanti per i diritti fondamentali vengono delegate a sistemi opachi, non intelligibili e potenzialmente discriminatori[72].

In questo scenario, il quadro europeo si distingue nettamente da quello statunitense per una scelta politica e costituzionale precisa: ricondurre la moderazione dentro un perimetro fondato su responsabilità, trasparenza e controllo umano. Il *Digital Services Act* costruisce un modello nel quale l'automazione non può agire in assenza di garanzie procedurali effettive, imponendo obblighi di motivazione, supervisione e revisione umana che mirano a impedire che le piattaforme esercitino un potere editoriale incontrollato e strutturalmente opaco[73]. Si tratta di un approccio che rifiuta la neutralità tecnologica come illusione e riconosce esplicitamente che l'algoritmo è una forma di potere che richiede limiti e contrappesi.

Il DSA, tuttavia, costituisce un punto di partenza più che un approdo definitivo.

L'evoluzione dell'IA generativa, dei sistemi predittivi e delle architetture dell'attenzione richiede interventi ulteriori volti ad assicurare livelli più elevati di trasparenza, *audit* indipendenti realmente incisivi e meccanismi di ricorso comprensibili e accessibili. Resta inoltre essenziale rafforzare l'*accountability* degli intermediari, soprattutto quando le decisioni riguardano contenuti politici o giornalistici, in cui gli effetti sulle dinamiche democratiche sono maggiormente evidenti[74].

Un fronte particolarmente sensibile riguarda la dimensione epistemica della moderazione: il modo in cui le piattaforme selezionano, ordinano e rendono visibili i contenuti produce effetti sistemici sulla percezione pubblica e sulla formazione dell'opinione. La riflessione di Floridi sull'esigenza di preservare un'infosfera «giusta» e inclusiva sottolinea che la questione non è soltanto tecnica, ma eminentemente politica e sociale[75].

La progettazione degli algoritmi, infatti, non determina solo ciò che è eliminato, ma anche ciò che è amplificato, reso credibile o relegato ai margini, con conseguenze profonde per la qualità della deliberazione pubblica.

In definitiva, la sfida per il legislatore, per le autorità garanti e per le stesse piattaforme consiste nel riconoscere che la moderazione non è un'attività ancillare, bensì una vera e propria funzione normativa privata, che richiede regole chiare, controlli pubblici efficaci e un grado minimo di trasparenza strutturale.

---

[72] J. GRIMMELMANN, *The Virtues of Moderation*, cit., pp. 157-165.

[73] Artt. 17-22 e 34-36 DSA; D. KELLER, *Automated Moderation and the DSA's Human Oversight Mandate*, cit.

[74] E. CELESTE, *Digital Constitutionalism and the Role of Internet Bills of Rights*, London, 2022.

[75] L. FLORIDI, *The Logic of Information*, cit., pp. 105-130.

Il superamento del mito della neutralità dell'intermediario è un passo necessario per costruire un ecosistema informativo nel quale innovazione e diritti fondamentali possano coesistere senza che l'una sacrifichi gli altri. Solo in questo modo l'intelligenza artificiale potrà trasformarsi da potenziale fonte di rischio in un elemento capace di rafforzare la qualità della democrazia digitale, contribuendo alla costruzione di un ambiente informazionale equo, affidabile e rispettoso della dignità delle persone.

ABSTRACT: La moderazione dei contenuti nelle piattaforme digitali rappresenta un nodo centrale della *governance* dell'infosfera contemporanea. Il saggio analizza il ruolo delle piattaforme come regolatori privati, l'evoluzione tecnico-giuridica della moderazione algoritmica, il quadro comparato tra UE, USA e Corte di giustizia, e le implicazioni per i diritti fondamentali, con un focus sugli obblighi del *Digital Services Act*.

PAROLE CHIAVE: moderazione dei contenuti; piattaforme digitali; DSA; IA; libertà di espressione; responsabilità degli intermediari.

From bulletin board to gatekeeper: digital platforms and moderation

ABSTRACT: *Content moderation on digital platforms represents a central node in the governance of the contemporary infosphere. The essay examines the role of platforms as private regulators, the techno-legal evolution of algorithmic moderation, the comparative framework between the EU, the United States, and the Court of Justice of the European Union, and the implications for fundamental rights, with a particular focus on the obligations established by the Digital Services Act.*

KEYWORDS: *content moderation; digital platforms; DSA; AI; freedom of expression; intermediary liability.*