Intelligenza artificiale e fonti del diritto: verso un nuovo concetto di *soft law*? La rimozione dei contenuti terroristici *online* come *case-study*

di Chiara Graziani

Abstract: Artificial intelligence and sources of law: towards a new concept of soft law? The removal of terrorist content online as a case-study – This article addresses the regulation of artificial intelligence aimed at detecting and removing terrorist content online. After pointing out the lack of traditional sources of law – both binding and non-binding – in this field, the author highlights a pervasive and potentially dangerous role of terms of service and community standards adopted by private actors, who play the role of (quasi) law-makers. The conclusions of this work argue in favour of a stronger involvement of traditional regulators and, hence, of traditional sources.

Keywords: Artificial intelligence; Counter-terrorism; Terrorist content online; Law-making; Sources of law; Soft law

1. Intelligenza artificiale e fonti del diritto: profili introduttivi

L'intelligenza artificiale è ormai impiegata nei contesti più diversi della vita e delle attività umane. Tra i moltissimi settori in cui gli algoritmi c.d. intelligenti vengono usati, alcuni risultano particolarmente delicati, come la tutela della sicurezza pubblica da gravi minacce, quale è il terrorismo internazionale[1].

Dal lato squisitamente tecnico, esistono una pluralità di sistemi automatizzati intelligenti che possono risultare strumentali all'antiterrorismo[2]. Il presente lavoro si concentra sugli algoritmi di identificazione e di rimozione dei contenuti terroristici sul *web*. L'intelligenza

[1] Per una disamina dell'evoluzione delle *counter-terrorism measures* nei primi anni dopo il 2001, A. Vedaschi, À la guerre comme à la guerre*? La disciplina della guerra nel diritto costituzionale comparato*, Torino, 2007. Sulle evoluzioni più recenti, G. Lennon, C. Walker (eds), *Routledge Handbook of Law and Terrorism*, London-New York, 2015; A. Vedaschi, K.L. Scheppele (eds), *9/11 and the Rise of Global Anti-Terrorism Law: How the UN Security Council Rules the World*, Cambridge, 2021.

[2] *Ex plurimis*, la c.d. sorveglianza algoritmica sui metadati delle comunicazioni o su altre operazioni (ad esempio, le transazioni finanziarie), il riconoscimento facciale, la *counter-radicalization* svolta da sistemi come Jigsaw.

DPCE online
ISSN: 2037-6677

artificiale può essere impiegata a vario titolo in tali operazioni. In particolare, gli algoritmi possono essere "allenati" a riconoscere contenuti (messaggi, *posts*, immagini), presenti sulle diverse piattaforme, correlati al terrorismo, per poi porre in essere azioni preventive e/o repressive: segnalare ad un operatore umano, rimuovere direttamente, impedire l'*upload* oppure permetterlo oscurando la parte di contenuto oggetto del *flag* di pericolosità[3]. A seguito di tali processi, l'autore del contenuto illecito potrà essere sottoposto a procedimenti penali, secondo le norme vigenti nell'ordinamento di riferimento, per reati come l'istigazione, l'apologia, la glorificazione del terrorismo[4]. Lo scenario illustrato comporta un "affidamento"[5] delle autorità pubbliche su soggetti privati (*services providers*, motori di ricerca, aziende di tecnologia in generale) che ospitano i contenuti sulle loro piattaforme e dispongono della tecnologia necessaria alla loro eventuale rimozione.

Dall'angolatura prettamente giuridica, si osserva una significativa confusione, se non veri e propri *gaps*, nella regolamentazione dell'intelligenza artificiale a fini di eliminazione di contenuti terroristici *online*.

In primo luogo, da una prospettiva multilivello, si rileva la scarsezza di c.d. fonti dure del diritto (*hard law*), in quanto poche di esse regolano nel dettaglio l'uso dell'intelligenza artificiale nell'antiterrorismo. In secondo luogo, alcuni soggetti istituzionali (ad esempio l'Unione europea o il Consiglio d'Europa) hanno adottato linee guida non vincolanti, classificabili come *soft law*, spesso vaghe e imprecise, più simili a dichiarazioni di principio che a vere e proprie fonti giuridiche e comunque quasi mai focalizzate sull'utilizzo dell'intelligenza artificiale nel *counter-terrorism*. In terzo luogo, vi sono invece moltissime *policies* (*terms of services* e *community standards*) elaborate da attori privati, come le aziende di *Information and Communication Technology (ICT)*, che gestiscono i *social media* e le piattaforme su cui l'intelligenza artificiale opera. Anche queste *policies* configurano una sorta di *soft law*, benché elaborata da soggetti privati e caratterizzata da peculiarità (e correlate questioni giuridiche) che si vedranno nel prosieguo.

Questo articolo si propone di discutere l'impatto di tale complesso *background* sul concetto teorico di "fonte del diritto", inteso nella sua

[3] In genere ciò avviene grazie al riconoscimento delle *keywords* o al sistema degli *hashes*. E.J. Lansó, *No amount of "AI" in content moderation will solve filtering's prior restraint problem*, in *Big Data & Society* 1 (2020).

[4] Sul tema, C. Walker, *The War of Words with Terrorism: An Assessment of Three Approaches to Pursue and Prevent*, in 22 *Journal of Conflict & Security Law* 1 (2017).

[5] Nel presente contributo si parla di "collaborazione" fra il settore pubblico e quello privato. Tale termine si preferisce a quelli – pure ricorrenti – di "cooperazione" o "*partnership*", poiché si ritiene renda maggiormente l'idea di due parti (quella pubblica e quella privata) che compartecipano al raggiungimento di un obiettivo, senza (eccessive) interferenze nelle rispettive funzioni. Cfr. J. Ridaura Martínez, *La colaboración entre seguridad pública y seguridad privada en materia de terrorismo*, in J. Lozano Miralles (coord.), *La lucha contra el terrorismo en el marco del sistema de seguridad nacional*, Cizur Menor, 2021, 313.

accezione di fonte-atto legale di matrice politica[6], sia essa vincolante o meno e a prescindere dal livello di governo che la adotti (internazionale, sovranazionale, interno[7]). Per facilità di esposizione, nel testo ci si riferisce a queste fonti come a "fonti tradizionali", "ordinarie" o "classiche", distinguibili in *hard law* e *soft law*. Questi aggettivi sono scelti sulla scorta della constatazione che, perlomeno negli ordinamenti dell'Europa occidentale improntati al *civil law*, gli atti con queste caratteristiche sono stati per lungo tempo considerati l'archetipo della nozione di fonte del diritto[8]. Un esame empirico basato sulla comparazione, tanto verticale quanto orizzontale, tra fonti che disciplinano l'intelligenza artificiale usata per individuare e rimuovere i contenuti potenzialmente "pericolosi" del *web* dimostra come, in tale settore, i caratteri "genetici" della nozione di fonte sopra delineata siano soggetti ad una progressiva erosione. Di riflesso, si argomenta a favore del necessario rafforzamento del ruolo delle istituzioni pubbliche e delle fonti da esse adottate.

Al fine di sviluppare l'analisi, il lavoro si divide in tre parti. La prima si concentra sulle poche e aspecifiche fonti tradizionali di carattere vincolante in materia di intelligenza artificiale ed eliminazione di contenuti terroristici sul *web*. L'esame di queste fonti parte dal versante internazionale e arriva a quello interno. La seconda, sempre da un angolo di osservazione multilivello, si focalizza sulle fonti non vincolanti (*soft law* in senso "proprio"), adottate nel corso del tempo dalle istituzioni per orientare l'azione degli organi politici e degli operatori tecnici. Sempre in quest'ottica si analizzano le fonti

[6] Si intende il termine "legale" in opposizione a "*extra ordinem*", che indica quelle fonti adottate al di fuori delle procedure tipiche del diritto (ad esempio, sulla base dello stato di necessità). Al proposito, S. Romano, *Sui decreti-legge e lo stato di assedio in occasione del terremoto di Messina e di Reggio-Calabria*, in *Rivista di diritto pubblico e della pubblica amministrazione in Italia*, 1909, 251.
La caratterizzazione come "fonte atto" è opposta alla categoria delle "fonti fatto", in cui rientrano le fonti (consuetudini, convenzioni, precedenti giudiziari) che derivano da attività non direttamente volta ad innovare il diritto. V., per questa distinzione, V. Crisafulli, *Lezioni di diritto costituzionale*, vol. 2, Padova, 1978, 32. Spec. sul precedente giudiziario, v. P. Calamandrei, *Appunti sulla sentenza come fatto giuridico* (1932), in *Opere giuridiche*, Napoli, 1965, 270.
La riconduzione alla matrice politica vuole porre l'accento sulle fonti prodotte da organi istituzionali e, pur in senso lato, ricollegati all'indirizzo politico. Ciò esclude dall'oggetto di questa ricerca un confronto con quelle fonti in cui prevale la componente religiosa, giurisprudenziale, consuetudinaria. Sulle fonti di tale derivazione, R. Sacco, *Diritti stranieri e sistemi di diritto contemporaneo*, in *Enciclopedia giuridica*, vol. XI, Roma, 1989, 6; L. Pegoraro, A. Rinella, *Costituzioni e fonti del diritto*, Torino, 2017, 19, 44.
Per uno studio comparatistico su sistemi e categorie di fonti, F. Palermo, *Le fonti normative*, in P. Carrozza, A. Di Giovine, G.F. Ferrari (cur.), *Diritto costituzionale comparato*, Roma-Bari, 2019, 952.
[7] Sui collegamenti tra la sistematica delle fonti nazionali, italiane o straniere, e quella di altri ordinamenti, quali quello internazionale e quello dell'Unione europea, A. Pizzorusso, *Fonti del diritto*, in A. Scialoja, G. Branca (cur.), *Commentario del Codice Civile*, Bologna-Roma, 2011, 896.
[8] In questo senso, L. Pegoraro, A. Rinella, *op. cit.*, 147, che parlano di «pervasività degli archetipi occidentali».

di "autoregolamentazione" (*soft law* "dei privati", al di fuori dell'idea classica di fonte), che gli operatori del *web* adottano per "normare" il proprio operato sulle relative piattaforme. La terza parte discute le implicazioni di tale panorama sul concetto di "fonte del diritto" nell'accezione spiegata. Seguono alcune riflessioni conclusive sul futuro della regolamentazione della rimozione dei contenuti terroristici *online* e, più in generale, dell'intelligenza artificiale a tutela della sicurezza pubblica.

2. Intelligenza artificiale, rimozione dei contenuti terroristici online e fonti giuridiche vincolanti: lo stato dell'arte

L'esame delle fonti vincolanti internazionali, sovranazionali e interne in materia di rimozione dei contenuti terroristici *online* evidenzia poche e spesso vaghe disposizioni sull'uso degli algoritmi. Anzi, questi ultimi non sono vero e proprio oggetto di regolazione, ma piuttosto appaiono come una scelta necessitata da parte della piattaforma *web*, che deve ricorrere all'automazione per procedere alla rimozione entro i brevi termini imposti dalle normative. La presente sezione dimostra quanto appena asserito.

2.1. La (scarna) regolamentazione internazionale del fenomeno

Il tema della rimozione dei contenuti terroristici sul *web* – e, ancor più, il peculiare aspetto dell'intelligenza artificiale strumentale a tale scopo – è poco regolamentato dalle fonti adottate da organismi internazionali dotati della possibilità di approvare strumenti giuridici vincolanti.

Conviene prendere in esame in prima battuta il quadro giuridico offerto dalle Nazioni Unite (nel prosieguo, ONU), organizzazione internazionale a carattere universale per eccellenza. Tra le risoluzioni del Consiglio di Sicurezza adottate *ex* capitolo VII della Carta ONU[9] ve ne sono alcune che si focalizzano sui rischi posti dallo sfruttamento di Internet da parte dei terroristi (c.d. *e-terrorism*) ed enfatizzano l'importanza della collaborazione fra il settore pubblico e quello privato per affrontare in maniera efficiente il problema[10]. Tuttavia, queste fonti mai fanno riferimento, neppure implicito, al possibile utilizzo di strumenti automatizzati – come gli algoritmi di intelligenza artificiale – né costruiscono un quadro giuridico chiaro e utile agli Stati al fine di gestire la rimozione dei contenuti terroristici *online*. Anzi, le risoluzioni si limitano perlopiù a richiamare – senza, però, istituzionalizzarli – alcuni accordi informali e attuati su base volontaria fra le più importanti aziende tecnologiche (Facebook, YouTube, ecc.). Queste ultime, grazie a tali schemi di cooperazione, si impegnano a fare il possibile per evitare – anche grazie a sistemi automatizzati, previsti dalle loro *policies*

[9] Trattasi delle risoluzioni giuridicamente vincolanti nei confronti degli Stati membri.
[10] Consiglio di Sicurezza delle Nazioni Unite, risoluzione n. 21129 del 17 dicembre 2013; *id.*, risoluzione n. 2354 del 24 maggio 2017; *id.*, risoluzioni nn. 2395 e 2396 del 21 dicembre 2017.

– che i loro spazi virtuali ospitino contenuti pericolosi[11].

Di conseguenza, quello della rimozione dei contenuti sembra uno dei pochi ambiti in cui il Consiglio di Sicurezza ONU non ha (ancora?) esercitato la propria *vis* espansiva, imponendo obblighi giuridici agli Stati membri[12].

Neppure si riscontrano, sempre nel diritto internazionale, trattati o convenzioni che si occupino nel dettaglio dell'argomento.

In relazione al diritto internazionale di portata regionale[13], si può soffermarsi sul Consiglio d'Europa. Quest'ultimo non ha adottato, finora, specifiche fonti di *hard law* in materia di rimozione di contenuti terroristici *online*. Vi è invero la Convenzione del Consiglio d'Europa sul Cybercrime (2001)[14], che disciplina la lotta al terrorismo via *web*, ma non affronta direttamente l'argomento della rimozione dei messaggi terroristici con tecniche automatizzate intelligenti. Va però notato che sono attualmente in atto i negoziati per un protocollo addizionale a questa Convenzione, il quale terrebbe in considerazione la necessità di porre in essere «direct cooperation with providers»[15] e dedicherebbe un articolo alle decisioni automatizzate, seppur solo limitatamente al settore delle c.d. *e-evidence*. Sempre in seno al Consiglio d'Europa, fa poi da sfondo, in qualità di strumento vincolante che si occupa di intelligenza artificiale, ma non precipuamente della rimozione di contenuti terroristici, la Convenzione 108 sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale del 1981[16]. Essa è stata peraltro aggiornata nel 2018, con un protocollo emendativo[17] che rafforza alcune tutele del soggetto sottoposto a decisione automatizzata, come il diritto a conoscere la logica del trattamento dei dati attuato dalla macchina e a non essere soggetti a decisioni puramente automatizzate[18].

---

[11] Come il Global Internet Forum to Counter Terrorism, creato nel 2017. M. Conway et al., *Disrupting Daesh: Measuring Take Down of Online Material and Its Impacts*, in 42 *Studies in Conflict and Terrorism* 141 (2018).

[12] Sul ruolo del Consiglio di Sicurezza ONU nei primi vent'anni dopo gli attacchi dell'11 settembre, v. i diversi contributi pubblicati in A. Vedaschi, K.L. Scheppele (eds), *op. cit.*

[13] Il termine qui è usato in opposizione a "universale", che identifica anzitutto le Nazioni Unite, per indicare le convenzioni internazionali gestite nell'ambito di sistemi che nascono nella sfera di particolari aree geografiche (perciò, regionali). A. Cassese, P. Gaeta, *Diritto internazionale*, Bologna, 2013.

[14] C.E.T.S. 185. Tale Convenzione è stata aperta nel 2001 alle firme sia degli Stati membri del Consiglio d'Europa sia di Paesi terzi e ha ricevuto, nel momento in cui si scrive, quarantotto ratifiche.

[15] Second Additional Protocol to the Convention on Cybercrime on enhanced cooperation and disclosure of electronic evidence, Draft Protocol version 2, 12 April 2021, rm.coe.int/2nd-additional-protocol-budapest-convention-en/1680a2219c.

[16] C.E.T.S. 108.

[17] C.E.T.S. 223.

[18] In tale disposizione risuona l'eco dell'art. 22 del 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, G.U.U.E. L 119 del 4.5.2016, 1-88 (c.d. General Data Protection Regulation).

Va segnalato che l'erosione delle fonti tradizionali di *hard law* raggiunge, al livello internazionale universale e regionale, un grado medio-alto. Ciò in quanto le poche fonti esistenti regolano più gli aspetti generali dell'intelligenza artificiale che i suoi utilizzi particolari.

2.2. La regolamentazione sovranazionale del fenomeno

Dal piano internazionale (universale e regionale), si può spostarsi su quello sovranazionale[19] e prendere in considerazione l'Unione europea (UE).

Vi sono almeno due recenti strumenti giuridici vincolanti che si soffermano sui contenuti terroristici in rete e sulla loro rimozione.

In primo luogo, va richiamata la direttiva (UE) 2017/541[20]. Essa è adottata dalle istituzioni dell'Unione in risposta ai sempre più frequenti attentati in area europea, perpetrati da gruppi terroristici riconducibili all'estremismo islamico. La direttiva non disciplina unicamente la rimozione di contenuti terroristici *online*. Al contrario, la si potrebbe definire una legislazione *omnibus*, perché le sue disposizioni toccano moltissimi ambiti di rilievo per la lotta al terrorismo internazionale: dalle indicazioni agli Stati membri sugli elementi costitutivi delle fattispecie di reati terroristici agli atti preparatori; dalla tutela delle vittime e dei loro congiunti al contrasto ai *foreign fighters*; dal finanziamento delle organizzazioni terroristiche alla radicalizzazione in rete[21].

Per quanto interessa questo studio, la direttiva (UE) 2017/541 si concentra sull'eliminazione dei messaggi terroristici sul *web* all'art. 21, il quale impone agli Stati membri di assicurare «la tempestiva rimozione dei contenuti online ospitati nel loro territorio che costituiscono una pubblica provocazione per commettere un reato di terrorismo», nonché di adoperarsi al fine di far rimuovere gli stessi contenuti ospitati al di fuori del loro territorio. La direttiva, però, nulla dice sul ricorso a strumenti automatizzati, nel concreto ampiamente usati. Questa fonte, quindi, affronta sì la rimozione dei contenuti pericolosi di stampo terroristico, ma lo fa in maniera blanda, sorvolando sul tema dei *tools*.

---

[19] Il termine è qui inteso nel senso di organizzazioni che si differenziano da quelle di diritto internazionale puro per il più marcato grado di integrazione tra i membri. Secondo questa definizione è da includere nel novero delle organizzazioni sovranazionali l'Unione europea. P. Costanzo, *Organizzazioni internazionali e sovranazionali in Europa (dalla "guerra fredda" al "confronto" per la crisi georgiana)*, Relazione al seminario di studi Italo-Brasileiro de Direito Constitucional (28 agosto-4 settembre 2008, Recife e Salvador di Bahia), in *Consulta OnLine*, 2008, 1.

[20] Direttiva (UE) 2017/541 del Parlamento europeo e del Consiglio, del 15 marzo 2017, sulla lotta contro il terrorismo e che sostituisce la decisione quadro 2002/475/GAI del Consiglio e che modifica la decisione 2005/671/GAI del Consiglio, G.U.U.E. L 88 del 31.3.2017, 6–21.

[21] Sulla direttiva 2017/541, G. De Minico, *La risposta europea al terrorismo del tempo ordinario: il lawmaker e il giudice*, in *Osservatorio sulle fonti*, 2/2017, 1; M.E. Gennusa, *Tutto in una definizione? La nuova direttiva antiterrorismo dell'Unione europea e i confini del terrorismo*, in *Quaderni costituzionali*, 3/2017, 651.

Più specifico è il regolamento (UE) 2021/784[22], di recente adozione e, nel momento in cui si scrive, non ancora applicabile negli Stati membri. Questo regolamento ha avuto un *iter* di approvazione lungo – essendo stato proposto dalla Commissione nel settembre 2018 e adottato il 29 aprile 2021 – e controverso. Tra i punti maggiormente discussi vi è proprio il ricorso a meccanismi automatizzati per la rimozione dei messaggi terroristici.

La prima bozza del regolamento, presentata dalla Commissione, permetteva esplicitamente la rimozione dei contenuti con strumenti automatizzati, anche in maniera proattiva[23] da parte dei *service providers*. Questa prima versione della proposta non poneva, a mo' di "contropartita", alcuna garanzia di revisione umana dell'operato della macchina, né di non obbligatorietà dell'automazione. A seguito di aspre critiche sia in sede di *iter* legislativo[24] sia, più in generale, mosse da gruppi di esperti[25] e da associazioni a tutela dei diritti[26], sono stati inseriti due *caveat* all'art. 5 del regolamento. Il primo consiste nel vincolo di sottoporre a «verifica umana» quanto svolto dall'algoritmo. Il secondo si sostanzia nell'esenzione, per il *provider*, dall'obbligo di monitorare costantemente i contenuti ospitati e di farlo grazie al ricorso all'automazione. In altre parole, il *provider* è tenuto – in tempi stretti, ossia entro un'ora dalla segnalazione[27] – a rimuovere un contenuto

[22] Regolamento (UE) 2021/784 del Parlamento europeo e del Consiglio, del 29 aprile 2021, relativo al contrasto della diffusione di contenuti terroristici online, G.U.U.E. L 172 del 17.5.2021, 79–109. Il regolamento in parola diverrà applicabile a partire dal 7 giugno 2022, ossia un anno esatto dopo la sua entrata in vigore. V. art. 24, regolamento (UE) 2021/784, *cit.*

[23] Ossia a prescindere da un ordine di rimozione proveniente da un'autorità pubblica.

[24] Vedasi il parere della Fundamental Rights Agency (FRA), richiesto dal Comitato LIBE ed espresso nel febbraio 2019. FRA, *Proposal for a Regulation on preventing the dissemination of terrorist content online and its fundamental rights implications. Opinion*, 12.2.2019 (fra.europa.eu/sites/default/files/fra_uploads/fra-2019-opinion-online-terrorism-regulation-02-2019_en.pdf).

[25] V. la lettera inviata alla Commissione da tre Special Rapporteurs delle Nazioni Unite (Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; Special Rapporteur on the right to privacy; Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism) il 7 dicembre 2018. Il testo della missiva è consultabile in spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=24234.

[26] V. la lettera inviata da una serie di associazioni a tutela dei diritti (fra cui Amnesty International e Human Rights Watch) il 25 marzo 2021 al Parlamento europeo. Il testo è consultabile in www.hrw.org/news/2021/03/25/joint-letter-eu-parliament-vote-against-proposed-terrorist-content-online. V., inoltre, Voxpol, "The EU Terrorist Content Regulation: Concerns about Effectiveness and Impact on Smaller Tech Platforms", 1.7.2020 (www.voxpol.eu/the-eus-terrorist-content-regulation-concerns-about-effectiveness-and-impact-on-smaller-tech-platforms/).

[27] Tale *deadline* così breve rende assai probabile, se non altro per motivi pratici, il ricorso all'automazione. M. Scheinin, *AI and the Presumption of Terrorist Nature of Expression: EU Regulation 2021/784 on Terrorist Content Online* (intervento di presentazione di *working paper* nell'ambito di ICON-S Mundo, 7 luglio 2021).

DPCE online
ISSN: 2037-6677

terroristico su ordine di una «autorità competente» di uno Stato membro[28], la quale avrà identificato tale messaggio come "terroristico" sulla base della definizione data dalla direttiva (UE) 2017/541; può (ma non deve) adottare misure proattive per individuare e rimuovere messaggi terroristici (ritenuti tali sulla base delle proprie *policies*); nel farlo, può (ma non deve) usare strumenti automatizzati.

Questa breve ricostruzione mostra un'erosione di stadio "medio" delle fonti tradizionali di tipo *hard law*, con l'UE che ha provato a disciplinare il tema, ma senza entrare nei dettagli, tecnici e giuridici, dell'impiego dell'intelligenza artificiale nel settore oggetto di normazione[29].

2.3. Alcuni tentativi di regolamentazione interna: il caso tedesco (Netzwerkdurchsetzungsgesetz 2017) e quello francese (Loi Avia 2020)

Si è visto che il regolamento UE sui contenuti terroristici *online* ha avuto un *iter* lungo e tortuoso. Forse è proprio per questo motivo che gli Stati membri non hanno disciplinato dettagliatamente la rimozione di contenuti terroristici su Internet grazie all'intelligenza artificiale. In altri termini, è possibile che i Legislatori interni abbiano preferito attendere il Legislatore dell'Unione, il quale era in procinto di adottare una normativa direttamente applicabile, invece di agire autonomamente con atti domestici che, successivamente, sarebbero stati a rischio di disapplicazione. Infatti, pur esistendo negli Stati membri delle disposizioni sull'eliminazione dei contenuti terroristici *online*, si tratta di fonti che non si soffermano sul ruolo dei privati e, in particolare, sull'approccio all'automazione. Germania e Francia, due Paesi che hanno legiferato autonomamente (o hanno tentato di farlo, v. *infra*) prima dell'adozione del regolamento (UE) 2021/784 offrono un interessante caso di studio. Pare dunque opportuno analizzare brevemente le rispettive legislazioni.

Il Netzwerkdurchsetzungsgesetz[30] è una legge del Bundestag tedesco

[28] Tali autorità non sono state ancora definite. Ogni Stato membro è tenuto ad individuare e comunicare la propria entro il 7 giugno 2022. La lista delle autorità competenti verrà pubblicata sulla Gazzetta Ufficiale dell'UE.

[29] È pure da notare che la Commissione ha, nell'aprile 2021, avanzato una proposta di regolamento in materia di intelligenza artificiale. Questa, se approvata, costituirebbe una normativa di base, anch'essa non settoriale per il *counter-terrorism*, ai cui principi queste operazioni, condotte con mezzi di intelligenza artificiale, dovrebbero sottostare. V. Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, COM(2021) 206 final.

[30] Netzwerkdurchsetzungsgesetz vom 1. September 2017 (BGBl. I S. 3352). Una traduzione in inglese della legge è disponibile al seguente indirizzo: germanlawarchive.iuscomp.org/?p=1245. V. sull'argomento J. Rinceanu, *Verso una forma di polizia privata nello spazio digitale? L'inedito ruolo dei* provider *nella disciplina tedesca dei* social network, in *Sistema penale*, 2021, 1. Questa legge è stata emendata già due volte, nel 2020 e nel 2021. V. Gesetz zur Änderung des Netzwerkdurchsetzungsgesetzes vom 9. Juni 2021 (BGBI I S. 1436), con l'inserimento, tra l'altro, di una procedura di appello nei confronti

entrata in vigore il 1° ottobre 2017. Al fine di favorire la trasparenza dei *media* digitali e contrastare i c.d. discorsi dell'odio[31], essa impone pesanti sanzioni nei confronti di quei *social networks* che non intervengano in maniera celere – entro 24 ore dalla segnalazione – a rimuovere qualsiasi contenuto «offensichtlich rechtswidrigen» ("manifestamente illegale")[32] sulla base delle disposizioni del codice penale tedesco che puniscono non solo l'incitamento al terrorismo, ma pure altre tipologie di reati collegati all'*hate speech*. La segnalazione può provenire tanto da autorità pubbliche quanto da privati (altri utenti) e deve essere valutata dal gestore della piattaforma stessa. Deve allora esserci un meccanismo (con ogni probabilità automatizzato, dato il poco tempo a disposizione) che verifichi se il messaggio pubblicato integri – manifestamente – la fattispecie di reato di cui al codice penale. Si noti che il Netzwerkdurchsetzungsgesetz dovrà essere parzialmente disapplicato[33], una volta che il regolamento UE sulla rimozione dei contenuti diventerà applicabile, perché la tempistica delle 24 ore è più dilatata rispetto a quella di un'ora, imposta dalla legislazione eurounitaria.

Una disciplina simile a quella tedesca è stata adottata in Francia con la legge n. 2020-766[34] (c.d. Loi Avia, dal nome della deputata proponente il relativo progetto di legge). Secondo questa normativa, i *social media* avrebbero dovuto rimuovere – entro 24 ore dalla segnalazione, dunque quasi sicuramente con mezzi automatizzati – i contenuti segnalati dalla polizia o dagli utenti come messaggi che costituiscono «provocation à des actes de terrorisme» o «apologie de tels actes» sulla base delle disposizioni del codice penale. Tuttavia, adíto in via preventiva circa la costituzionalità della legge, il Conseil Constitutionnel ha dichiarato queste disposizioni in violazione della libera manifestazione del pensiero, poiché la valutazione è rimessa all'apprezzamento della polizia, quando non addirittura dell'operatore tecnologico, senza previo scrutinio di un giudice né di un'autorità amministrativa indipendente[35]. Alla luce di questa decisione, sarà

delle misure adottate direttamente dagli enti privati. Si segnala, inoltre, che anche in Austria, sulla scia di quanto fatto in Germania, è stato adottato il Kommunikationsplattformen-Gesetz (BGBl. I Nr. 151/2020), anch'esso piuttosto scarno per quanto riguarda la specifica regolamentazione degli strumenti automatizzati usati per la rimozione dei contenuti pericolosi. Sulla proposta della normative austriaca, v. M.J. Riedl, *A primer on Austria's 'Communication Platforms Act' draft law that aims to rein in social media platforms*, in *LSE Blog*, 14 September 2020 (blogs.lse.ac.uk/medialse/2020/09/14/a-primer-on-austrias-communication-platforms-act-draft-law-that-aims-to-rein-in-social-media-platforms/).

[31] Nella cui categoria è possibile includere il discorso terroristico. G. Rollnert Liern, *Incitación al terrorismo y libertad de expresión: el marco internacional de una relación complicada*, in *Revista de Derecho Político*, 2014, vol. 91, 250 ss.

[32] Sez. 3.2, Netzwerkdurchsetzungsgesetz, cit. Sono invece da rimuovere entro 7 giorni i contenuti semplicemente «rechtswidrigen» ("illegali").

[33] Oppure il Legislatore tedesco potrebbe decidere di modificarlo prima del 7 giugno 2022.

[34] Loi n° 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet.

[35] Conseil Constitutionnel, décision n° 2020-801 DC du 18 juin 2020. P. Mouron, *La censure*

DPCE online
ISSN: 2037-6677

interessante vedere come il Conseil si relazionerà con il nuovo regolamento dell'Unione europea, che, nell'indicare un'«autorità competente» come soggetto deputato a segnalare i contenuti da rimuovere (obbligatoriamente), non ne impone il carattere giurisdizionale (o, perlomeno, amministrativo indipendente).

È proprio sul piano nazionale che il grado di erosione delle fonti tradizionali di *hard law* risulta manifesto, poiché il diritto interno, pur dove presente, sembra – giustamente – preoccuparsi affinché il messaggio radicalizzante *non si diffonda*, ma non interessarsi dei *mezzi tecnologici* da usare per raggiungere lo scopo.

3. La regolamentazione dell'intelligenza artificiale a fini di rimozione di contenuti terroristici online: fonti di *soft law* "tradizionale" e "dei privati"

Visto che le fonti tradizionali vincolanti volte a regolare l'eliminazione dei contenuti terroristici *online* sono poche e vaghe, è utile comprendere cosa avviene a livello di *soft law*. Il concetto di *soft law* rimanda a quegli atti con valore orientativo e non giuridicamente vincolante[36].

Il presente paragrafo studia, in prima battuta, se e in quale misura tali strumenti non vincolanti siano stati approvati da istituzioni pubbliche (siano esse internazionali, sovranazionali o nazionali). Si rimane nel novero delle fonti classiche e si fa riferimento alla *soft law* "tradizionale". In seconda battuta, si considera la regolamentazione del *topic* da parte di atti anch'essi non vincolanti, almeno formalmente, ma adottati dai soggetti che operano in rete (*providers*, *social media* e altre piattaforme di comunicazione e condivisione di contenuti via *web*)[37]. Per indicare queste fonti che esorbitano dal perimetro delle fonti classiche si parla, in questo scritto, di *soft law* "dei privati".

3.1. Fonti di *soft law* "tradizionale": prospettive multilivello

Le fonti di *soft law* tradizionale, non vincolanti ma adottate da soggetti istituzionali, proliferano in materia di intelligenza artificiale, ma non si soffermano molto sui suoi usi nel *counter-terrorism*.

(prévisible) de la loi Avia par le Conseil constitutionnel, in *Revue européenne des médias et du numérique*, 2020, 17. È tuttavia opportuno ricordare che, in Francia, è stata approvata la loi n° 2021-1109 du 24 août 2021 confortant le respect des principes de la République (anche detta *loi contre le séparatisme*), la quale riprende alcune disposizioni della loi Avia. Sulla proposta di questa legge, v. P. d'Iribarne, *Un projet de loi*, in *Commentaire*, 2/2021, 426.

[36] E. Mostacci, *La* soft law *nel sistema delle fonti*, Padova, 2008.

[37] Va dato conto che, secondo un certo orientamento, tali atti, pur non essendo di matrice pubblicistica, possono essere inquadrati nell'idea di *soft law*. Per tale dottrina, in questa nozione non rientrerebbero tanto gli atti non vincolanti, quanto quelli che costituiscono un'autoregolamentazione dei soggetti coinvolti (di matrice non governativa e, più ampiamente, non pubblicistica). J.J. Kirton, M.J. Trebilcock, *Introduction: Hard Choices and Soft Law in Sustainable Global Governance*, in J.J. Kirton, M.J. Trebilcock (eds), *Hard Choices, Soft Law: Voluntary Standards in Global Trade, Environment and Social Governance*, Aldershot, 2004, 9.

Non è possibile – né sarebbe utile ai fini dell'analisi – elencare tutte queste fonti. Basti ricordare, sul piano internazionale e sovranazionale, gli *OECD Principles on AI*, approvati nel 2019 dagli Stati membri dell'Organizzazione per la cooperazione e lo sviluppo economico, gli *Orientamenti etici per un'intelligenza artificiale affidabile*, adottati nel 2019 dal Gruppo indipendente di esperti sull'intelligenza artificiale, istituito dalla Commissione europea nel giugno 2018[38], e le *Linee guida in materia di intelligenza artificiale e protezione dei dati personali* del Comitato consultivo del Consiglio d'Europa per la protezione degli individui nell'ambito delle operazioni di trattamento dei dati, risalenti anch'esse al 2019[39]. Gli *OECD Principles on AI* raccomandano l'adozione di sistemi automatizzati «human-centred» e sono rivolti sia ai *policy-makers* sia a chi concretamente costruisce meccanismi algoritmici. Hanno gli stessi destinatari gli *Orientamenti etici*, che pongono principi di base a cui, sia in fase di programmazione sia di regolazione, ci si dovrebbe attenere. Tra di essi, vale la pena ricordare la trasparenza, la revisione umana e il rispetto del principio di non discriminazione. Le *Linee guida* del Consiglio d'Europa sono rivolte sia a soggetti pubblici preposti alla regolazione sia ad enti privati incaricati della predisposizione degli algoritmi e sono più focalizzate su una prospettiva di *data protection*.

A livello interno, vanno ricordate, *ex multis*, le linee guida *AI in the UK: ready, willing and able?*[40], a cura del Select Committee on Artificial Intelligence della House of Lords britannica nel 2018, che danno indicazioni su come sviluppare sistemi di intelligenza artificiale – per gli scopi più svariati – nel rispetto dei principi di trasparenza e *accountability*.

Ebbene, nessuna delle menzionate fonti di *soft law* – come pure altre qui non approfondite per ragioni di sintesi[41] – prende in seria considerazione l'*artificial intelligence* finalizzata a identificare e rimuovere messaggi terroristici sul *web*. Vi è solo un rapidissimo cenno, negli *Orientamenti etici*, alla capacità dell'algoritmo di «identificare una persona […] ad esempio nei casi di […] finanziamento del terrorismo»[42]. È indubbio che molte raccomandazioni generali contenute in questa *soft law* siano applicabili agli

[38] Consultabili in op.europa.eu/it/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1#:~:text=Un'IA%20affidabile%20si%20basa,punto%20di%20vista%20tecnico%20e.

[39] Consultabili in rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8.

[40] Consultabili in publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.

[41] Conviene dare menzione almeno dei principali. Per l'Unione europea, si vedano: Gruppo indipendente di esperti sull'intelligenza artificiale, *Policy and investment recommendations for trustworthy Artificial Intelligence*, 2019; Id., *The assessment list for trustworthy Artificial Intelligence*, 2019; Id., *Sectoral Considerations on the Policy and Investment Recommendations*, 2019. Per il Consiglio d'Europa: Comitato dei Ministri, *Recommendation to member States on the human rights impacts of algorithmic systems*, CM/Rec(2020); Directorate General of Human Rights and Rule of Law, *Guidelines on facial recognition*, 2021.

[42] *Orientamenti etici*, cit., par. 130.

utilizzi in parola[43], ma la mancanza di un quadro *ad hoc* – pur meramente orientativo e non pienamente vincolante – è indicativo di quanto il *topic* sia lasciato all'intervento di soggetti privati, dissociati dal piano istituzionale. Pertanto, persino quello che si poteva definire come "l'ultimo baluardo" delle fonti classiche, ossia la *soft law* tradizionale, viene ad essere escluso (*rectius*, si autoesclude) dal disciplinare questo peculiare tema.

3.2. La *soft law* "dei privati": uno scenario complesso

L'intervento del settore privato è in effetti preponderante, nonché precedente rispetto ai tentativi, più o meno blandi, di regolazione istituzionale. Bisogna allora comprendere, in prima battuta, *chi siano* i soggetti privati che intervengono in tale materia e, in secondo luogo, *su quali aspetti* essi si focalizzino e *con quali strumenti*.

I soggetti coinvolti sono essenzialmente di tre categorie: i c.d. *Internet service providers*, ossia le aziende preposte a fornire connettività a Internet; gli *hosting providers*, cioè quelle imprese in grado di collocare su un *server* i dati relativi ad un sito *web*; e le *ICT companies*, società che offrono la possibilità di pubblicare contenuti e/o comunicare con altri utenti *online*.

Con riguardo agli ambiti in cui essi intervengono, per quanto interessa questo studio, questi soggetti tendono a regolamentare le definizioni (che cos'è il messaggio terroristico o, secondo diversa terminologia, radicalizzante)[44], le conseguenze dell'eventuale caricamento o condivisione in rete di contenuti rientranti in tali definizioni[45], i mezzi di reclamo esperibili dall'utente nel caso in cui ritenga che un proprio contenuto sia stato ingiustamente o per errore eliminato[46]. Tali *policies* non rinviano a fonti istituzionali per l'aspetto definitorio o per quello della fattispecie che andrà rimossa, mentre sono spesso esplicite circa gli algoritmi e i casi in cui il mezzo automatico può operare[47].

---

[43] Si pensi al rispetto della *privacy* (*rectius*, al principio di proporzionalità nelle interferenze con la vita privata), alla trasparenza e all'*accountability*.

[44] Ad esempio, stando alle norme della *community* YouTube, sono vietati, tra l'altro: «Contenuti prodotti da organizzazioni criminali violente o terroristiche; […] che elogiano o commemorano figure di spicco degli ambienti terroristici e criminali allo scopo di incoraggiare altri a commettere atti di violenza».

[45] Sempre sull'esempio delle *policies* di YouTube, i contenuti che violano le norme, intercettati da un algoritmo, verranno eliminati e al soggetto che li ha caricati verrà inviata un'email. In caso di ripetute violazioni, l'*account* YouTube – e con esso la possibilità di condividere contenuti – verrà disabilitato.

[46] Come la compilazione di moduli che permettano di esporre quanto accaduto e sottoporre a valutazione dell'azienda ICT stessa (YouTube) oppure la possibilità di "ricorrere" a organismi interni (es. il Facebook Oversight Board).

[47] Facebook, ad esempio, impiega l'intelligenza artificiale per calcolare la probabilità che un *post* contenga un messaggio che supporti il terrorismo e, quando tale probabilità è alta, la rimozione è automatizzata, senza intervento dell'operatore umano. V. Facebook, *Hard Questions: What Are We Doing to Stay Ahead of Terrorists?*, 2018, about.fb.com/news/2018/11/staying-ahead-of-terrorists/.

Gli strumenti con cui queste "norme" (in senso ovviamente lato e atecnico) vengono adottate dai summenzionati soggetti privati sono, di regola, i relativi *terms of services*, contenenti i c.d. *community standards*. Si tratta di documenti elaborati dagli organi delle singole società con la collaborazione di soggetti, solitamente non identificati, che si qualificano come esperti del settore[48].

Dunque, sono proprio questi atti non ascrivibili alle fonti classiche a giocare il ruolo di *dominus*, rimpiazzando il regolatore pubblico che si ritrae.

4. L'uso dell'intelligenza artificiale per la rimozione dei contenuti terroristici *online*: impatto sul concetto di "fonte del diritto"

Dinanzi a tali constatazioni, è opportuno domandarsi quali siano le ripercussioni sull'idea di fonte del diritto come richiamata in Introduzione. Per riflettere sul tema, la presente sezione prende atto che il *background* sopra presentato evidenzia l'assorbimento di (rilevanti) "porzioni" di potere pubblico da parte di entità private, quali *providers* e *ICT companies*. In particolare, in questa sede ci si occupa dell'"esternalizzazione" del potere legislativo (inteso in senso ampio come "normativo", ossia di porre norme generali e astratte). In parole più chiare, i privati, nel definire nelle proprie *policies* (*terms of services*, *community standards*) le fattispecie in cui un certo messaggio va rimosso in quanto "terroristico", assumono i connotati di inediti Legislatori, che non si muovono nelle aule delle assemblee legislative, bensì estrinsecano la propria opera sul *web*. I paragrafi seguenti si concentrano unicamente sul *law-making* e non esaminano le ulteriori "esternalizzazioni", sempre a favore dei c.d. giganti del *web*, di potere esecutivo – nel rimuovere concretamente i contenuti – o giudiziario – nel decidere su eventuali reclami[49]. Né vengono approfondite le pur importanti conseguenze sulla libera manifestazione del pensiero o sulla *privacy*, che esulano dal *focus* del lavoro[50].

4.1. La "privatizzazione" del potere legislativo: uno scenario (quasi) senza precedenti

Si è detto che *providers* ed altri operatori del *web*, costruendo e impiegando

[48] Si legge sulla pagina YouTube dedicata alle norme della *community* che esse sono «sviluppate in collaborazione con numerosi esperti di norme indipendenti specializzati nel settore, oltre ai creator di YouTube».

[49] Per approfondimenti, A. Vedaschi, *Sicurezza e diritti nella digital age. La tecnologia: un'arma a doppio taglio nella lotta al terrorismo internazionale*, in A. Somma, L. Lloredo Alix (cur.), *Scritti in onore di Mario G. Losano. Dalla filosofia del diritto alla comparazione giuridica*, Torino, [illegible] 518. V. inoltre K.E. Eichensehr, *Public-Private Cybersecurity*, in 95 *Texas Law Review* 467 (2017).

[50] Su questi profili, R. Cohen-Almagor, *The Role of Internet Intermediaries in Tackling Terrorism Online*, in 86 *Fordham Law Review* 425 (2017).

DPCE online
ISSN: 2037-6677

algoritmi di rimozione dei contenuti pericolosi per la sicurezza nazionale, esercitano funzioni sostanzialmente legislative. Si rinvengono almeno tre caratteri propri del *law-making*. Il primo, già richiamato, è la generalità e astrattezza delle "norme" (*lato sensu*) imposte dall'operatore tecnologico. Nel definire quali contenuti vanno rimossi e in che circostanze, non si agisce per affrontare un caso specifico, ma si delinea una fattispecie generale. Il secondo è che tali "norme" sono, nei fatti, vincolanti. La natura vincolante delle "norme" in parola non deriva – come accade con le fonti tradizionali – dal loro fondamento su una fonte gerarchicamente sovraordinata che le rende formalmente valide[51] e idonee a vincolare i singoli (dopo eventuali passaggi integrativi dell'efficacia). Diversamente, essa parrebbe discendere dalla forza contrattuale di *terms of service* e *community standards*: se non si accettano o si violano le "regole del gioco", non si può usufruire dei servizi messi a disposizione dalle aziende della tecnologia. Il terzo carattere comune è costituito dalle ripercussioni sulla sfera giuridica del singolo, cosa che accade sia per le fonti tradizionali – le quali esplicano i propri effetti giuridici in capo ai soggetti ricadenti nel loro ambito di applicazione – sia per le norme "atipiche" dei privati, potenzialmente implicanti la rimozione di un messaggio e, quindi, la limitazione della libera espressione individuale.

Le fonti "dei privati" non condividono con quelle tradizionali il fatto di essere il risultato – perlomeno a partire dallo stato liberale – di un processo deliberativo che si possa dire, in modo più o meno marcato, "democratico". Le fonti "classiche" possono essere dotate di un più evidente carattere rappresentativo – come la legge, che è di regola il prodotto dell'assemblea rappresentativa – oppure tale caratteristica può essere meno spiccata – si pensi alle fonti di adozione dell'esecutivo, o addirittura unicamente di alcuni suoi esponenti. Pure in questi ultimi casi, però, vi è un *link* tra quanto approvato e la maggioranza politica, cosa che li inserisce, almeno parzialmente, all'interno del circuito democratico rappresentativo. Tale *trait d'union* non esiste, come si è visto, nelle *policies* dei giganti della tecnologia, in cui l'*iter* di produzione e i soggetti coinvolti restano, il più delle volte, oscuri, minando peraltro i principi di trasparenza e di pubblicità che dovrebbero caratterizzare l'adozione delle norme vincolanti.

È da chiedersi se questo *shift* di potere legislativo a beneficio di entità private costituisca un *unicum* che contraddistingue solo l'oggetto di questo studio, ossia gli algoritmi per l'identificazione e rimozione dei contenuti terroristici *online*, oppure se vi siano state precedenti esperienze assimilabili. La risposta è piuttosto articolata. Sicuramente vi sono casi in cui norme approvate da soggetti privati hanno assunto un rilievo tale da essere, in pratica, vincolanti *erga omnes* (*rectius*, per tutti coloro che si trovano nel perimetro di applicabilità di queste regole). A titolo di mero esempio, si

[51] Ciò perlomeno a voler abbracciare un'impostazione positivistica di stampo kelseniano. H. Kelsen, *Teoria generale del diritto e dello Stato* (1945), Milano, 1994.

possono citare i principi contabili internazionali (International Accounting Standards – IAS, poi sostituiti dagli International Financial Reporting Standards – IFRS). Gli IAS-IFRS sono stati elaborati dall'International Accounting Standard Committee, un gruppo di esperti in materia di contabilità, e poi recepiti dal diritto comunitario (successivamente, dell'Unione europea)[52]. Il caso degli IAS-IFRS presenta analogie, ma anche differenze con quello degli *standards* per la rimozione dei contenuti terroristici sul *web*. Per quanto riguarda le analogie, si tratta di due circostanze in cui organismi privati, e non soggetti pubblici, sono intervenuti a normare una data situazione. La più significativa differenza sta nel fatto che gli IAS-IFRS sono stati recepiti istituzionalmente (configurando una legislazione *bottom-up*, ossia che ha impulso da una porzione di società civile e viene poi accolta dalle istituzioni con i relativi strumenti normativi). Ciò ha contribuito all'uniformazione delle regole contabili internazionali. Lo stesso non si può dire per l'eliminazione dei contenuti terroristici, soprattutto con strumenti automatizzati[53]. Infatti, non vi sono stati atti di recepimento da parte delle istituzioni che hanno portato ad uniformare i *community standards* delle diverse piattaforme, cosa, tra l'altro, d'impatto negativo sulla parità di trattamento. Anzi, queste *policies*, che proliferano e vengono modificate in maniera fluida e non sempre chiara agli utenti, evidenziano non marginali differenze le une rispetto alle altre, sotto il profilo sia della fattispecie di messaggio da rimuovere, sia degli algoritmi usati, differenti a seconda dell'azienda di cui si parla.

Pertanto, si può dire che l'intricato rovo della normazione di origine "privatistica" dei contenuti pericolosi in rete rappresenti una peculiarità nel panorama del modello di regolazione "ibrido" pubblico-privato, non trovando, nelle recenti esperienze, alcun termine di paragone.

4.2. Un nuovo concetto di "*soft law*" …o, più in generale, di "fonte del diritto"?

Si sono viste, nel § 3.1, talune analogie e differenze tra il processo di elaborazione dei *terms of service* privati e il *law-making* tradizionale. Tali constatazioni riflettono in modo speculare alcune analogie e differenze tra i *terms of service* stessi e le fonti classiche del diritto. Vi sono altri elementi, però, più strettamente attinenti al concetto di fonte e che meritano di essere analizzati per capire come inquadrare questi atti.

Anzitutto, la distinzione tra fonti c.d. dure e fonti c.d. *soft* non sembra tenere. Da un lato, *ictu oculi*, i *terms of service* (che includono, lo si ricorda, i *community standards*) appaiono come atti non vincolanti: si è perciò parlato di *soft law* dei privati, se non addirittura di disposizioni di autonomia

[52] Regolamento (CE) n. 1606/2002 del Parlamento europeo e del Consiglio, del 19 luglio 2002, relativo all'applicazione di principi contabili internazionali, in G.U.C.E. L 243 dell'11.9.2002, 1-4.

[53] C. Graziani, *Removing Terrorist Content Online: The Intersection between the International, Regional, and Domestic Level*, in A. Vedaschi, K.L. Scheppele (eds), *op. cit.*, 222.

DPCE online
ISSN: 2037-6677

contrattuale. Invero, però, ad un esame più attento si nota come essi finiscano per vincolare la generalità dei consociati, data l'amplissima diffusione che i servizi della tecnologia hanno. Si tratterebbe quindi di fonti "ibride", perché rapportabili all'autonomia privata, non vincolanti in apparenza – chi non intende rispettarle potrebbe ben scegliere di non sottoscrivere i servizi *ICT* – ma vincolanti nei fatti, dato che operazioni come la messagistica via *web*, la pubblicazione di contenuti *online*, la connessione ai *social network* fanno ormai parte della vita quotidiana.

In seconda battuta, la natura di questi atti è ibrida sotto il profilo della loro sindacabilità. Normalmente, per le fonti del diritto sono previste procedure volte a contestarne la compatibilità con le fonti sovraordinate. Si pensi ai modelli esistenti di sindacato di costituzionalità[54], ma pure alla possibilità di mettere in dubbio la legalità di atti amministrativi, determinandone l'annullamento da parte del giudice competente[55]. Nessuna di queste eventualità è attuabile per le "fonti" di cui si discute. I *terms of service* di un operatore tecnologico non potrebbero essere sottoposti ad un giudizio di costituzionalità né al vaglio di un giudice amministrativo, che decida di annullarli, o ancora a meccanismi assimilabili. E ciò nonostante siano disposizioni che, soprattutto quando comportano l'impiego di mezzi automatizzati, possono incidere fortemente sulle libertà individuali. Si potrebbe ipotizzare di demandare queste peculiari fonti allo scrutinio di un giudice ordinario, trattandole alla stregua di disposizioni contrattuali ed agendo con i relativi rimedi. Ad esempio, nel caso di interferenza sproporzionata nella *privacy* di un individuo a causa di un algoritmo utilizzato per identificare e rimuovere contenuti, l'utente potrebbe convenire in giudizio il *provider*, denunciare la *non-compliance* con la normativa di *data protection* e chiedere l'annullamento della relativa clausola dei *terms of services*. Nondimeno, è possibile identificare diverse criticità. Un siffatto algoritmo – per ragioni ascrivibili vuoi al segreto industriale vuoi alle sue caratteristiche tecniche[56] – potrebbe non essere conosciuto né conoscibile, rendendo

[54] Sul tema, A. Vedaschi, *La giustizia costituzionale*, in P. Carrozza, A. Di Giovine, G.F. Ferrari (cur.), *op. cit.*, 1087.

[55] Similmente, sul piano UE, si pensi al rinvio pregiudiziale di validità o al ricorso per annullamento. Il discorso è parzialmente diverso per i trattati internazionali. Per essi non esiste una corte che ne dichiari l'invalidità, ma sono pur sempre sottoposti alla volontà di Stati sovrani, che potrebbero in ogni momento recedere, effettuare la denuncia o, in sede giurisdizionale, attivare meccanismi assimilabili ai controlimiti negli Stati ove ciò è previsto. Per uno studio comparato sui controlimiti, G. Martinico, *Is the European Convention Going to Be 'Supreme'? A Comparative-Constitutional Overview of ECHR and EU Law before National Courts*, in 23 *European Journal of International Law* 401 (2017).

[56] Esistono algoritmi così "avanzati" da raggiungere un grado di autonomia per cui neanche chi li programma è in grado di conoscere l'intero processo logico che l'algoritmo applica a fini decisionali. Trattasi dei c.d. *black boxes*. L'uso di questi algoritmi è tipico dell'*intelligence* francese dopo un'importante riforma del 2015, ma non è escluso che essi vengano impiegati in ulteriori contesti. W. Mastor, *The French Intelligence Act: 'The French Surveillance State?'*, in 23 *European Public Law* 707 (2017).

estremamente complicato per il giudice assumere una decisione motivata. In aggiunta, potrebbero porsi non marginali problemi circa gli effetti di una sentenza di una corte. Piattaforme tecnologiche come YouTube, Facebook e molte altre *big tech* lavorano – con i propri algoritmi – su un mercato che non conosce i confini nazionali e, dunque, neanche le regole del riparto di competenza giurisdizionale *ratione loci*. L'eventuale decisione di un giudice statale su un singolo caso, invece, difficilmente si applicherebbe con una logica *cross-border*. Potrebbero, a questo proposito, essere d'aiuto le corti sovranazionali, soprattutto in area europea. La Corte di giustizia dell'Unione europea ha già in altri casi emanato sentenze con ripercussioni sui comportamenti dei giganti del *web* quando operano negli Stati membri UE[57]. Tuttavia, la difficile conoscibilità degli algoritmi sembra imporre considerevoli barriere nei confronti di decisioni giurisdizionali precise e puntuali.

Le criticità riscontrate sono veri e propri campanelli d'allarme, che dovrebbero far propendere per la non inclusione di questi atti di autonomia privata nel novero delle fonti del diritto, né di tipo vincolante né di tipo meramente orientativo. Aprire eccessivamente ai soggetti privati e alle loro regole porrebbe esageratamente in tensione le garanzie connesse ad un ordinario processo di *law-making*.

5. Riflessioni conclusive

L'analisi condotta porta a due principali osservazioni, una di carattere teorico e una di natura pratica.

Sul versante teorico, si è visto che gli evidenziati rischi scaturenti dalla *soft law* dei privati escludono l'ingresso di questi atti in una sistematica ordinaria delle fonti, come si trattasse di una nuova categoria di quelle fonti qui definite tradizionali. In termini più generali, non è la nozione classica di fonte che deve essere dilatata, finanche stravolta per includere forzatamente i *terms of services*. Al contrario, tale concetto, con le connesse garanzie, deve essere salvaguardato, così da fare da spartiacque tra quegli atti che *possono* incidere su situazioni giuridiche soggettive e quelli che non sono abilitati a farlo.

Sul piano pratico, poi, è auspicabile che le fonti classiche del diritto intervengano in modo più incisivo a regolare algoritmi finalizzati a rimuovere i contenuti *web* potenzialmente lesivi della sicurezza. Il *law-maker*

[57] Si pensi al caso *Google Spain*, in tema di diritto all'oblio. Corte di giustizia dell'Unione europea (Grande Sezione), *Google Spain SL e Google Inc. c. Agencia Española de Protección de Datos (AEPD) e Mario Costeja González*, Causa C-131/12, sentenza del 13 maggio 2014. Sul tema, G. Resta, V. Zeno-Zencovich (cur.), *Il diritto all'oblio dopo la sentenza Google Spain*, Roma, 2015. V. la successiva sentenza con cui la Corte di giustizia stessa ha poi limitato l'ambito di applicazione del diritto all'oblio, Corte di giustizia dell'Unione europea (Grande Sezione), *Google LLC c. Commission nationale de l'informatique et des libertés (CNIL)*, Causa C-507/17, sentenza del 24 settembre 2019.

DPCE online
ISSN: 2037-6677

pubblico dovrebbe disciplinare le caratteristiche basiche dei sistemi algoritmici da sfruttare, compresi i loro aspetti tecnici, con regole precise anziché con mere enunciazioni di principio. Opportuna sarebbe anzitutto una normativa internazionale, in modo da far corrispondere alla globalità dell'oggetto la globalità degli strumenti regolatori[58]. A questa andrebbero affiancate convenzioni a carattere internazionale regionale, che tengano conto delle specificità delle singole aree, nonché mirate azioni di matrice sovranazionale. Va altresì valorizzato il ruolo dei Legislatori nazionali, che potrebbero intervenire non solo con il mero recepimento o applicazione delle fonti sovrastatali, ma anche, nei limiti da queste consentiti, per adattarle alle peculiarità dei propri ordinamenti. Nell'*iter* normativo, a tutti i livelli, si potrebbero coinvolgere a vario titolo esponenti del settore privato, sempre facendo in modo che le linee guida da loro elaborate siano recepite e cristallizzate in atti giuridici tradizionali, vincolanti per il maggior numero di Stati possibili. In altre parole, la collaborazione dei privati, sicuramente essenziale, va riportata negli argini di processi decisionali aperti, trasparenti e, in definitiva, democratici.

Gli approcci sinora osservati – si pensi al regolamento UE o alle legislazioni di respiro nazionale richiamate nei precedenti paragrafi – hanno dimostrato una vaghezza che non solo è difficilmente accettabile in un settore così delicato quale quello della sicurezza, ma è pure potenzialmente foriera di ampia discrezionalità dei privati. Tale stato delle cose è sicuramente correlato al fatto che ci si trova in una fase non ancora avanzata dell'uso di algoritmi al fine di contrastare il discorso terroristico sul *web*. Infatti, l'automazione intelligente compie quotidianamente sviluppi difficili da incasellare in atti giuridici formali. Di conseguenza, l'autoregolamentazione dei privati ha costituito, fino a questo momento, una soluzione temporanea, ma deve necessariamente essere superata da una disciplina il più possibile condivisa e sottoposta a procedure deliberative inclusive e rappresentative di tutti i soggetti coinvolti.

Chiara Graziani
Dip.to di Giurisprudenza
Università degli Studi di Milano-Bicocca
chiara.graziani@unimib.it

[58] Sui c.d. sistemi regolatori globali e la possibilità di dare vita ad una «*rule of law*» globale, S. Cassese, *Il diritto amministrativo globale. Una introduzione*, in *Rivista trimestrale di diritto pubblico*, 2005, 331.