RIVISTA ITALIANA DI INFORMATICA E DIRITTO
PERIODICO INTERNAZIONALE DEL CNR-IGSG

ISSN 2704-7318
Fascicolo 2-2021
DOI: 10.32091/RIID0045

Social media, profili artificiali e tutela della reputazione. Come l'avvento dei *social bot* per la gestione dei profili social possa rappresentare una grave minaccia per la reputazione delle persone e quali potrebbero essere le risposte a tale pericolo

Alessandro Tedeschi Toschi • Giampaolo Berni Ferretti

L'utilizzo dei social media ha permesso a chiunque di diffondere il proprio pensiero con una velocità ed una capillarità che eclissano quelle dei grandi organi di stampa tradizionali. All'interno di queste "piazze virtuali" operano anche i cosiddetti *social bot*, dei programmi che, una volta forniti delle credenziali di accesso di un account, sono in grado di gestirlo in autonomia, dando però l'impressione di essere una persona vera. La rapidità e precisione di reazione sui social network rendono questi "strumenti digitali" pericolosamente utili per la diffusione di insinuazioni, maldicenze ed offese con un'ampiezza ed una velocità tali da distruggere irrimediabilmente la reputazione di chiunque venga preso di mira dai loro amministratori. L'attuale impianto giuridico di tutela delle vittime di diffamazione, nonostante le recenti pronunce giurisprudenziali in tema di diffusione online di falsità e l'estensione dell'applicazione di strumenti come il sequestro preventivo all'ambito dei social network, rischia comunque di essere inefficace di fronte ad un utilizzo ben concertato di questi *socialbot*. Nemmeno le contromisure introdotte dai proprietari dei social media sono state in grado di arginare efficacemente questo fenomeno. Risulta, così, necessario un ripensamento della tutela della reputazione che si estenda anche alla prevenzione dall'utilizzo improprio degli strumenti informatici oggi alla portata di chiunque. Si avanzano, quindi, in questo articolo varie ipotesi per risolvere il problema, tra le quali: l'estendere anche ai proprietari dei social media l'obbligo della "adeguata verifica della clientela" già esistente in capo agli operatori finanziari o l'istituzione di un obbligo di approntare un sistema interno di segnalazione ai gestori e di contestuale marcatura, visibile agli altri utenti, dei contenuti reputati diffamanti.

Diffamazione – Social network – Automazione – Bot

Sommario: *1. Introduzione: l'avvento dei social media – 2. I bot, natura e azioni compiute sui social media – 3. I potenziali pericoli rappresentati dai* social bot *sui social media – 4. L'inquadramento della diffamazione sui social media in Italia – 5. La problematica attuazione della tutela della reputazione lesa tramite social media – 6. Tentativi di risposta al problema della presenza dei* social bot *sui social media – 7. La necessità di nuovi strumenti di tutela – 7.1. Riforme legislative – 7.2. Riformulazione delle interpretazioni giurisprudenziali in materia di responsabilità degli ISP – 7.3. Innovazioni digitali – 8. Conclusioni*

A. Tedeschi Toschi è tirocinante presso il Tribunale per i Minorenni di Milano. Oltre alla laurea in giurisprudenza in Italia, ha avuto modo di approfondire le proprie conoscenze informatiche tramite dei corsi della University of Michigan e della Harvard University. G. Berni Ferretti è avvocato iscritto all'Ordine degli Avvocati di Milano ed è presidente dell'associazione culturale senza scopo di lucro "Milano Vapore".

1. Introduzione: l'avvento dei social media

«Ripetete una bugia cento, mille, un milione di volte e diventerà una verità». Questa frase viene spesso attribuita a Joseph Goebbels, ministro della propaganda del Terzo Reich dal 1933 al 1945, e, sebbene per ironia della sorte non si trovino conferme che l'abbia effettivamente pronunciata, rimane comunque emblematica di come un sistema ben organizzato possa distorcere la verità fino a convincere il pubblico che gli avvenimenti siano diversi o addirittura l'opposto della realtà. Se un tempo per raggiungere questi risultati era necessario avere a disposizione un intero apparato burocratico con migliaia di funzionari (il Ministero del Reich per l'istruzione pubblica e la propaganda, noto anche come *Propagandaministerium*, ne contava circa 2.000 nel 1939) oggi, grazie soprattutto alle nuove tecnologie e all'evoluzione delle dinamiche sociali che da esse ne discendono[1], raggiungere i medesimi effetti è diventato estremamente più semplice e decisamente meno costoso. Sono, infatti, i social network, con la loro diffusione globale e la loro facilità di utilizzo, nonché la loro gratuità, ad aver permesso la propagazione di interpretazioni, opinioni ed inesattezze da parte dei loro utenti in tempi e modi impensabili fino a pochi decenni fa[2]. La loro strutturazione, poi, e la facilità di accesso ad essa hanno permesso ad un'ampia pletora di programmatori, anche amatoriali, di sviluppare soluzioni digitali per incrementare ulteriormente la diffusione di contenuti su queste piattaforme di interazione sociale.

Nati come spazi digitali di condivisione e comunicazione, in grado di permettere contatti immediati anche con persone a migliaia di chilometri di distanza, i social network sono diventati soprattutto negli ultimi tempi un terreno fertile per la propaganda ed il proselitismo di tutti coloro che ne hanno intuito le potenzialità e hanno un minimo di comprensione del loro funzionamento[3]. Pur essendo dei servizi informatici volti alla diffusione e alla consultazione di contenuti digitali, quali video, immagini e testi da parte dei loro utenti, alla base delle dinamiche di queste "piazze virtuali" permangono gli stessi principi che hanno governato le interazioni umane nei contesti politici e sociali fin da quando ne abbiamo memoria (il carisma, le capacità oratorie e l'intuizione dei sentimenti dei propri uditori rimangono sempre gli elementi fondanti della fama e del successo di un personaggio pubblico)[4]. In altre parole, i social media hanno portato a dei veri e propri sconvolgimenti nei modi e nei tempi di propagazione del pensiero, grazie a delle architetture digitali sviluppate negli ultimi anni e ospitate su infrastrutture capaci di gestire l'elevata mole di dati generati (si calcola che nel solo corso dell'agosto 2021 siano stati pubblicati 575.000 tweet ogni minuto)[5].

Per di più, l'architettura fisica di qualsiasi social network non è quella di una grande infrastruttura "ministeriale", dove migliaia di persone devono coordinarsi assiduamente tramite una puntigliosa e ramificata burocrazia per fare in modo che anche solo uno slogan raggiunga un vasto pubblico. Essa risiede, invece, in una rete di server, grandi computer interconnessi tra loro e capaci di svolgere miliardi di calcoli e di condividerne i risultati in una frazione di secondo che si limitano a registrare e a diffondere in modo automatico qualsiasi contenuto venga pubblicato dai milioni di utenti ad essi collegati. È dentro i confini di questa architettura che avvengono le continue e frenetiche azioni di pubblicazione e condivisione di contenuti e la loro perenne e scrupolosa analisi statistica da parte delle macchine. Tutto questo avviene sotto l'occhio vigile ma totalmente acritico dei programmi creati per la gestione e la trasmissione dei dati, per i quali le parole, le immagini ed i fotogrammi dei video condivisi non sono altro che una lunga serie di zero ed uno (i bit).

2. I bot, natura e azioni compiute sui social media

È proprio l'architettura di queste "piazze", governate da software acritici e volti a permettere al maggior numero possibile di utenti la condivisione di contenuti e a misurarne le interazioni, ad aver permesso a soggetti in possesso di capitali relativamente contenuti (sicuramente rispetto a quelli necessari per il funzionamento di un ministero) e di un certo grado di conoscenza del funzionamento di queste reti di computer di crearsi letteralmente un proprio seguito, anche molto ampio, sempre pronto a promuovere e condividere immediatamente qualsiasi cosa essi pubblichino sui social. Quella che a prima vista pare una schiera di fedelissimi seguaci profondamente convinti della validità dei contenuti pubblicati da questi soggetti non è, tuttavia, effettivamente composta da esseri umani assiduamente presenti su queste piattaforme. Si tratta, invece, dei cosiddetti *social bot* o nella vulgata più comune, criticata da diversi ricercatori per la sua portata confusiva all'interno del dibattito, semplicemente bot (dalla contrazione delle parole software robot).

Prima ancora di descrivere nel dettaglio la natura e le azioni compiute da questi *social bot*, pare qui opportuno compiere in via preliminare un chiarimento sul lessico utilizzato all'interno del più ampio di-

scorso sui bot e sul discendente cosiddetto "problema *bot*". Questo perché già da tempo è stata sottolineata l'incertezza generata dall'impiego disordinato di termini e definizioni per descrivere i fenomeni legati al loro uso (e abuso) in ambito informatico[6].

Il termine bot è stato utilizzato per indicare una varietà estremamente estesa di software, la cui ampiezza è mutata a seconda del grado di conoscenza tecnica e del periodo storico. All'inizio della diffusione dei personal computer, il termine era stato impiegato per riferirsi a una pluralità di sistemi software che comunicavano messaggi di avviso agli utenti umani[7], mentre all'inizio degli anni Duemila venne associato, all'interno degli ambienti della sicurezza informatica, a computer compromessi e controllati da remoto da dei malware[8]. Successivamente, con l'ascesa di Twitter come social network di ampio utilizzo e base principale per l'automazione dei profili, alcuni ricercatori hanno cominciato ad utilizzare questo termine per indicare gli account gestiti da un software[9], mentre informatici attivi nel campo della sicurezza informatica hanno preferito l'utilizzo del termine *sybil*, già usato in quell'ambiente per descrivere attori o nodi compromessi all'interno di una rete[10]. A voler dare una definizione generale di bot, pare qui opportuno riferirsi a quella proposta da Tsvetkova, Garcia-Gavilanes, Floridi e Yasseri nel 2017[11], ossia di un codice di programmazione (si veda l'esempio di codice sorgente presente più avanti in Figura 2) che viene eseguito continuamente agendo senza alcun tipo di intervento umano e reagendo autonomamente ai cambiamenti dell'ambiente (informatico) in cui opera.

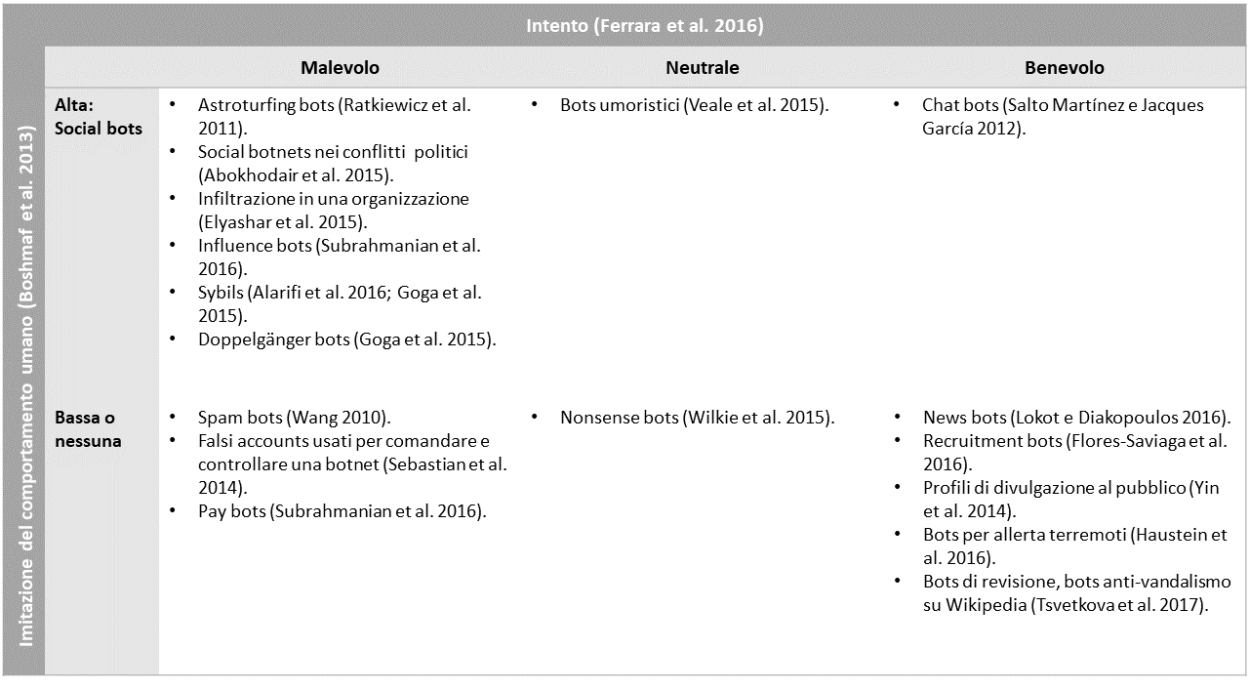
Nella categoria dei bot sopra delineata rientrano, così, moltissime tipologie di programmi per il compimento automatico di determinate azioni (si veda l'organizzazione di questa categoria in Figura 1 operata da Stieglitz, Brachten, Ross e Jung). Fra queste vi rientra anche quella degli strumenti informatici volti ad automatizzare varie attività sui social media, ossia quella dei *social bot*. Pure questa categoria, però, non trova una propria definizione univoca e precisa: alcuni ricercatori usano il termine *social bot* per indicare qualsiasi account sui social media gestito da un algoritmo[12], mentre altri lo usano per indicare solamente i «programmi per computer progettati per utilizzare i social network simulando il modo in cui gli esseri umani comunicano e interagiscono insieme»[13].

Infine, alcuni hanno costruito l'ulteriore ed ancora più specifica categoria dei *socialbot* (una parola unica), all'interno della quale vengono fatti rientrare gli account automatizzati che, assumendo un'identità inventata, possono infiltrarsi in reti di utenti reali e diffondere link o contenuti dannosi[14]. Essi vengono anche definiti in termini di sicurezza informatica come "avversari" e spesso vengono chiamati *sybil* (riprendendo la terminologia usata per descrivere gli attori o i nodi compromessi di una rete).

Date, quindi, le molteplici definizioni che sono state proposte e gli utilizzi spesso lamentati come impropri dei termini, appare qui opportuno sottolineare che in questa sede verranno presi in considerazione solo i *social bot* (due parole distinte) ed i *socialbot* (una parola unica). In particolare, ci si concentrerà sulle attività di diffusione di insinuazioni, maldicenze ed offese operata da questi ultimi (che formano una sotto-categoria specifica dei primi). Si tenga anche presente che verranno prese in considerazione le solo ipotesi di gestioni automatizzate di profili di persone inesistenti e non anche quelle di creazione non autorizzata di profili nuovi di individui reali, per le quali si configura anche il reato di sostituzione di persona previsto dall'articolo 494 del nostro codice penale[15]. In altre parole, verranno qui considerati solo quei software che, una volta forniti delle credenziali di accesso di un account di un social media, sono in grado di compiere le stesse azioni di un utente umano, che agiscono in modo da dare l'impressione che dietro al profilo che gestiscono ci sia una vera persona e che hanno il fine specifico di diffondere insinuazioni, maldicenze ed offese.

Visto tutto quanto fin qui considerato, è quindi facile immaginare come i profili social particolarmente attivi nel supportare e condividere continuamente i contenuti pubblicati da un altro, invece che appartenere a dei sostenitori entusiasti, ben potrebbero essere gestiti da un algoritmo capace di produrre in automatico contenuti e interazioni e progettato per emulare il comportamento di un essere umano[16], seguendo con una precisione meccanica gli argomenti ed i soggetti che è stato programmato a seguire.

Nonostante il funzionamento dei *social bot* possa inizialmente sembrare complesso, bisogna considerare che esso si fonda sulla ripetizione di azioni preimpostate dalla piattaforma stessa su cui opera (si pensi ai "mi piace" e "condividi" di Facebook o ai tweet e retweet di Twitter), tutte azioni standardizzate e ripetitive per le quali esistono appositi e ben identificati pulsanti virtuali, e sulla raccolta dei dati caricati dagli utenti e liberamente accessibili da parte di chiunque vi sia iscritto[17]. È, quindi, decisamente semplice istruire un programma, ossia il *social bot*, ad individuare le informazioni da copiare e le azioni da compiere, ossia i tasti virtuali da premere, e a compierle solo al verificarsi di determinate condizioni che esso è ben capace di riconoscere (ad esempio, il fatto che ad un determinato post il profilo che questo particolare bot gestisce non abbia ancora messo un



Imitazione del comportamento umano (Boshmaf et al. 2013)	Intento (Ferrara et al. 2016)		
	Malevolo	**Neutrale**	**Benevolo**
Alta: Social bots	• Astroturfing bots (Ratkiewicz et al. 2011). • Social botnets nei conflitti politici (Abokhodair et al. 2015). • Infiltrazione in una organizzazione (Elyashar et al. 2015). • Influence bots (Subrahmanian et al. 2016). • Sybils (Alarifi et al. 2016; Goga et al. 2015). • Doppelgänger bots (Goga et al. 2015).	• Bots umoristici (Veale et al. 2015).	• Chat bots (Salto Martínez e Jacques García 2012).
Bassa o nessuna	• Spam bots (Wang 2010). • Falsi accounts usati per comandare e controllare una botnet (Sebastian et al. 2014). • Pay bots (Subrahmanian et al. 2016).	• Nonsense bots (Wilkie et al. 2015).	• News bots (Lokot e Diakopoulos 2016). • Recruitment bots (Flores-Saviaga et al. 2016). • Profili di divulgazione al pubblico (Yin et al. 2014). • Bots per allerta terremoti (Haustein et al. 2016). • Bots di revisione, bots anti-vandalismo su Wikipedia (Tsvetkova et al. 2017).

Figura 1: *Categorizzazione dei bot proposta da Stieglitz, Brachten, Ross e Jung nel 2020. Gli account vengono differenziati in base al loro intento e grado di imitazione del comportamento umano (si noti come gli autori abbiano preferito il termine* sybil *a* socialbot*)*

"mi piace" o non l'abbia ancora condiviso). A ciò si aggiunga, poi, che è possibile trovare su Internet le istruzioni, corredate anche di esempi, per la creazione di versioni più o meno complesse di questi *social bot*, che è pure possibile reperire intere librerie con licenze d'uso gratuito[18] di comandi preimpostati da poter impartire loro (in quest'ambito una delle librerie più diffuse si chiama *Selenium* e la si può recuperare con una semplice ricerca tramite Google)[19] e che è persino possibile recuperare liberamente e gratuitamente il codice sorgente di questi programmi su siti Internet ampiamente frequentati[20].

La prosopopea che ammanta il dibattito sulle capacità delle nuove tecnologie informatiche non deve però trarre in inganno: tutti i bot sono dei semplici programmi che si limitano a compiere delle azioni (o dei cicli di azioni) sulla base delle istruzioni che hanno ricevuto, o meglio, che si limitano a compiere le operazioni inserite all'interno del loro stesso codice sorgente (si veda l'esempio in Figura 2). Ciò significa che sullo sfondo di tutta questa struttura permane sempre la presenza di almeno un essere umano, un programmatore che ha deciso come questi bot debbano comportarsi, ossia che ha deciso quali operazioni inserire all'interno del loro codice e che questi avrebbero poi svolto[21]. In altre parole, dietro a ciascun bot (o a ciascuna squadra di bot) vi è sempre una persona che ha consciamente deciso quali eventi (e quali prevedibili conseguenze) avrebbero dovuto verificarsi tramite l'avvio di questi suoi "strumenti digitali" e che, quindi, difficilmente non potrà essere considerato il responsabile di qualsiasi risultato finisca col verificarsi[22]. Ciò comporta, ovviamente, che la questione della responsabilità debba estendersi anche a tutti coloro che abbiano commissionato[23] la creazione o abbiano ottenuto per sé o per altri da lui diretti la gestione dei bot, risultandone, così, in qualche forma i mandanti ed i beneficiari (anche indiretti) responsabili di tutte le azioni compiute da questi strumenti (si pensi, ad esempio, a degli influencer che commissionino la creazione di *social bot* che seguano e supportino i loro profili, accrescendo artificialmente il numero dei propri follower e finendo così con il violare i termini contrattuali di utilizzo di una piattaforma social).

Questo fenomeno, poi, si giova di un'altra importantissima caratteristica delle tecnologie digitali: la ripetibilità. Infatti, una volta creato un *social bot*, un programma piuttosto "leggero" per le capacità di calcolo dei moderni computer, è possibile copiarlo e dargli le credenziali di un altro account nel giro di pochissimi minuti, creando così un secondo follower,

che, come il primo, sarà un fedelissimo sostenitore di qualsiasi profilo sia programmato a seguire. Questo processo può essere quindi ripetuto un'altra volta e un'altra ancora, tanti quanti sono i profili che si hanno a disposizione e, dato che non esiste un vero e proprio limite al numero di profili *social* che si possono aprire, è concettualmente possibile continuare all'infinito (in una ricerca pubblicata dal *The New York Times* viene fatto riferimento ad un software che «potrebbe creare fino a 100.000 nuovi accounts in cinque giorni»)[24], creando così folle oceaniche di sostenitori artificiali di chiunque possa permettersi di pagare per questo procedimento.

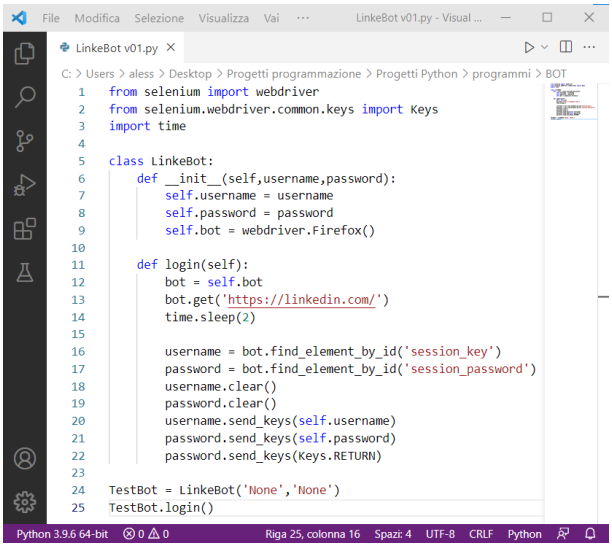


Figura 2: *Codice sorgente di un* social bot *che utilizza la libreria Selenium ed il driver GeckoDriver per compiere l'accesso ad un profilo LinkedIn attraverso il browser Firefox (codice realizzato da uno degli autori a riprova che chiunque può imparare a creare un* social bot*)*

Insomma, oggigiorno per essere al centro dell'interesse pubblico e poterne orientare i sentimenti non servono più schiere di funzionari statali o di dipendenti privati lautamente pagati (si pensi, ad esempio, all'impero della carta stampata di William Randolph Hearst – che ancora oggi conta circa 20.000 dipendenti – la cui influenza a fine Ottocento fu tale da spingere l'opinione pubblica statunitense a chiedere l'entrata in guerra contro il Regno di Spagna per supportare l'indipendenza di Cuba), ma solo un modesto capitale da poter spendere ed un ristretto gruppo di programmatori neanche necessariamente brillanti.

3. I potenziali pericoli rappresentati dai *social bot* sui social media

L'aspetto più problematico di un tale processo è il fatto che chiunque abbia o sia in grado di reperire (magari anche grazie a gruppi di interesse stranieri)[25] un capitale adeguato sarà in grado di crearsi una rete di fedelissimi follower artificiali programmati per osannarlo e per ripetere prontamente qualunque cosa dica, senza che questa sia effettivamente convincente, veritiera o anche solo rispettosa dei più basilari principi del rispetto della persona (finendo, così, col configurarsi come dei *socialbot*), e che diano ai meno esperti l'illusione che sia invece il grande pubblico a condividere e a seguire la sua dottrina[26]. In altre parole, come ben spiegato da Menczer, professore di informatica presso l'Indiana University, i *social bot* «possono distorcere la visione del mondo che abbiamo online, manipolando così l'opinione pubblica. Possono creare l'impressione che un'idea o una persona siano popolari quando non lo sono»[27]. Non solo, chiunque si possa permettere di creare in questo modo dei sostenitori, potrà anche creare un esercito di feroci oppositori ad un determinato soggetto, i quali non faranno altro che attaccarlo, denigrarlo e diffamarlo incessantemente, attuando così quello che già alcuni definiscono come una sorta di "squadrismo digitale".

Questa realtà tecnologica, ormai diffusa e ben radicata nel panorama dei social network (secondo uno studio congiunto del 2017 della University of Southern California e della Indiana University[28], circa il 15 per cento dei profili Twitter, ossia circa 48 milioni, sono gestiti da *social bot*), ha colorato di nuove e più fosche sfumature il dibattito sempre attuale sul bilanciamento tra la libertà di espressione e la tutela della dignità e della reputazione delle persone. Se in passato tale dibattito, concentrato soprattutto sulla necessità di trovare un equilibrio tra la pubblicazione di notizie e il diritto alla riservatezza, nonché sulla difesa da diffamazioni e falsità, vedeva come principali protagonisti giornalisti e personaggi pubblici, oggi gli attori di queste azioni propagandistiche o diffamatorie possono essere anche dei privati cittadini, privi di organizzazioni strutturate, nonché privi delle autorizzazioni o delle identificazioni spesso richieste per poter pubblicare opinioni e notizie[29].

A rendere ancora più preoccupante l'avvento di questi sostenitori artificiali nell'ambito dei social network è il fatto che i *social bot* a cui sono dati in gestione i profili possono essere installati su qualsiasi computer che abbia accesso a Internet, senza quindi la necessità che queste macchine si trovino all'interno di un qualche tipo di redazione, né addirittura che

si trovino all'interno del paese in cui risiedono il loro finanziatore o il loro creatore. Questo significa, ad esempio, che, se un soggetto che risiede nell'Unione europea volesse crearsi o comprarsi una schiera di follower per il proprio profilo social, non solo non avrebbe la necessità di installare i *social bot* su dei computer posti nel suo domicilio ma neppure entro i confini comunitari potendoli così tenere al di fuori della giurisdizione di qualsiasi autorità europea e rendendoli per queste di fatto irraggiungibili o difficili anche solo da identificare. Alla possibile extraterritorialità della sede fisica dei *social bot* si aggiungono, poi, la facilità con cui essi possono essere trasferiti in altre località (installarli su di un altro computer è un'operazione che può essere eseguita in brevissimo tempo) e, soprattutto, la facilità quasi elementare con cui si possono dare loro le credenziali di accesso di altri profili. Fattore quest'ultimo che risulta estremamente utile nel caso in cui i profili che questi *social bot* gestiscono vengano sospesi o banditi dai social network in cui operano.

Di fatto, quindi, chi sappia creare e ben orchestrare queste folle di utenti artificiali ben potrebbe apparire completamente estraneo agli umori e alle direzioni presi da esse, pur essendone in realtà l'unico vero direttore. Così, celandosi dietro a queste reti digitali di programmi e profili, chiunque abbia le giuste disponibilità potrebbe sia fingersi portavoce dei sentimenti delle masse attive online, anche se in verità ne sta dettando gli orientamenti in maniera unilaterale (potenzialmente sia a favore che contro chiunque), sia mostrarsi lontano se non addirittura contrario a qualsiasi affermazione, anche diffamatoria, scaturita "spontaneamente" da esse. Siamo, quindi, di fronte al rischio concreto che soggetti privi di scrupoli ma in possesso delle giuste conoscenze possano presentarsi come dei moderni vati della società oppure degli strenui difensori della dignità e della reputazione delle persone, quando in realtà, per usare le parole di Ferrara, Varol, Davis, Menczer e Flamini[30], sono dei burattinai che muovono a proprio piacimento i fili di migliaia di pupazzi digitali, liberi anche di orientarli contro quegli stessi avversari di cui in pubblico difendono l'onore. Così, chiunque finisca al centro dell'attenzione di questo megafono digitale – ed artificiale – può diventare la vittima di un'autentica tempesta di denigrazioni e calunnie, che può travolgerlo e farlo naufragare ancora prima che si renda conto di cosa gli stia capitando e alla quale, allo stato attuale delle normative e delle tecnologie, non sembra possibile opporre dei rimedi rapidi e realmente efficaci per tutelare la sua reputazione.

4. L'inquadramento della diffamazione sui social media in Italia

Non che la nostra normativa nazionale difetti di riconoscere la lesività della diffusione di false informazioni sul conto di una persona: l'articolo 595 ("Diffamazione") del nostro codice penale è espressamente dedicato a sanzionare chi, «comunicando con più persone, offende l'altrui reputazione». In particolare, il terzo comma di detto articolo, nell'individuare gli strumenti con i quali è possibile compiere il reato di diffamazione in forma aggravata, apre le porte all'inclusione tra questi dei mezzi informatici e digitali. Infatti, questo comma considera genericamente che l'offesa può essere compiuta «con qualsiasi altro mezzo di pubblicità». Quindi, all'interno di questa categoria residuale degli strumenti di comunicazione (e, di conseguenza, di pubblicità) con cui è possibile diffondere opinioni lesive della dignità altrui già da tempo la nostra giurisprudenza nazionale ha fatto rientrare i social media[31], riconoscendo loro una velocità ed una capillarità di diffusione di interpretazioni, opinioni ed inesattezze che eclissano con facilità addirittura quelle dei più grandi e strutturati organi di stampa tradizionali e, così, anche la pericolosità che essi possono comportare per la reputazione di chiunque[32].

Di conseguenza, come ben rilevato dalla Corte di cassazione, «la diffusione di un messaggio diffamatorio attraverso l'uso di una bacheca "Facebook" integra un'ipotesi di diffamazione aggravata»[33]. Ciò avviene, nello specifico, in ragione del fatto che la pubblicazione di un contenuto, quale un testo, un'immagine o un video, all'interno di un qualsiasi social network comporta la sua immediata visibilità da parte di tutti i membri di questa "piazza virtuale" (con delle modalità che variano leggermente a secondo della struttura del social su cui è stato pubblicato) e, quindi, risulta essere una «condotta potenzialmente capace di raggiungere un numero indeterminato o comunque quantitativamente apprezzabile di persone» e, così, «ampliando – e aggravando – in tal modo la capacità diffusiva del messaggio lesivo della reputazione della persona offesa»[34]. Queste considerazioni sono state poi riprese dai giudici anche in sentenze successive, consolidando l'orientamento secondo cui la diffusione di un messaggio diffamatorio attraverso l'uso di un social network integra le previsioni di cui al terzo comma dell'art. 595 del codice penale[35].

La nostra giurisprudenza, insomma, ha già riconosciuto da tempo come l'utilizzo di social media per la diffusione di insinuazioni, maldicenze ed offese integri non solo il reato di diffamazione, ma che lo faccia in forma aggravata, dato che questi sono un

"mezzo di pubblicità" con il quale è facile raggiungere rapidamente un alto numero di persone. Tuttavia, la nostra stessa giurisprudenza ha mostrato come, per via delle particolari funzionalità tecniche del *medium* interessato, già solo perseguire gli autori umani delle denigrazioni ha presentato delle difficoltà non indifferenti.

5. La problematica attuazione della tutela della reputazione lesa tramite social media

Per riparare alle lesioni aggravate della reputazione il nostro ordinamento predispone una serie di strumenti di tutela per il soggetto colpito dalle dichiarazioni diffamatorie. In particolare, di fronte ad una lesione del proprio onore, è possibile proporre denuncia-querela e chiedere così alle autorità di ricercare l'asserito colpevole oppure è possibile citarlo innanzi al giudice civile per chiederne la condanna al risarcimento dei danni patiti. Parallelamente, sempre a tutela della reputazione, il nostro ordinamento prevede anche degli strumenti cautelari volti a prevenire l'aggravamento della lesione avvenuta: in sede penale la vittima potrà chiedere il sequestro preventivo della pagina social o del profilo che contenga il contenuto diffamatorio[36], mentre in ambito civilistico la vittima potrà richiedere il procedimento cautelare d'urgenza previsto dall'articolo 700 del codice di procedura civile (che potrà essere instaurato ancor prima di avviare il giudizio ordinario) per far cancellare il contenuto diffamatorio in tempi brevissimi, così scongiurandone l'eccessiva diffusione.

Tuttavia, come già accennato, i procedimenti volti a dare tutela al giusto credito sociale di ciascuno incorrono comunque in numerose problematiche di attuazione materiale per quanto riguarda i casi digitali. Innanzitutto, l'identificazione del titolare di un normale profilo social non è di immediata realizzabilità: sebbene sui social media le persone tendano a creare dei profili personali utilizzando i propri nomi e cognomi reali, spesso per rendersi facilmente riconoscibili dai membri di una determinata cerchia sociale (si pensi agli "amici" su Facebook o alle "connessioni" su LinkedIn), ben poco impedisce che qualcuno crei un profilo fornendo delle credenziali differenti da quelle della propria reale identità[37] o, peggio ancora, che impersoni un altro soggetto[38]. Con questa libertà di iscrizione è così possibile dare in gestione ad un *social bot* il profilo creato senza incontrare alcun vero ostacolo (come un effettivo accertamento dell'identità o della natura del suo utilizzatore)[39]. La conseguenza di tutto ciò è, così, l'impossibilità di utilizzare le credenziali del profilo per identificare l'autore degli atti diffamatorii e tantomeno il creatore dei *socialbot* che li abbiano compiuti.

L'impossibilità di avere certezza dell'identità del diffamatore tramite le credenziali fornite in rete da lui stesso ha spinto la giurisprudenza italiana ad attribuire rilevanza ad un altro elemento caratteristico delle interazioni online: l'*Internet Protocol Address* (detto anche indirizzo IP), un codice numerico che viene assegnato in automatico a qualsiasi dispositivo connesso ad Internet o ad una rete locale per identificarlo in maniera univoca e consentirgli di comunicare con altri dispositivi[40]. La riconosciuta imprescindibilità di questo indirizzo per avere certezza della corretta individuazione dell'autore delle lesioni all'altrui onore deriva dal fatto che esso viene attribuito dalle compagnie telefoniche ad ogni cliente che abbia una connessione Internet e che, di conseguenza, rivolgendosi ad esse è possibile ottenere informazioni circa l'identità del titolare di questo indirizzo. Tuttavia, un indirizzo IP permette al massimo di identificare il titolare della linea telefonica associata e non l'identità dell'autore dei contenuti diffamanti, non consentendo, quindi, di escludere che un soggetto diverso dal titolare dell'utenza abbia sfruttato la linea telefonica[41].

Inoltre, come precedentemente sottolineato, il dispositivo da cui vengono diffusi (anche automaticamente) i contenuti diffamatorii ben potrebbe trovarsi al di fuori dei confini territoriali dello Stato in cui si trova il suo utente nonché la vittima di tali espressioni, rendendo così ancora più complesso risalire all'identità personale dell'autore del reato. Di fronte ad una situazione del genere, la via da percorrere sarebbe quella di richiedere ai gestori dei social media – solitamente molto più attenti a tutelare la privacy che la reputazione dei propri utenti – di fornire copia dei dati di identificazione del profilo in loro possesso e, di fronte ad un loro probabile rifiuto, questa strada diventerebbe ancora più lunga e tortuosa: l'unica cosa da fare sarebbe rivolgersi alla magistratura requirente, la quale a sua volta dovrebbe presentare una rogatoria internazionale, che ben potrebbe ricevere risposta negativa da parte dei suoi omologhi stranieri (ad esempio perché in molti degli Stati USA, dove la maggior parte dei social media ha la propria sede legale, la libertà di parola occupa un ruolo decisamente preminente rispetto alla tutela della reputazione)[42]. Si noti, però, che negli Stati Uniti, per far fronte a questo possibile – e probabile – rifiuto dei gestori, le normative statali e le corti hanno costruito delle procedure antecedenti l'apertura di un processo per scoprire l'identità dei soggetti che pubblichino dei contenuti contestati come diffamato-

ri. In particolare, la *Rule* 224 adottata dalla Corte Suprema dello Stato dell'Illinois ha istituito una procedura precisa che permette di presentare un ricorso (*petition*) per ottenere, qualora vi siano *prima facie* le condizioni, che il gestore della piattaforma online comunichi i dati sensibili degli utenti che hanno pubblicato i contenuti contestati e poterli così citare in giudizio[43]. Tutto ciò, però, non esclude che chiunque gestisca detto profilo (sia esso una persona fisica o una macchina addestrata) abbia fornito dei dati falsi o inutilizzabili per rintracciarlo, ad esempio celando l'indirizzo IP dell'utenza telefonica usata tramite l'uso di un *Virtual Private Network* (VPN)[44] o di un *server proxy*[45] e rendendosi così irreperibile[46].

6. Tentativi di risposta al problema della presenza dei *social bot* sui social media

Dalle considerazioni fin qui esposte in merito all'inquadramento della diffamazione compiuta a mezzo social dalle persone – ormai pacificamente acquisito dalle nostre corti nazionali – non può che discenderne la logica conseguenza che anche l'utilizzo dei *socialbot* per la diffusione di messaggi e contenuti lesivi della reputazione altrui vi rientri completamente. Anzi, data la loro rapidità e precisione di reazione, deriva necessariamente non solo che integri pienamente una forma aggravata del reato previsto dall'articolo 595 c.p. ma, anzi, che, accrescendo in modo esponenziale gli effetti diffamatori della loro diffusione tramite le piattaforme digitali, comporti ancora più che in altri casi la necessità di realizzare una rapida ed efficace tutela delle vittime.

Infatti, un soggetto umano che decida di pubblicare contenuti diffamatorii all'interno di un social network, per quanto possa essere ampia la sua rete di contatti (il *network*, appunto) ed assidua la sua attività di diffusione digitale, incontrerà sempre dei limiti di tempo, di spazio e fisiologici con cui, invece, una macchina non deve confrontarsi. Similmente, poi, anche nelle attività di supporto dei contenuti pubblicati da altri (si pensi ai "mi piace" e "condividi" di Facebook o al retweet di Twitter) la capacità di interazione di un *social bot* risulta scevra dalle costrizioni biologiche di un utente umano. In aggiunta, è soprattutto nel caso di queste attività di "supporto" che si possono esprimere pienamente le potenzialità dei *socialbot*: essi, dato che sono quasi sempre usati contemporaneamente in numero elevato e che hanno una reattività (detta anche *engagement*) immediata e completa a qualunque contenuto pubblicato dal profilo che sono programmati a seguire e supportare, possono facilmente dare fin da subito l'illusione che quanto espresso sia popolare (e, quindi, molto spesso erroneamente anche considerato veritiero, come emerso dalla ricerca condotta da Luo, Hancock e Markowitz nel 2020)[47] e portare così ad una sua ancor più celere diffusione.

In merito alle diverse forme di diffusione sui social media di contenuti diffamatorii da parte dei *socialbot*, il discorso trova delle particolari declinazioni per ciascuna piattaforma considerata ma è comunque possibile identificare tre diverse tipologie di azioni che questi programmi "avversari" possono compiere: (1) la produzione automatica di contenuti originali (la pubblicazione di post, tweet e commenti); (2) la condivisione di contenuti di altri (le azioni "condividi" o "retweet" dei contenuti di terzi, anche esterni alla piattaforma); (3) l'apposizione di "reazioni" a sostegno dei contenuti di terzi (ad esempio, i "mi piace" di Facebook o i "consiglia" di LinkedIn). Questa distinzione, però, non sembra comportare delle profonde diversità nell'integrazione delle condotte diffamatorie. Infatti, la nostra giurisprudenza ha già messo in luce come tutte queste attività possano pienamente integrare il reato di diffamazione aggravata previsto dal terzo comma dell'articolo 595 c.p.: per quanto riguarda la prima tipologia di azioni si è già visto come da tempo per la Corte di cassazione essa «integra un'ipotesi di diffamazione aggravata» sia nei casi di pubblicazione di contenuti[48] che di commenti ad essi[49]; relativamente alla seconda tipologia di condotte diffusive (la condivisione di contenuti di terzi)[50] si tenga presente che esse consistono nella esposizione dei contenuti prodotti da terzi all'interno della propria personale rete sociale, di fatto diffondendo tra i propri contatti le espressioni lesive dell'altrui onore e, quindi, secondo le parole usate dagli Ermellini nella loro decisione del 29 gennaio 2016, n. 3981, «dimostrando di volerle amplificare attraverso il proprio comportamento»; da ultima, l'apposizione di "reazioni" ai contenuti di terzi è già stata riconosciuta dai Giudici di Piazza Cavour come potenzialmente rilevante per la nostra normativa penale, riconoscendo come l'opzione di Facebook "mi piace" abbia carattere diffusivo, «attesa la . . . funzione propalatrice svolta in tale contesto dal social network» e sia, quindi, idonea ad avere portata offensiva[51].

Tuttavia, come visto, anche già solo perseguire gli autori umani delle denigrazioni ha presentato delle difficoltà non indifferenti, sia a livello concettuale che a livello pratico, che si ripresentano in forma ancora più complessa nei confronti degli amministratori dei *socialbot*. Infatti, gli ostacoli tecnici fin qui visti nel risalire anche solo all'identità di una persona fisica, dati i caratteri sopra illustrati di velocità, ripetibilità

e delocalizzazione propri di questi programmi, possono portare con estrema facilità a rendere lettera morta qualsiasi richiesta di tutela e giustizia di chiunque sia stato travolto dalla tempesta di denigrazioni e calunnie realizzata attraverso questo megafono digitale ed artificiale. Di conseguenza, non sembra irragionevole auspicare una maggior responsabilizzazione ed un maggior coinvolgimento dei gestori di questi media nelle attività di contrasto di questo fenomeno e di tutela della reputazione di quelli che molto spesso sono anche i loro utenti.

Bisogna comunque riconoscere come di fronte a questo fenomeno le compagnie proprietarie dei social network non siano rimaste completamente inerti ed abbiano implementato una serie di misure volte ad evitare o comunque a limitare la diffusione dei *social bot* che ingigantiscano i numeri delle reazioni e delle azioni che sono programmati a compiere. Tuttavia, queste misure in massima parte si limitano ad un'analisi acritica delle condotte e delle abitudini dei profili, tramite la valutazione di elementi statistici quali la velocità e la frequenza di azione[52], e ad una successiva eventuale sospensione dei profili che compiono azioni non in linea con le interazioni mediamente compiute da un essere umano. Tutto ciò, come riconosciuto dagli stessi autori del già citato studio delle attività dei *social bot* sui social media[53], non è un approccio definitivo, in grado di risolvere alla radice i rischi della presenza dei *socialbot* sulle piattaforme social, anche a causa del fatto che è sempre possibile affinare questi ultimi perché abbiano dei comportamenti sempre più umani[54,55].

Insomma, le misure adottate dalla maggior parte dei gestori non sono risultate risolutive nell'affrontare il problema di fondo creato dai *socialbot* sui social network: l'accrescimento artificiale della diffusione di contenuti dannosi per la reputazione di persone o, addirittura, di interi gruppi sociali o etnici.

7. La necessità di nuovi strumenti di tutela

Sebbene sembri possibile far rientrare questa problematica all'interno del già ampio dibattito relativo al delicato equilibrio tra libertà di parola e diritto al riconoscimento del giusto credito sociale, a parere di chi scrive, questa questione ha dei confini maggiormente sfumati sul fronte della semplice libertà di parola e degli strumenti con i quali sia possibile esprimerla. Infatti, come già precedentemente illustrato, i *social bot* si limitano ad essere dei semplici strumenti volti ad accrescere artificialmente la diffusione e la visibilità di opinioni e contenuti e, quindi, il dibattito in merito ad essi dovrebbe riguardare soprattutto la libertà del loro utilizzo.

Non si vuole certo suggerire l'introduzione di un divieto legale di utilizzo dei *social bot* sui social network, sia perché già le condizioni di utilizzo delle piattaforme prevedono dei divieti o delle limitazioni al loro uso (ma ciò non ha sortito risultati soddisfacenti)[56], sia perché in alcuni casi i *social bot* vengono usati anche per finalità perfettamente lecite e addirittura utili[57]. Riprendendo un esempio già fatto da altri[58], si pensi ai servizi di allerta automatizzati in caso di emergenza, talmente utili da essere addirittura integrati all'interno di alcune applicazioni delle piattaforme social (come gli "avvisi locali importanti" di Facebook) o ad un quotidiano che, per poter immediatamente informare i suoi lettori tramite il proprio profilo social sulle variazioni di borsa, abbia dato ad un *social bot* le credenziali di accesso ad un suo "sotto-profilo" e gli abbia dato il compito di pubblicare automaticamente un post ad ogni fluttuazione degli indici azionari (nessun essere umano, infatti, sarebbe capace di svolgere lo stesso compito di controllo e pubblicazione alla medesima velocità e con gli stessi ritmi ed orari né avrebbe molto interesse ad impegnarsi a fare ciò per il resto della propria vita lavorativa).

Tuttavia, la facilità con cui è possibile affidare uno o più profili a dei *socialbot* e la minaccia che questi possono rappresentare come elevatori a potenza della diffusione sui social media di contenuti lesivi per la reputazione delle persone impone una loro attenta considerazione all'interno del dibattito su libertà di parola e tutela della reputazione in rete. In molti hanno già da tempo rilevato la sussistenza di un "problema *bot*", spesso generalizzando la questione o riferendosi ad essa con termini impropri, e hanno avanzato diverse soluzioni (spesso mostrando una comprensione non proprio piena del funzionamento di questi software) che sarebbero di difficile – se non impossibile – implementazione e avrebbero conseguenze variatamente profonde sulla struttura ed il funzionamento dei social media e, più in generale, dell'intera economia digitale. Si è voluto, quindi, esporre nelle pagine successive alcune possibili soluzioni – o mitigazioni – ad un fenomeno che, sebbene non capillarmente diffuso, potrebbe facilmente causare, per riprendere le parole della Consulta[59], un'amplificazione degli addebiti diffamatori e il cui carattere lesivo per la vittima – in termini di sofferenza psicologica e di concreti pregiudizi alla propria vita privata, familiare, sociale, professionale e politica – risulterebbe grandemente potenziato.

Sebbene sia sempre opportuno quantomeno valutare la possibilità di introdurre modifiche o innova-

zioni all'interno del nostro sistema giuridico, prendendo in considerazione le soluzioni adottate – o anche solo proposte – da legislazioni straniere o dagli stessi gestori di alcuni social media, si deve ammettere come in quest'ambito domini ancora l'incertezza circa la loro effettiva adottabilità ed efficacia. In particolare, la maggior parte delle soluzioni di seguito illustrate affronta principalmente la più ampia questione della diffamazione – anche anonima – a mezzo social e solo indirettamente ha riflessi di incentivazione per i gestori delle piatteforme ad affrontare il "problema *bot*".

In ogni caso, di seguito vengono prese in considerazione delle riforme legislative (anche sull'esempio delle proposte avanzate nel contesto statunitense), i possibili perimetri di aggiornamento delle interpretazioni giurisprudenziali di normative ormai datate rispetto al progresso tecnologico e, infine, le soluzioni tecniche già adottate da alcuni dei prestatori di servizi della società dell'informazione.

7.1. Riforme legislative

Con riferimento alle possibili modifiche al contesto normativo, sebbene non siano ancora state promulgate delle disposizioni legislative specificamente mirate a contrastare le attività dei *socialbot*, vi sono già state delle proposte – sia da parte di un legislatore che della dottrina – che potrebbero avere degli effetti deterrenti rispetto all'utilizzo di questi algoritmi "malevoli". C'è tuttavia da ammettere che nella maggior parte dei casi le previsioni considerate sono volte ad attribuire con maggior facilità la responsabilità in capo ai prestatori di servizi informatici per le condotte dei loro utenti, lasciando così a questi l'effettivo compito di adottare delle misure che impediscano l'amplificazione della diffusione di contenuti diffamatorii sui social media tramite questi particolari strumenti digitali.

(a) *Divieto di utilizzo dei bot, l'esempio della* Feinstein Bill *del 2018* – In ambito normativo la soluzione più incisiva – e probabilmente di più difficile attuazione concreta – volta a contrastare le attività dei *socialbot* che sia stata presentata è quella della *Section 5* del proposto *Bot Disclosure and Accountability Act of 2018* (detto anche *Feinstein Bill*)[60]. La sezione di questa proposta di legge americana promuove l'introduzione di un divieto per candidati e partiti politici, commissioni politiche, corporazioni e organizzazioni del lavoro di utilizzare o far utilizzare da altri «programmi software automatizzati destinati a impersonare o replicare attività umane per pubblicità politica online». L'ampiezza di questa disposizione, che emenderebbe il Titolo III del *Federal Election Campaign Act of 1971* (52 U.S.C. 30101 et seq.), tuttavia, appare piuttosto scarsa dato che limiterebbe il divieto di utilizzo di questi strumenti informatici solamente a determinate categorie di persone e di associazioni e solo in relazione ad una specifica attività (la propaganda politica), nulla prevedendo per altri argomenti oggetto di pubblicazioni online, quali le condotte di privati cittadini o l'attendibilità di certe asserzioni sugli effetti di prodotti farmaceutici.

Un aspetto di questo disegno di legge che è, comunque, di grande interesse in questa sede è rappresentato dai divieti, formulati alla *Section 5*[61], di «utilizzare o far utilizzare [ad altri] qualsiasi programma o processo software automatizzato inteso a impersonare o replicare l'attività umana online per creare, amplificare, condividere o diffondere in altro modo qualsiasi comunicazione pubblica» oppure di «sollecitare, accettare, acquistare o vendere programmi o processi software automatizzati destinati a impersonare o replicare attività umane online per qualsiasi scopo». Infatti, tali proibizioni, che ben si possono estendere ad attività esterne ai social media (come, ad esempio, la firma di petizioni online su piattaforme non progettate per delle interazioni sociali), pongono come proprio oggetto di regolamentazione non solo l'uso – che sarebbe sanzionato come illecito – di strumenti digitali che impersonano in qualsiasi ambito informatico degli esseri umani ma anche altre forme di direzione delle attività di questi mezzi, rilevando, così, come la responsabilità per gli effetti da essi prodotti debba estendersi anche a tutti coloro che ne risultino in qualche forma i mandanti o i beneficiari (anche indiretti). Tuttavia, questa previsione non sembra tener conto dei limiti tecnici rilevati *supra* al paragrafo 5 di identificazione degli amministratori dei *social bot* installati su macchine con un indirizzo IP celato da un server proxy o poste al di fuori della giurisdizione dello Stato che verrebbe raggiunto da tali attività di diffusione di contenuti. Inoltre, questo disegno di legge andrebbe incontro a non poche difficoltà di applicazione pratica, dato che la definizione di "politica" nel contesto sociale di una nazione può ricomprendere argomenti molto variegati a seconda anche del periodo storico (si pensi solo a come negli Stati Uniti l'utilizzo di dispositivi medici come le mascherine durante la pandemia da Covid-19 si sia caricato di significati politici sconosciuti ad altre società)[62]. Infine, si noti anche come una trasposizione di questa proposta di modifica legislativa all'interno del nostro ordinamento, oltre a non coprire l'ambito della tutela dalla diffusione di contenuti diffamatorii e ad incappare nelle medesime ambiguità terminologiche già espresse riguardo a cosa costituisca "attività politica", finirebbe anche con

l'accendere una controversia sulle possibili restrizioni della libertà sindacale che ne potrebbero derivare (decisamente meno tollerate dal nostro ordinamento rispetto a quello statunitense).

In ogni caso, nonostante gli evidenti limiti del proprio contenuto, la *Feinstein Bill* del 2018 ha il pregio, almeno agli occhi di chi scrive, non solo di identificare un problema quantomai attuale (quello della diffusione automatizzata, concertata ed anonima di interpretazioni, opinioni ed inesattezze) ma anche di riconoscere come soggetti ed organizzazioni possano utilizzare o far utilizzare da altri degli strumenti «per creare un effetto carrozzone, per costruire false tendenze sui social media diffondendo automaticamente hashtag e persino per sopprimere le opinioni dell'opposizione»[63], che con tutta facilità potrebbe essere anche diretto verso la distruzione della reputazione di qualcuno.

(b) *Attribuzione di responsabilità agli Internet Service Provider, l'abolizione della* Section 230 *del* Communications Decency Act – Una seconda radicale soluzione che è stata formulata all'interno del contesto dell'equilibro tra la libertà di parola (la *Freedom of speech* del Primo Emendamento alla Costituzione degli Stati Uniti, concetto estremamente caro sia al sistema giuridico che alla società civile americani) e la tutela della reputazione operato dalla legislazione americana è l'abrogazione del cosiddetto "Privilegio del Buon Samaritano" (in inglese *Good Samaritan privilege*) previsto dalla *Section 230* del *Communications Decency Act* (CDA) statunitense[64]. Questa proposta, avanzata con dovizia di ragionamenti da B. Storm nel 2013[65], è posta in conclusione di un'articolata analisi della "lodevole" ma "eccessivamente ampia" tutela garantita ai prestatori di servizi informatici (detti *Internet Service Provider* o ISP) che «ignora una complessa storia di *common law* e di conseguenza scompagina l'attento atto di bilanciamento che è la legge sulla diffamazione» e che sembra essere «un passo in avanti verso l'assenza di una normativa [contro la] diffamazione». Infatti, come rilevato da più autori, sin dall'approvazione del CDA questo privilegio ha fermamente tutelato presso le corti americane gli ISP dalla responsabilità per le dichiarazioni diffamatorie di terzi, conosciuti o anonimi, pubblicate attraverso i loro servizi online[66] ed anche da altre tipologie di azioni civili[67].

Una proposta di modifica così profonda della normativa che regola la responsabilità di alcuni tra i servizi digitali più utilizzati nel mondo deriva anche dalla considerazione dell'ambiguità tenuta dalla giurisprudenza della Corte Suprema americana riguardo al cosiddetto "diritto all'anonimato" (*right of anonimity*), il quale è stato ricavato dal Primo Emendamento. Infatti, se da tempo oltreoceano sono stati indicati chiaramente come abusi del diritto alla *Freedom of speech* le attribuzioni obiettivamente false di caratteri o condotte personali, quali crimini, malattie o mancanze di integrità o abilità professionali, che vengono considerate forme di diffamazione *per se*[68], il diritto implicitamente riconosciuto da questo emendamento di esprimere le proprie opinioni senza essere identificati ha sempre ricevuto un'interpretazione restrittiva nei casi affrontati, senza però essere mai oggetto di una trattazione sistematica e chiarificatrice. Sebbene, infatti, in diverse sentenze la Corte abbia sancito il diritto all'anonimato solamente per i discorsi di carattere politico[69], essa ha solo implicitamente riconosciuto che detto diritto non si possa applicare anche a discorsi fraudolenti o diffamatorii[70], rendendo di fatto incerta la possibilità di reperire l'identità del diffamatore, nonostante il riconosciuto diritto di citare in giudizio convenuti sconosciuti[71]. Sommata a questa insicurezza, la disposizione di tutela dei *provider* della *Section 230* del CDA ha lasciato – argomenta sempre Storm – alle vittime di diffamazione «poche probabilità di avere successo in una legittima azione di [tutela dalla] diffamazione».

Una soluzione invero decisamente massimalista che molti contesterebbero comportare un totale ripensamento dei modelli economici di qualsiasi impresa che operi su Internet e permetta interazioni pubbliche tra e con i propri utenti, se non addirittura il loro fallimento. Tuttavia, questa ipotesi non sembra essere aprioristicamente rigettata da alcuni tribunali nel mondo. Ad esempio, i recenti casi *Fairfax Media Publications Pty ltd v. Dylan Voller*, *Nationwide News Pty limited v. Dylan Voller* e *Australian News Channel Pty ltd v. Dylan Voller* decisi dall'Alta Corte d'Australia[72], sebbene incentrati sulla responsabilità dei titolari di pagine su di un social network e non del titolare della piattaforma stessa, si sono conclusi con il riconoscimento delle ricorrenti (le società d'informazione giornalistica) come «editori dei commenti delle terze parti [ossia gli] utenti di Facebook». Ciò in quanto, argomenta la Corte, dal momento che «la responsabilità di una persona come editore dipende dal fatto che tale persona, facilitando e incoraggiando la relativa comunicazione, abbia "partecipato" alla comunicazione della questione diffamatoria a un terzo» e, dal momento che nel caso di specie avevano creato delle pagine pubbliche su Facebook e vi avevano pubblicato dei contenuti, esse avevano, di conseguenza, «facilitato, incoraggiato e quindi assistito la pubblicazione di commenti da terze parti utenti di Facebook». In altre parole, i giudici australiani hanno ritenuto responsabili dei soggetti che avevano semplicemente facilitato ed incoraggia-

to la pubblicazione di contenuti diffamatorii di terze parti, aprendo così la strada ad una possibile futura estensione di una responsabilità di questo tipo anche in capo ai gestori delle piattaforme *social*.

Una riforma in termini simili a quelli appena visti non sarebbe da sola in grado di fungere da ostacolo alle attività diffamatorie compiute dai *socialbot* e avrebbe certamente dei risvolti significativi sulle dinamiche delle attività d'impresa dei cosiddetti prestatori di servizi informatici. Ma essa avrebbe comunque il merito, da un lato, di indurre queste imprese ad adottare in autonomia delle soluzione al contrasto sia delle condotte diffamatorie perpetrate sulle loro piattaforme sia della presenza di profili gestiti artificialmente – e potenzialmente pericolosi – e, dall'altro, di assicurare una maggiore facilità alle vittime di tempeste orchestrate di denigrazioni e calunnie di ottenere un risarcimento in caso di mancata implementazione di procedure o strumenti volti a evitare la diffusione di contenuti lesivi del loro buon nome.

A voler provare a tradurre all'interno della nostra legislazione nazionale l'implementazione di una proposta di questo genere si dovrebbe compiere un vero e proprio ribaltamento delle disposizioni contenute nel d.lgs. 9 aprile 2003, n. 70[73]. In quanto ciò significherebbe istituire in capo agli ISP una responsabilità generale per tutti i contenuti memorizzati a richiesta di un destinatario dei loro servizi – che al momento è esclusa dell'articolo 16 di questo decreto[74] – e non solo per quelli della cui illiceità «sia effettivamente a conoscenza» finendo con l'istituire di fatto in capo agli ISP anche un vero e proprio obbligo generale di sorveglianza dei contenuti memorizzati o messi in rete da terzi (obbligo oggi espressamente escluso dall'articolo 17)[75].

In quest'ambito vi è, comunque, da sottolineare come la nostra giurisprudenza nazionale abbia già da tempo abbandonato la puntigliosa interpretazione del dato normativo, al fine di rendere maggiormente efficace la tutela di diritti violati da terzi tramite l'utilizzo di queste piattaforme. Infatti, sebbene la disposizione di cui all'articolo 16 del d.lgs. 70/2003 stabilisca una responsabilità del provider per tali violazioni solamente qualora rimanga inerte a seguito di «comunicazione delle autorità competenti», ossia a seguito di una sua "conoscenza qualificata", la giurisprudenza – invero per lo più nell'ambito della tutela dei diritti di proprietà intellettuale – ha preferito dare un'interpretazione di questo articolo maggiormente responsabilizzante, in base alla quale l'obbligo di rimozione del contenuto scatta anche quando l'ISP ha avuto una conoscenza "generica" della sua illiceità tramite comunicazioni di natura differente come, ad esempio, la diffida di chi si dichiara danneggiato dai contenuti[76]. Inoltre, anche la neutralità degli ISP è stata da tempo circoscritta dalla nostra giurisprudenza, la quale ha creato la figura del c.d. "hosting provider attivo"[77], ossia un prestatore che non si limita a ricevere passivamente i dati dai propri clienti (come fanno i cosiddetti "prestatori di servizi che agiscono come intermediari" o semplicemente "prestatori intermediari") ma che svolge su di essi attività ulteriori, quali indicizzazione, selezione, organizzazione e filtraggio, che non gli permettono di essere esonerato da responsabilità per i contenuti che ha processato[78].

L'evoluzione dell'interpretazione delle disposizioni del d.lgs. 70/2003, quindi, non pare configurare tanto la necessità di una vera e propria abolizione – come suggerito da Storm per la *Section 230* del CDA – della neutralità degli ISP rispetto alle situazioni di diffamazione compiute tramite *socialbot* sulle loro piattaforme quanto, piuttosto, una maggiore estensione dei principi espressi nelle decisioni dei tribunali in merito alla tutela dei diritti di proprietà intellettuale appena viste, che già da sola potrebbe spingere i gestori dei social ad una più attenta e rapida risposta alla lamentata diffusione di contenuti diffamatorii.

(c) *Obbligo di adeguata verifica della clientela, l'estensione dell'articolo 18 del d.lgs. 21 novembre 2007, n. 231* – Una ulteriore forma di intervento legislativo, meno stravolgente ma comunque molto incisiva, di garanzia potrebbe essere l'imposizione di controlli per l'iscrizione ad un social medium, tramite i quali compiere un'attenta verifica dell'identità della persona (o dell'ente) che voglia aprirvi un profilo.

Sebbene questa soluzione possa sembrare di complessa realizzazione (vi sono, infatti, diversi social network con centinaia di milioni di utenti) ed invasiva della privacy degli utenti, essa non sarebbe né la più innovativa né la più indiscreta di quelle proposte (vi è, infatti, chi ha addirittura proposto l'obbligo di creazione di un profilo sulle piattaforme social che abbia il proprio vero nome)[79]. Inoltre, questo genere di procedura non è nuovo nell'accesso a servizi online: l'esempio più immediato è quello dei *Digital Currency Exchanges* (DCE), delle piazze virtuali di scambio di criptovalute ed altre valute digitali accessibili al pubblico tramite Internet. Per poter usufruire di questi servizi, infatti, è necessario compiere un percorso di registrazione che comporta anche la presentazione – tramite, nella maggior parte dei casi, la scansione e l'invio con un messaggio di posta elettronica – di documenti d'identità ed altri certificati, come quello di residenza. L'obbligo di compiere questa procedura di "adeguata verifica della clientela" deriva

in Italia dalle disposizioni previste dall'articolo 18 del d.lgs. 21 novembre 2007, n. 231 (il cosiddetto "decreto antiriciclaggio"), e dal d.lgs. 25 maggio 2017, n. 90[80], che ha incluso anche gli operatori di queste piazze di scambio, detti *exchanger*, all'interno della categoria degli "operatori non finanziari soggetti alle disposizioni del decreto antiriciclaggio" (tuttavia, l'articolo 1, comma 2, lettera *ff* del d.lgs. del 2007 include in tale categoria solo coloro che permettono la «conversione da ovvero in valute aventi corso legale»). L'adozione di una procedura di tal genere farebbe sì che qualsiasi profilo che operi su di un social medium, anche qualora venga dato in gestione ad un *social bot*, sarebbe comunque riferibile ad una persona fisica che ne sarebbe il responsabile e risponderebbe personalmente degli eventuali illeciti compiuti dal software. Di conseguenza, pur non facendo da blocco all'utilizzo di algoritmi di gestione dei profili, l'adozione di procedure con questi caratteri fungerebbe da deterrente rispetto ad un loro utilizzo inappropriato, data la riferibilità di ciascuno di essi.

Ovviamente, una raccolta di dati tanto sensibili dovrebbe essere a sua volta sottoposta a garanzie di tutela della privacy degli utenti. Così, al fine di equilibrare istanze di tutela della reputazione e della privacy, sarebbe opportuno prevedere una forma di divulgazione dell'identità degli asseriti diffamatori solo a seguito di un procedimento innanzi ad un giudice che ordini ai gestori delle piattaforme la condivisione con il ricorrente delle informazioni d'identità richieste. Questa forma di bilanciamento delle garanzie è già da tempo stata adottata (senza, però gli obblighi di "adeguata verifica" sopra suggeriti) negli Stati Uniti, con formule diverse per ciascuno Stato[81], che sono state, però, anche oggetto di critiche per la mancanza di uniformità tra di loro e di chiarezza di alcune di esse.

Infine, non si può ignorare come l'implementazione da parte dei gestori dei social network in cui potrebbero essere attivi dei *social bot* – anche in violazione delle loro condizioni di utilizzo – di un sistema di accertamento così incisivo potrebbe danneggiare la loro attività d'impresa, dato che questo potrebbero facilmente scoraggiare l'iscrizione di nuovi utenti, diminuire le loro interazioni ovvero indurre alcuni a cancellarsi dalle piattaforme. Per mitigare questo rischio sarebbe anche possibile ipotizzare di imporre tale verifica solo per determinate categorie di utenti "qualificati" (quali i gestori delle cosiddette "pagine" su Facebook) ovvero di condizionare la possibilità per gli utenti di compiere determinate azioni sulle piattaforme, quali la condivisione di contenuti di terzi (ossia il "condividi" di Facebook o il "retweet" di Twitter) all'avvenuto compimento di tale verifica.

(d) *Procedimento sommario per l'ottenimento dell'identità del diffamante, la proposta del* Good Samaritan Amendment – Uno degli strumenti più utilizzati all'interno della vigente normativa americana per riuscire a perseguire gli autori di diffamazioni anonime sui social media (e su Internet in generale) è uno strumento pre-processuale (*pre-suit tool*) sviluppato in diverse declinazioni dalle corti statali degli Stati Uniti. La creazione di queste procedure di scoperta (*discovery*) dell'identità deriva, innanzitutto, dal fatto che nell'ordinamento americano la diffamazione non è considerata a livello federale un reato (anzi, in alcuni casi la Corte Suprema è addirittura intervenuta dichiarando incostituzionali alcune normative statali che sanzionavano penalmente la diffamazione)[82] e che oggi solo 15 tra Stati e Territori degli Stati Uniti hanno normative penali riguardanti la diffamazione[83]. La prima conseguenza di tale (non) inquadramento a livello federale delle condotte diffamatorie è stata quella di relegare agli aspetti del risarcimento del danno in sede civile la difesa della reputazione. Inoltre, questo sistema di reperimento dell'identità ha origine anche, da una parte, dalla riconosciuta facilità con cui è possibile pubblicare su Internet contenuti diffamanti in modo anonimo e dalle speculari difficoltà che le vittime si trovano a dover affrontare per ottenere un risarcimento e, dall'altra, dalle previsioni della *Rule* 26(d)(1) delle Regole Federali di Procedura Civile degli Stati Uniti, in base alla quale l'unico modo per poter scoprire, in caso di incertezza, chi sia la persona da far convenire in giudizio è di presentare una mozione *ex parte* (*ex-parte motion*) per ottenere un ordine giudiziale (*subpoena*) che costringa l'ISP a rivelarla (questo anche in ragione della relativa normativa federale[84] che proibisce ad un ISP di rivelare le generalità di un utente in assenza di un ordine di un tribunale e della notifica di esso all'interessato). Tuttavia, queste procedure di scoperta locali risultano spesso poco pratiche per ottenere l'identità di coloro che hanno pubblicato i contenuti, divenendo anche motivo di varie richieste di riforme legislative[85].

Una delle procedure statali che maggiormente ha incontrato il favore sia dei commentatori sia di altre corti statali è quella sviluppata dalla Corte Suprema dello Stato del Delaware[86], che ha delineato la struttura di un procedimento sommario di cognizione nella propria decisione sul caso *Doe vs. Cahill* del 2005[87]. Essa è stata successivamente suggerita come base per una riforma della *Section 230* del *Communications Decency Act* che «renderà la disposizione del "Buon Samaritano" ciò che pretende di essere: protettiva degli interessi del Primo Emendamento e degli ISP, ma comunque accomodante per meritorie

affermazioni di diffamazione»[88]. Il modello di giudizio sommario delineato dalla decisone *Cahill* comprende due presupposti perché l'attore ottenga un ordine di rivelazione: (1) che egli si sia «impegnato per informare il pubblicatore anonimo che egli è il destinatario di un mandato di comparizione *o* [l'oggetto di una] richiesta di ordine di divulgazione»; (2) che egli presenti prove sufficienti a sostenere *prima facie* ogni elemento della sua richiesta per la successiva instaurazione del processo per diffamazione.

Dato il favore incontrato dal modello di giudizio sommario tratteggiato nel caso *Cahill*, ritenuto capace di bilanciare in modo molto efficace gli interessi concorrenti di querelanti feriti ed oratori anonimi e l'intento originale del Congresso nell'emanare la *Section 230*[89], T.E. Lynch ha proposto nel 2008 un emendamento a tale sezione del CDA che codifichi al suo interno questa forma di giudizio sommario[90]. Nello specifico, il proposto *Good Samaritan Amendment*, riprendendo le disposizioni contenute nella *Section 512* del *Digital Millennium Copyright Act* (DMCA) per l'emanazione di un ordine giudiziale (*subpoena*) di rivelazione nei casi di violazione dei diritti di proprietà intellettuale da parte di soggetti anonimi, prevede che l'attore che presenti un ricorso fornisca: (1) l'indicazione del materiale ritenuto illecito, per aiutare l'ISP ad identificare l'anonimo autore del danno; (2) il testo dell'ordine da emanare; (3) prove sufficienti a sostenere *prima facie* ogni aspetto del danno rivendicato per il quale sta ricercando l'identità dell'anonimo responsabile; (4) una dichiarazione di accuratezza delle informazioni presentate; (5) la sottoscrizione dell'attore; (6) una dichiarazione giurata che la citazione è richiesta per identificare l'anonimo responsabile dell'illecito e che tali informazioni saranno utilizzate solo per indicarlo come convenuto in un'azione di risarcimento.

Tuttavia, anche in questo caso, una riforma di questa portata si mostrerebbe probabilmente più utile a scoprire più velocemente l'identità di soggetti che sui *social* fanno uso di un *alias* ma che gestiscono personalmente un profilo, piuttosto che ad arginare l'attività malevola di un *socialbot*. Anche se non si può escludere che con uno strumento di questo genere si possano comunque ottenere informazioni utili per la scoperta di chi stia utilizzando in modo non troppo esperto questo tipo di algoritmi come, ad esempio, un indirizzo IP non celato e la derivante intestazione del contratto telefonico (elementi che però, come visto al paragrafo 5 potrebbero non risultare sufficienti per la giurisdizione italiana all'identificazione dell'effettivo diffamante).

Tradurre all'interno della nostra legislazione nazionale una proposta di questo genere avrebbe il pregio di presentare con maggiore autorevolezza agli ISP stranieri una richiesta di comunicazione dei dati identificativi in loro possesso. Infatti, come già precedentemente accennato (si veda quanto rilevato al paragrafo 5, in particolare alla nota 42), al momento le indagini compiute dalla magistratura requirente – e ancor più di queste quelle difensive di parte – risultano molto spesso frustrate dai rifiuti o dalle inerzie delle loro controparti o delle imprese straniere interpellate. Per questi motivi non pare inopportuno ipotizzare l'introduzione di un procedimento sommario di cognizione che, scorrendo parallelamente alle attività realizzate dalle Autorità a seguito di querela ovvero svolgendosi in forma prodromica all'instaurazione di un giudizio civile di risarcimento, permetta a chi si dichiari vittima di diffamazione da parte di soggetti anonimi di rivolgersi direttamente ad un giudice per ottenere un ordine nei confronti dell'ISP di comunicazione dei dati in suo possesso sull'identità (o su ogni elemento utile a pervenire ad essa) dell'asserito diffamante. Per poter ottenere una pronuncia in tal senso, il ricorrente dovrebbe ovviamente presentare prova della necessità di ottenere tali informazioni ed in particolare: (1) copia della richiesta di oscuramento dei contenuti ritenuti illeciti inviata all'ISP; (2) l'indicazione di tutti i possibili riferimenti al diffamante che è stato in grado di raccogliere (primo fra tutti il profilo registrato sulla piattaforma); (3) richiesta all'ISP di informare l'utente accusato di diffamazione; (4) copia del contenuto ritenuto diffamante; (5) indicazione dei motivi che rendono *prima facie* sussistente la diffamazione a danno dell'ingiungente; (6) assunzione di responsabilità nei casi di uso o trattamento scorretti dei dati sensibili eventualmente ricevuti dall'ISP. Inoltre, in attesa che l'ISP consegni quanto richiesto (in questo caso copia di dati sensibili dell'asserito diffamante), sarebbe opportuno prevedere che in questa procedura il giudice possa anche emanare un provvedimento cautelare d'urgenza con cui ordinare il provvisorio ed immediato oscuramento dei contenuti indicati nel ricorso (ciò in ragione del fatto che si tratta di un caso di lesione di diritti della personalità, i quali potrebbero risultare irrimediabilmente pregiudicati e non più suscettibili di reintegrazione)[91]. Dovrebbe, quindi, essere previsto, in caso di mancata opposizione da parte dell'ISP o dell'accusato entro un termine ragionevole, l'obbligo per il gestore di comunicare i dati richiesti con conseguente responsabilità in caso di inadempimento. Infine, sia in caso di inutilizzabilità dei dati che di conclusione dei procedimenti di tutela o risarcimento per diffamazione, dovrebbe essere previsto un obbligo in capo all'ingiungente di cancellazione della copia dei dati ricevuti.

Sebbene, come detto, questa innovazione normativa non avrebbe le stesse ragioni di esistere che ha la sua fonte d'ispirazione d'oltreoceano, date la rilevanza penale delle diffamazioni nel nostro ordinamento ed il conseguente intervento della magistratura requirente nei casi portati alla sua conoscenza e la possibilità di ottenere un provvedimento cautelare d'urgenza ex art. 700 c.p.c., essa avrebbe due indubbi vantaggi: il primo sarebbe quello di sgravare le Procure della Repubblica di una parte delle indagini da compiere per risalire alla vera identità dell'anonimo diffamante, che verrebbe, invece, addossata alla vittima della diffamazione (il contraltare di questa previsione sarebbe, però, il fatto che i tribunali si vedrebbero investiti di un notevole numero di richieste di emanazione di decreti di "rivelazione" d'identità); il secondo sarebbe il fatto che, a differenza di quanto accade adesso, il Dipartimento di Giustizia degli Stati Uniti (dove ha sede la maggior parte dei gestori di social network) e, più in generale, le autorità di altri paesi dovrebbero valutare se adeguarsi alla decisione di un'autorità giudicante e non solo come rispondere alle richieste di un pubblico ministero.

(e) *Obbligo di adozione di strumenti di rilevazione di sospette attività bot, l'esempio della* Feinstein Bill *del 2018* – La già citata *Feinstein Bill* del 2018 prevede, tra le altre disposizioni, anche l'istituzione di un obbligo in capo ai gestori dei social media di predisporre «un processo per identificare, valutare e verificare se l'attività di qualsiasi utente del sito web dei social media è condotta da un programma software automatizzato o da un processo intesi a impersonare o replicare l'attività umana online»[92]. Il rispetto di quest'obbligo, che funge da contrafforte al divieto già analizzato *supra* alla lettera (a), nella pratica si tradurrebbe nell'adozione da parte dei prestatori di servizi informatici di metodologie e strumenti molto simili – se non uguali – a quelli utilizzati dai ricercatori che hanno già compiuto studi in questo campo[93]. Nello specifico, per giungere all'accertamento della presenza di un *social bot* dietro ad un profilo online la maggior parte degli studi condotti si è concentrata sull'analisi delle attività e delle interazioni compiute dai profili presenti su di un social medium e sulla successiva comparazione delle loro condotte con quelle di alcuni modelli comportamentali tipici degli utenti umani sulla piattaforma. Da questa comparazione è quindi possibile identificare i profili che deviano sensibilmente dagli standard utilizzati, facendo supporre con sufficiente accuratezza (tutti gli studi hanno comunque evidenziato un certo margine di incertezza) che siano gestiti da un algoritmo e non da una persona. Per rispettare, poi, l'ultimo dei termini indicati nella Feinstein Bill del 2018, ossia la verifica della presenza di un *bot* o di un essere umano dietro al profilo "deviante", la tipologia di strumenti che più facilmente potrebbe essere adottato sarebbe quella già molto diffusa dei test CAPTCHA (sebbene, come rilevato alla nota 39, alcuni di questi non siano perfettamente affidabili contro gli algoritmi più avanzati). Sebbene questi passaggi del procedimento ipotizzato possano sembrare particolarmente impegnativi ed invasivi, si consideri come l'analisi delle condotte dei profili sia già alla base dell'attività economica dei gestori delle piattaforme[94] e, quindi, ampiamente sviluppata ed affinata e come anche gli strumenti di verifica della natura umana siano ormai una costante dell'accesso ai servizi online.

Dato, però, che al momento non si ha un ampio consenso su quale sia lo strumento migliore per l'individuazione dei *social bot* sulle piattaforme[95] e che la struttura estremamente mutevole di questi algoritmi (i quali, come si è brevemente accennato, nelle loro forme più raffinate sono anche in grado di adattarsi autonomamente ai cambiamenti del contesto digitale in cui operano) e la loro rapidissima evoluzione difficilmente renderanno possibile giungere all'individuazione di uno strumento o un processo perfettamente affidabili per la loro rilevazione, pare opportuno evitare di legare la conformità ad una prescrizione di tal genere all'adozione di uno specifico strumento individuato da una fonte normativa o da un'autorità indipendente. Questo in quanto ciò indicherebbe con chiarezza anche a soggetti o entità – anche stranieri – quale sarebbe l'unico ostacolo digitale da dover aggirare per poter nuovamente infiltrare i propri *socialbot* sulle piattaforme, spingendoli ad agire con quest'unico scopo, e di conseguenza, finirebbe col permettere di frustrare con eccessiva facilità l'intrinseco scopo di efficace e rapida tutela degli utenti umani della normativa. Di conseguenza, parrebbe opportuno istituire, piuttosto, un obbligo in capo agli ISP di «adottare ed efficacemente attuare un processo idoneo a valutare, identificare e verificare se l'attività di qualsiasi utente del sito web dei social media è condotta da un programma software automatizzato o da un processo intesi a impersonare o replicare l'attività umana online».

L'introduzione di un obbligo che riecheggia quello di adozione di «modelli di organizzazione e di gestione idonei a prevenire reati» sancito dal d.lgs. 8 giugno 2001, n. 231[96], avrebbe, dato il carattere estremamente dinamico e mutevole appena visto dei fenomeni che intende contrastare, il pregio di spingere alcune delle più grandi società del settore informatico ad impegnarsi attivamente nel contrasto ai *socialbot*, lasciandogli comunque la flessibilità di scegliere quali strumenti adoperare – o anche creare –

per meglio rispondere alle loro esigenze in quest'ambito e di modificarli ed adattarli con grande flessibilità ed immediatezza. In aggiunta, questo tipo di attività di valutazione dei profili sui social avrebbe l'indubbio vantaggio, rispetto ad uno di "adeguata verifica della clientela" ipotizzata *supra* alla lettera (c), di non rallentare l'iscrizione alle piattaforme da parte di nuovi utenti né darebbe loro l'impressione di subire una profonda intrusione nella loro privacy.

7.2. Riformulazione delle interpretazioni giurisprudenziali in materia di responsabilità degli ISP

Al di fuori dell'ambito legislativo, una considerazione rilevante nell'ambito della tutela della reputazione sui social media, che tuttavia avrebbe al massimo un effetto indiretto sul contrasto alla presenza ed alle attività dei *socialbot*, riguarda i casi in cui possa effettivamente sussistere un'esenzione di responsabilità in capo agli ISP (aspetto che, appunto, può avere solamente la conseguenza riflessa di spronarli ad agire con maggiore impegno sulle proprie piattaforme per arginare l'attività di questi algoritmi "malevoli"). Infatti, relativamente alla già citata figura di genesi giurisprudenziale del c.d. "hosting provider attivo", ossia di un prestatore di servizi della società dell'informazione che non si limita a ricevere passivamente i dati dai propri clienti ma che svolge su di essi attività ulteriori, vi è preliminarmente da rilevare come, sebbene non sia prevista dalla nostra legislazione, essa trovi comunque giustificazione nel considerando 42 della direttiva 2000/31/CE, il quale dispone che l'assenza di responsabilità per gli ISP riguarda esclusivamente il caso in cui la loro attività «è di ordine meramente tecnico, automatico e passivo, il che implica che il prestatore di servizi della società dell'informazione non conosce né controlla le informazioni trasmesse o memorizzate». È proprio sulla scorta di tale considerazione che la Corte di giustizia dell'Unione europea ha stabilito nella propria decisione sul caso *L'Oréal SA c. eBay*[97] (invero relativa alla violazione dei diritti di proprietà intellettuale e industriale) che la neutralità dell'ISP non può essere invocata se «il prestatore del servizio, anziché limitarsi ad una fornitura neutra di quest'ultimo, mediante un trattamento puramente tecnico e automatico dei dati forniti dai suoi clienti, svolge un ruolo attivo atto a conferirgli una conoscenza o un controllo di tali dati» e che il ruolo del fornitore di servizi si considera attivo «allorché presta un'assistenza che consiste in particolare nell'ottimizzare la presentazione delle offerte in vendita di cui trattasi o nel promuoverle».

Le considerazioni formulate ormai dieci anni fa all'interno della decisone *L'Oréal* dovrebbero oggi essere rivalutate alla luce dei più recenti sviluppi nell'ambito delle attività informatizzate di analisi dei contenuti e spingere le interpretazioni giurisprudenziali delle vigenti normative a riconsiderare entro quale tipologia di attività rientrino oggi quelle compiute da gestori delle piattaforme social. Se, infatti, ai tempi della citata sentenza e delle successive sui casi *Scarlet*[98] e *Papasavvas c. Fileleftheros*[99] un sistema informatico di controllo completo dei contenuti avrebbe potuto risultare complesso e costoso e, quindi, non imponibile da una normativa nazionale (in quanto contrario alle condizioni stabilite dall'art. 3, n. 1, della direttiva 2004/48), l'attuale avanzamento delle tecnologie digitali non può che mettere in dubbio tali assunti[100], in particolare per quanto riguarda i contenuti di tipo testuale e sonoro.

Infatti, negli ultimi anni si è assistito ad un notevole incremento d'interesse circa lo sviluppo degli algoritmi di analisi del linguaggio[101], tanto che oggi si suole identificare come branca autonoma dell'informatica lo studio degli algoritmi di elaborazione del linguaggio naturale (in lingua inglese *natural language processing* o NLP) con cui i computer possono essere usati per comprendere e manipolare testi o discorsi umani[102]. In quest'ambito, poi, la raffinatezza di certi algoritmi di analisi testuale è tale da averli resi oggi anche degli utili e sufficientemente affidabili strumenti di analisi psicometrica delle persone attive sui social media[103], tanto che diverse compagnie li impiegano per migliorare la loro offerta di servizi di micro-targeting politico (in lingua inglese *political microtargeting* o PMT) con un'efficacia tale da aver fatto tendere alcuni a qualificare questa pratica come una manipolazione dell'opinione pubblica[104]. Di conseguenza, pare oggi difficile che un ISP, per quanto riguarda almeno i contenuti testuali pubblicati su una sua piattaforma, possa argomentare che, riprendendo le parole dell'articolo 17 del d.lgs. 70/2003, l'adozione di strumenti di «sorveglianza sulle informazioni che trasmette o memorizza» sia per lui oggi eccessivamente complessa ed onerosa (ciò anche alla luce del fatto che, come evidenziato alla nota 101, diverse migliaia di progetti NLP sono disponibili gratuitamente e corredati di documentazioni che ne illustrano le funzionalità)[105].

In linea con l'idea della necessità di una rivalutazione della responsabilità degli ISP alla luce delle odierne tecnologie informatiche si sono recentemente mostrati anche i giudici di Roma (invero sempre con riferimento a casi di violazione dei diritti di proprietà intellettuale)[106]. In particolare, nella loro decisione del 10 gennaio 2019 i magistrati capitolini hanno ri-

levato che, come già precisato dalla Corte di giustizia dell'Ue nelle citata sentenza *L'Oréal*, la responsabilità del prestatore attivo di servizi hosting sorge ogni qual volta vi sia anche «un pur minimo contributo all'editing del materiale memorizzato lesivo di diritti tutelati», ossia anche quando questi «interviene nell'organizzazione e selezione del materiale trasmesso», e che, in conseguenza di ciò, «non appare condivisibile quella giurisprudenza che limita il ruolo attivo dell'hosting provider al solo caso in cui il gestore operi sul contenuto sostanziale del video caricato sulla piattaforma». Inoltre, il Tribunale della capitale, facendo riferimento alla dettagliata C.T.U. relativa alla natura dell'attività dell'impresa convenuta (Vimeo, una controllata del gruppo InterActiveCorp), ha evidenziato come questa gestisca un «sito di rete sociale ... dove i contenuti audiovisivi sono precisamente catalogati, indicizzati e messi in correlazione tra loro dalla stessa convenuta» anche attraverso l'utilizzo di algoritmi automatizzati di terze parti (nello specifico il servizio *AdSense* di Google), giungendo a concludere che «un sistema tecnologico così avanzato e sofisticato è del tutto incompatibile con la figura dell'hosting provider "passivo" ... ed integra invece quella di content-provider» e, soprattutto, che perché sorga la responsabilità del prestatore attivo di servizi hosting «è sufficiente che i mezzi tecnologici dallo stesso utilizzati siano comunque idonei a conferirgli la conoscenza o il controllo dei dati memorizzati».

Di simile avviso, precisamente nell'ambito della diffamazione esasperata dall'utilizzo di algoritmi per la propagazione dei contenuti, è stata anche la Corte Suprema dello Stato australiano di Victoria, la quale nella sua decisione sul caso *Trkulja v. Google Inc LLC*[107] ha concordato con la giuria nel ritenere che ad un motore di ricerca online (nel caso di specie Google) possa essere attribuita la responsabilità da editore quando la diffusione di contenuti diffamatorii derivi da un funzionamento del suo algoritmo coerente con la sua programmazione, ancor prima di aver ricevuto alcuna notificazione in merito al carattere illecito di questi. In altre parole, il giudice australiano ha ritenuto che l'utilizzo di software per la catalogazione ed organizzazione automatica di contenuti pubblicati da terze parti, data la prevedibilità dei risultati ottenibili originata dalla progettazione del loro codice sorgente, impedisca agli ISP di beneficiare delle esenzioni di responsabilità garantite ai prestatori intermediari di servizi informatici.

Tuttavia, nonostante gli indubbi progressi compiuti nell'ambito delle strumentazioni di sorveglianza ed analisi dei contenuti pubblicati sui social media, non pare comunque opportuno attribuire in modo automatico la responsabilità da "hosting provider attivo" a qualunque impresa che utilizzi software ed algoritmi gestionali per l'amministrazione dei contenuti caricati dagli utenti, in quanto i gestori delle piattaforme non tendono ad operare alcun tipo di *technical disclosure* delle tecnologie impiegate, impedendo così un giudizio *ex ante* sul loro effettivo contributo all'editing del materiale memorizzato[108].

7.3. Innovazioni digitali

Oltre alle ipotesi di interventi legislativi pare qui opportuno accennare al fatto che, allo stato attuale, vi siano già delle soluzioni tecniche messe a punto da alcuni gestori di piattaforme social che meritano un'attenta considerazione, se non anche un invito ad adottarle diretto agli altri gestori da parte delle Autorità competenti, dato il loro possibile carattere di argine alle attività dei *socialbot*, nonché la loro dimostrata possibilità di essere implementate senza risultare eccessivamente complesse ed onerose.

(a) *Indicazione degli utenti la cui identità è stata verificata, l'esempio di Tinder* – In termini non troppo dissimili da quelli ipotizzati *supra* al paragrafo 7.1.(c) nell'agosto del 2021 il gruppo InterActiveCorp (IAC), gestore dell'applicazione per smartphone Tinder, ha autonomamente introdotto a livello globale la verifica dell'identità dei propri utenti tramite l'invio di copia di un loro documento d'identità[109]. Tuttavia, a differenza dell'ipotizzata istituzione di un vero e proprio obbligo generale di "adeguata verifica della clientela", il compimento di tale accertamento attualmente avviene solo su base volontaria e la sua mancata realizzazione non comporta alcuna limitazione nell'utilizzo dei servizi offerti dalla piattaforma, anche se aggiunge al profilo di coloro che si sono sottoposti a tale procedura un marchio, visibile a tutti gli altri, che lo certifica. Inoltre, vi è da considerare come sia la natura stessa del servizio offerto da quest'applicazione, ossia permettere di stringere nuove conoscenze in rete e facilitare gli incontri di persona, a rendere non solo opportuno ma anche maggiormente rassicurante per i suoi utenti un accertamento preventivo dell'identità degli altri utilizzatori. In altre parole, l'esempio di Tinder, sebbene volto ad una maggiore tutela dei propri utenti, trova una ragione per la propria implementazione anche nel fatto che in questo modo è in grado di offrire un servizio più affidabile e protetto e, di conseguenza, attrarre più clienti.

Nonostante le particolarità della piattaforma appena considerata, vi è comunque il fatto che l'applicazione di un marchio di avvenuto controllo del profilo a seguito di invio volontario dei propri documenti avrebbe senz'altro l'effetto di rendere maggiormente credibili i contenuti pubblicati tramite es-

so[110], ponendosi, quindi, come incentivo a sottoporsi a questo genere di procedura e controbilanciando il rischio di rinuncia dovuta ad attese eccessivamente lunghe perché venga completata. In ogni caso, il programma di verifica della IAC dimostra la fattibilità tecnica di una soluzione di accertamento dell'identità degli utenti di un social network che permetterebbe così, almeno, di identificare con certezza il responsabile di qualsiasi illecito compiuto tramite uno o più profili aperti sulla piattaforma senza risultare eccessivamente complessa ma solo lunga.

(b) *Strumenti di segnalazione di contenuti accusati di essere diffamatorii* – Infine, una soluzione adottata da un altro prestatore di servizi della società dell'informazione che merita di essere considerata in questa sede è quella di Twitter rispetto ai contenuti pubblicati sulla sua piattaforma in violazione delle sue condizioni di utilizzo da parte di determinati profili. Nello specifico, la soluzione di apposizione di un pannello di avvertimento che nascondesse il contenuto di alcuni dei tweet pubblicati dall'allora Presidente degli Stati Uniti Donald J. Trump durante l'ultimo periodo del suo mandato (si veda l'esempio in Figura 3)[111].

Una soluzione di questo genere, impiegata originariamente solo per contrastare la disinformazione propugnata dal profilo di un soggetto pubblico sempre al centro dell'attenzione, potrebbe essere adottata da qualsiasi altro social network ed estesa a qualsiasi contenuto pubblicato, la cui attivazione, però, potrebbe essere demandata ad una verifica "dal basso". Se, infatti, i tweets dell'allora POTUS venivano controllati "dall'alto", ossia direttamente da degli addetti della piattaforma, la "copertura" con un pannello dei contenuti contestati come diffamatorii potrebbe avvenire tramite selezione di un'apposita opzione da parte di qualsiasi altro utente (azione che potrebbe essere definita con i termini inglesi "*wall-by-click*"). Questo permetterebbe un'interruzione decisamente più subitanea della spirale di propagazione di contenuti illeciti o lesivi, dato che anche la ri-condivisione del medesimo post o tweet mostrerebbe comunque questo pannello o *wall*, prevenendo una diretta visibilità di termini, immagini o video capaci di danneggiare la reputazione della persona contro cui sono diretti. Questa previsione, in particolare, sarebbe un valido contrasto alla sovradiffusione ottenibile tramite *socialbot* programmati, appunto, a ri-condividere i contenuti dei profili che sono programmati a seguire con assiduità.

Per evitare comunque una totale paralisi della diffusione di qualsiasi tipo di contenuto, sarebbe opportuno adottare anche la possibilità (visibile anche in Figura 3) di accedere comunque al contenuto contestato tramite la selezione di un apposito tasto (nell'esempio riportato molto semplicemente etichettato con la parola "View" ossia "visualizza"). Questa previsione, informando preventivamente gli eventuali lettori della possibile illiceità del contenuto e, nei casi di diffamazione, di possibile falsità dello stesso, permetterebbe così un bilanciamento tra la libertà di espressione e la tutela della reputazione di tutti coloro che vengano menzionati in dette espressioni; reputazione che verrebbe in parte protetta dalla messa in dubbio della veridicità delle parole contrassegnate come diffamatorie.

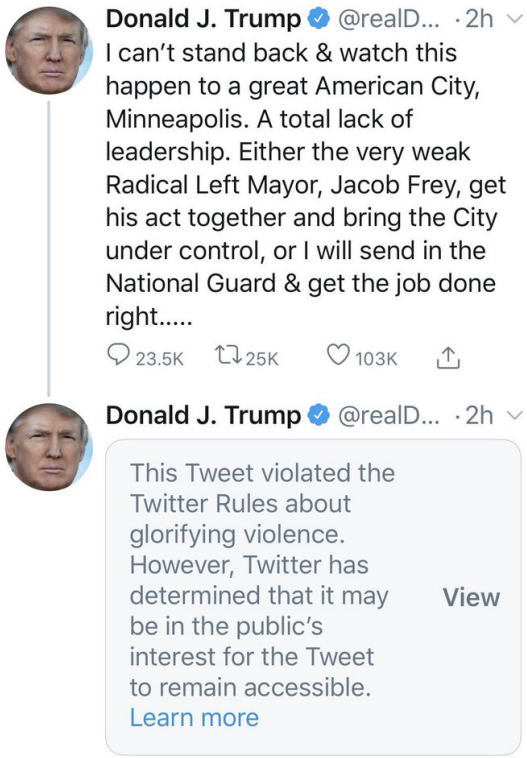


Figura 3: *La piattaforma Twitter optò per un "oscuramento" dei tweet dell'allora Presidente degli Stati Uniti che violavano i suoi termini di utilizzo. Il secondo tweet, in particolare, era ritenuto un incitamento alla violenza. Si noti come, comunque, il contenuto fosse visualizzabile tramite il tasto "View" posto sulla destra*

Infine, allo scopo di evitare abusi di un'opzione di questo genere, che già in parte si potrebbero predire sulla base delle critiche mosse ad altre previsioni come il *Digital Millennium Copyright Act* (DMCA)[112], sarebbe comunque necessario costituire un sistema interno all'organizzazione del gestore di revisione del contenuto segnalato, che permetterebbe anche una eventuale eliminazione del *wall* nel caso in cui esso

sia valutato come non offensivo (e possibili sanzioni, quali la sospensione temporanea, nei casi di abuso di questo strumento). Sebbene ciò possa sembrare poco incisivo nel tutelare la reputazione di chi si ritenga diffamato, si consideri che con questa soluzione rimarrebbe, comunque, sempre la possibilità di contestare la lesività del contenuto di fronte ad un tribunale con, assieme, la garanzia di evitare un'eccessiva diffusione di un contenuto fuorviante circa le qualità della persona coinvolta. Inoltre, con questo strumento il gestore della piattaforma non potrebbe opporre di non essere stato posto a conoscenza dell'illiceità del contenuto, essendo stato avvisato tramite uno strumento da lui stesso predisposto, e finirebbe così con l'integrare pienamente i presupposti sanciti in modo estensivo dalla giurisprudenza[113] per l'attribuzione di responsabilità – prevista dall'articolo 16 del d.lgs. 70/2003 – da mancata rimozione di contenuti di cui gli ISP siano stati portati a conoscenza. Inoltre, l'ipotizzato strumento di "*wall-by-click*" istituirebbe uno standard univoco di comunicazione al prestatore di servizio che consentirebbe di superare le problematiche di genericità delle diffide extragiudiziali, riguardo alle quali la giurisprudenza si è interrogata se la necessità di indicare i contenuti fosse soddisfatta «mediante la mera indicazione del nome ... e simili elementi descrittivi, oppure occorra anche la precisa indicazione del cd. Indirizzo URL»[114].

Ovviamente, questo sistema di controllo "dal basso" dei contenuti dovrebbe trovare dei limiti nei confronti dei profili di certe figure professionali – riconosciute e verificate – come i giornalisti, altrimenti ci sarebbe il concreto rischio che qualsiasi articolo inerente a personaggi pubblici o a questioni controverse (o semplicemente al centro del dibattito) verrebbe da questi soggetti o dai loro sostenitori (veri o artificiali che siano) immediatamente "segnalati" pubblicamente come diffamatorii, finendo con ostacolare il diritto di cronaca ed erodere la credibilità di determinati professionisti, già sottoposti ai limiti dei codici di condotta ed ai controlli degli ordini professionali, e finendo anche col rischiare di spostare il cosiddetto "problema *bot*" dalla condivisione di contenuti illeciti all'ostruzionismo verso contenuti legittimi.

8. Conclusioni

In conclusione, non si può ignorare come il mutato contesto degli strumenti di propagazione del pensiero con il quale ci ritroviamo a confrontarci oggigiorno ci ponga nell'urgenza di non rimanere più solo concentrati su quale sia il corretto bilanciamento tra libertà di pensiero e tutela della reputazione di una persona ma, soprattutto, di estendere il dibattito anche a quali strumenti implementare per tutelare effettivamente questi diritti fondamentali. Questo perché, mentre ci si interroga su quale sia il giusto punto di equilibrio tra queste due istanze, il progresso tecnologico e l'attuale strutturazione dei social network permettono a chiunque, grazie a strumenti relativamente facili da usare quali sono i *social bot*, di diffondere nel giro di poche ore insinuazioni, maldicenze ed offese con una tale ampiezza e velocità da distruggere irrimediabilmente la reputazione non solo di una persona ma addirittura di interi gruppi sociali (portando in certi casi perfino a conseguenze violente per le vittime e a veri e propri squarci nel tessuto sociale)[115], rimanendo comodamente nell'anonimato e irraggiungibile per le autorità a cui le vittime potrebbero rivolgersi nel (vano) tentativo di ottenere giustizia o almeno l'interruzione delle condotte poste in essere in loro danno (come chi scrive ha visto succedere).

Si è voluto, così, presentare diversi possibili interventi – sulla effettiva validità nella pratica di alcuni dei quali si ammette una certa incertezza – che potrebbero mitigare un aspetto particolare del cosiddetto "problema *bot*", quello della diffamazione dei singoli, nella speranza di poter fornire dei validi suggerimenti per affrontare una questione spesso ignorata o mal conosciuta o, quantomeno, portare ad una maggiore consapevolezza dei rischi e delle possibili soluzioni alla presenza sui social media di profili gestiti da algoritmi e nascosti dietro a false identità, riservandosi per il futuro di trattare con sufficiente approfondimento anche i possibili strumenti di contrasto alle attività di diffusione di falsità riguardanti eventi o verità scientifiche.

Note

[1]Si vedano, *ex multis*, N. Newman, *The rise of social media and its impact on mainstream journalism*, in "Reuters Institute for the Study of Journalism: Working Papers", 2009, 9 p.; H.S. Noor Al-Deen, J.A. Hendricks (eds.), *Social Media usage and impact*, Lexington Books, 2011, 328 p.; A. Guzman, F. Vis, *6 Ways Social Media is Changing the World*, 2016; S. Siddiqui, T. Singh, *Social Media its Impact with Positive and Negative Aspects*, in "International Journal of Computer Applications Technology and Research", vol. 5, 2016, n. 2, p. 71-75; S. Stieglitz, F. Brachten, B. Ross, A.-K. Jung, *Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts*, in "Proceedings of the 28th Australasian Conference on Information Systems (ACIS)", 2017, n. 89.

[2]Si vedano, *ex multis*, D. Rolph, *Defamation by Social Media*, in "Precedent", 2013, n. 117, p. 16-21; H.F. Kuehl, *Free Speech and Defamation in an Era of Social Media: An Analysis of Federal and Illinois Norms in the Context of Anonymous Online Defamers*, in "Northern Illinois University Law Review", 2016, vol. 36, n. 3, p. 28-56; E. Frederick, *Malice in the Digital Palace: A Commentary on Athletes, So-*

cial Media, and Defamation, in "Journal of Legal Aspects of Sport", vol. 27, 2017, n. 1, p. 79-89.

[3]Si vedano, *ex multis*, S. Utz, *The (Potential) Benefits of Campaigning via Social Network Sites*, in "Journal of Computer-Mediated Communication", vol. 14, 2009, n. 2, p. 221-243; S. Stieglitz, L. Dang-Xuan, *Social media and political communication: a social media analytics framework*, in "Social Network Analysis and Mining", vol. 3, 2013, n. 4, p. 1277-1291; G.S. Enli, E. Skogerbø, *Personalized Campaigns in Party-centred Politics Twitter and Facebook as Arenas for Political Communication*, in "Information, Communication & Society", vol. 16, 2013, n. 5, p. 757-774; G.S. Enli, *Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election*, in "European Journal of Communication", vol. 32, 2017, n. 1, p. 50-61; B. McLaughlin, T. Macafee, *Becoming a Presidential Candidate: Social Media Following and Politician Identification*, in "Mass Communication and Society", vol. 22, 2019, n. 5, p. 584-603.

[4]Si noti come in S. Stieglitz, L. Dang-Xuan, *op. cit.*, gli autori, facendo riferimento a S.L. Wartick, J.F. Mahon, *Toward a substantive definition of the corporate issue construct. A review and synthesis of the literature*, in "Business & Society", vol. 33, 1994, p. 293-311, e a D. Ingenhoff, U. Röttger, *Issues management Ein zentrales Verfahren der Unternehmenskommunikation*, in M. Meckel, B.F. Schmid (Hrsg), "Unternehmenskommunikation. Kommunikationsmanagement aus Sicht der Unternehmensführung", Gabler, 2008, p. 323-354, riconoscano il ruolo fondamentale occupato in politica dall'intuizione dei sentimenti dell'elettorato e facciano presente come oggi sia più semplice intercettarli tramite l'utilizzo di strumenti di analisi informatica dei contenuti pubblicati sui *social media*.

[5]Fonte: Statista.

[6]S. Stieglitz, F. Brachten, B. Ross, A.K. Jung, *op. cit.*; R. Gorwa, D. Guilbeault, *Unpacking the Social Media Bot: A Typology to Guide Research and Policy*, in "Policy & Internet", vol. 12, 2020, n. 2, p. 225-248, in cui gli autori sottolineano come molte delle soluzioni proposte a questo problema «si basino nella migliore delle ipotesi su delle scarse capacità tecniche».

[7]A. Leonard, *Bots: The Origin of the New Species*, Penguin Books, 1998, 269 p.

[8]Z. Yang, C. Wilson, X. Wang et al., *Uncovering Social Network Sybils in the Wild*, in "ACM Transactions on Knowledge Discovery from Data", vol. 8, 2014, n. 1, 29 p.

[9]R. Gorwa, D. Guilbeault, *op. cit.*

[10]Si vedano in merito L. Alvisi, A. Clement, A. Epasto et al., *SoK: The Evolution of Sybil Defense via Social Networks*, in "IEEE Symposium on Security and Privacy", 2013, p. 382-396; E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, in "Communications of the ACM", vol. 59, 2016, n. 7, p. 96-104.

[11]M. Tsvetkova, R. Garcia-Gavilanes, L. Floridi, T. Yasseri, *Even good bots fight: The case of Wikipedia*, in "Plos One", vol. 12, 2017, n. 2, che riprende quanto affermato in precedenza in S. Franklin, A. Graesser, *Is It an Agent, or Just a Program? A Taxonomy for Autonomous Agents*, in J.P. Müller, M.J. Wooldridge, N.R. Jennings (eds.) "Intelligent Agents III. Agent Theories, Architectures, and Languages", Springer, 1996, p. 21-35.

[12]M. Forelle, P. Howard, A. Monroy-Hernández, S. Savage, *Political Bots and the Manipulation of Public Opinion in Venezuela*, in "SSRN Electronic Journal", 2015, p. 1-8.

[13]N. Abokhodair, D. Yoo, D.W. McDonald, *Dissecting a Social Botnet: Growth, Content and Influence in Twitter*, in "Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)", Association for Computing Machinery, 2015, p. 839-851.

[14]Y. Boshmaf, I. Muslukhov, K. Beznosov, M. Ripeanu, *The Socialbot Network: When Bots Socialize for Fame and Money*, in "Proceedings of the 27th Annual Computer Security Applications Conference (ACSAC '11)", Association for Computing Machinery, 2011, p. 93-102.

[15]Cfr. quinta sezione penale della Corte di cassazione, sentenza del 23 luglio 2020 n. 22049, in cui viene specificamente affermato che «il reato di sostituzione di persona è integrato da colui che crea ed utilizza un profilo su social network, utilizzando abusivamente l'immagine di una persona del tutto inconsapevole, trattandosi di condotta idonea alla rappresentazione di una identità digitale non corrispondente al soggetto che lo utilizza» richiamando anche la propria precedente sentenza del 08 giugno 2018 n. 33862.

[16]E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, cit.

[17]*Ibidem*; D. Guilbeault, *Growing Bot Security: An Ecological View of Bot Agency*, in "International Journal of Communication", vol. 10, 2016, p. 5003-5021. In quest'ultimo l'autore sottolinea come allo stato attuale i bot non siano dotati dai loro creatori di una vera e propria intelligenza artificiale, dando comunque atto che le loro versioni più raffinate sono in grado addirittura di sviluppare in autonomia cicli di vita dinamici con cui imitare i comportamenti umani.

[18]Esistono diverse forme di licenze d'uso gratuito di software: dalla licenza *freeware* per i software proprietari alla licenza *GNU General Public License* per i software liberi.

[19]La *suite di strumenti per l'automazione dei browsers web Selenium* è sottoposta alla licenza d'uso *Apache license 2.0*, una licenza libera che consente agli utenti di usare il software per ogni scopo, di distribuirlo, modificarlo e di distribuire versioni modificate del software e che impone solamente di preservare l'informativa di diritto d'autore e d'esclusione di responsabilità.

[20]Si veda a questo proposito B. Kollanyi, *Where Do Bots Come From? An Analysis of Bot Codes Shared on GitHub*, in "International Journal of Communication", vol. 10, 2016, p. 4932-4951 in cui viene illustrata l'ampiezza dei numeri di codici sorgente disponibili sulla piattaforma GitHub (l'autore ne ha contati più di 4.000 solo per i bot utilizzabili su Twitter).

[21]Si veda a questo riguardo J. Messias, L. Schmidt, R. Oliveira, F. Benevenuto, *You followed my bot! Transforming robots into influential users in Twitter*, in "First Monday", vol. 18, 2013, n. 7, in cui gli autori descrivono la creazione di due differenti bot, dotati di due differenti (sebbene non troppo) codici sorgente e conseguentemente capaci di compiere azioni diverse all'interno dei social network in cui hanno operato.

[22]A queste conclusioni è giunta anche la Corte Suprema dello Stato australiano di Victoria nella propria decisione *Trkulja v. Google Inc. & Anor* [2012] VSC 533. Nel caso di specie la giuria ha ritenuto che i motori di ricerca siano responsabili della pubblicazione dei materiali diffamatorii che vengono assemblati in modo automatizzato dai loro software.

[23]Secondo una ricerca pubblicata sul *The New York Times*, a cura di Nicole Perlroth e condotta dagli informatici Andrea Stroppa e Carlo de Micheli, il mercato della vendita di questi seguaci artificiali ha ormai raggiunto dimensioni multimilionarie, cfr. N. Perlroth, *Fake Twitter Followers Becomes Multimillion Dollar Business*, April 5, 2013. In merito, alcuni hanno ipotizzato l'esistenza di un sottotipo di mercato dei bot, specializzato nella diffusione di disinformazione politica. Si vedano K. Starbird, J. Maddock, M. Orand et al., *Rumours, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing*, in "iConference 2014 Proceedings", 2014, p. 654-662; E. Ferra-

ra, *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election*, in "First Monday", vol. 22, 2017, n. 8.

[24]N. Perlroth, *op. cit.*

[25]Si veda a questo riguardo quanto accuratamente descritto in T.C. Helmus, E. Bodine-Baron, A. Radin et al., *Russian Social Media Influence Understanding Russian Propaganda in Eastern Europe*, Rand Corporation, 2018, 149 p.

[26]Y. Xie, F. Yu, Q. Ke et al., *Innocent by association: Early recognition of legitimate users*, in "Proceedings of the 2012 ACM Conference on Computer and Communications", Association for Computing Machinery, 2012, p. 353-364; D.W. Westerman, P.R. Spence, B. Van Der Heide, *A social network as information: The effect of system generated reports of connectedness on credibility on Twitter*, in "Computers in Human Behavior", vol. 28, 2012, n. 1, p. 199-206; M. Forelle, P. Howard, A. Monroy-Hernández, S. Savage, *op. cit.*; E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, cit.; S.C. Woolley, *Automating power: Social bot interference in global politics*, in "First Monday", vol. 21, 2016, n. 4; D. Guilbeault, *op. cit.*

[27]T. Santamato, *Su Twitter 48 milioni di profili sono 'robot'*, 2017, intervista rilasciata dal professor Filippo Menczer all'ANSA.

[28]O. Varol, E. Ferrara, C. A. Davis et. al., *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, in "Proceedings of the 11th International AAAI Conference on Web and Social Media", AAAI Press, 2017, p. 280-289.

[29]La legge italiana del 08 febbraio 1948, n. 47, prevede l'obbligo di registrazione di qualsiasi pubblicazione periodica presso la cancelleria del Tribunale nella cui circoscrizione la pubblicazione deve effettuarsi. Per le pubblicazioni periodiche diffuse esclusivamente online la registrazione è obbligatoria solo qualora intendano avvalersi delle provvidenze in favore dell'editoria previste dalla legge del 7 marzo 2001, n. 62.

[30]E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, cit.

[31]Infatti, come rilevato dai giudici della Cassazione nella sentenza del 29 gennaio 2015 n. 31022 (in particolare, al punto 18 della motivazione), i social media rientrano tra i «nuovi mezzi, informatici e telematici, di manifestazione del pensiero». Inoltre, la classificazione del «vasto ed eterogeneo ambito della diffusione di notizie ed informazioni» operata dalla Cassazione è stata da questa successivamente posta alla base della propria più recente decisione del 20 giugno 2019 n. 27675.

[32]Questi aspetti di velocità e diffusività sono stati di recente sottolineati dalla Corte costituzionale nella sua ordinanza del 26 giugno 2020 n. 132, in cui viene rilevata la «rapidissima e duratura amplificazione degli addebiti diffamatori determinata dai social networks e dai motori di ricerca in Internet, il cui carattere lesivo per la vittima – in termini di sofferenza psicologica e di concreti pregiudizi alla propria vita privata, familiare, sociale, professionale, politica – e per tutte le persone a essa affettivamente legate risulta grandemente potenziato rispetto a quanto accadeva anche solo in un recente passato».

[33]Cfr. prima sezione penale della Corte di cassazione, sentenza del 2 gennaio 2017 n. 50. Negli stessi termini si sono poi espressi anche i giudici della quinta sezione penale nella loro sentenza del 23 gennaio 2017 n. 8482.

[34]Cfr. prima sezione penale della Corte di cassazione, sentenza del 2 gennaio 2017 n. 50.

[35]Cfr. quinta sezione penale della Corte di cassazione, sentenza del 1° febbraio 2017 n. 4873. Questo inquadramento del reato di diffamazione tramite l'uso dei social media è stato poi ripreso nella sentenza del 6 settembre 2018 n. 40083, che ha precisato che «l'eventualità che fra i fruitori del messaggio vi sia anche la persona a cui si rivolgono le espressioni offensive» non permette comunque di «mutare il titolo del reato nella diversa ipotesi di ingiuria», e dalla Corte d'appello civile di Napoli nella sua sentenza del 16 gennaio 2019.

[36]In particolare, come rilevato dalla quinta sezione penale della Corte di cassazione nella sentenza del 15 maggio 2018 n. 21521, il giudice può disporre il sequestro preventivo di gruppi Facebook, qualora ritenga che vi sia il pericolo di aggravamento o di protrazione delle conseguenze del reato o di agevolazione dello stesso.

[37]Si veda a questo proposito la quinta sezione penale della Corte di cassazione, sentenza del 22 novembre 2017 n. 5352, in cui i giudici riconoscono questa eventualità e che, quindi, «a prescindere dal nickname utilizzato, l'accertamento dell'IP di provenienza del post può essere utile».

[38]Si veda riguardo a questa seconda eventualità quanto rilevato dalla quinta sezione penale della Corte di cassazione nella sentenza dell'8 giugno 2018 n. 33862 e nella sentenza del 6 luglio 2020 n. 22049.

[39]Anche strumenti di verifica come il CAPTCHA non sono immuni da possibili forzature e superamenti tramite l'utilizzo di algoritmi ed intelligenze artificiali. Al riguardo si vedano N. Perlroth, *op. cit.*, e A. Geitgey, *How to break a CAPTCHA system in 15 minutes with Machine Learning*, December 13, 2017, nonché un progetto open-source di un algoritmo volto ad infrangere il sistema CAPTCHA usato dal servizio di *micro-blogging* Weibo.

[40]Cfr. quinta sezione penale della Corte di cassazione, sentenza del 22 novembre 2017 n. 5352). Si veda anche quanto riconosciuto dalla Corte Suprema dello Stato del Delaware in *Doe v. Cahill*, 884 A.2d 451 (Del. 2005).

[41]Cfr. quinta sezione penale della Corte di cassazione, sentenza del 29 febbraio 2016 n. 8275, in cui i giudici hanno rilevato come questa eventualità sia materialmente possibile, sebbene nel caso di specie non configurasse un ragionevole dubbio, data la sussistenza di altri fattori.

[42]Il Dipartimento di Giustizia degli Stati Uniti, in una nota del 2016 alla Procura della Repubblica presso il Tribunale di Roma, ha affermato il proprio rifiuto a fornire informazioni in questi casi. Sull'argomento si veda il breve accenno fornito in V. Pacelli, *Diffamazione, la giustizia Usa: "Per Facebook sono opinioni". E gli haters restano impuniti*, in "Il Fatto Quotidiano", 25 marzo 2018.

[43]H.F. Kuehl, *op. cit.*; B. Storm, *The Man Behind the Mask: Defamed Without a Remedy*, in "Northern Illinois University Law Review", vol. 33, 2013, n. 2, p. 393-422.

[44]Vi sono anche dei browser di utilizzo comune, come Opera e Firefox, che hanno questa funzione di mascheramento dell'indirizzo IP già integrata e disponibile gratuitamente (seppure con dei limiti di traffico dati).

[45]Questi strumenti con i quali i dati scambiati in rete vengono fatti transitare attraverso dei server che fungono da intermediari e che forniscono al server che ospita materialmente il social media i propri indirizzi IP in vece di quello del client finale (il gestore del profilo diffamante). Così, ad esempio, un soggetto italiano potrebbe avere profilo su di un social media e accedervi attraverso un computer localizzato in Italia (ed avente così un indirizzo IP italiano) ma che, facendo passare i propri dati tramite un *server* intermediario fisicamente localizzato al di fuori dell'Unione europea (e quindi con un indirizzo IP estero), darebbe l'impressione al gestore del social (ad esempio, Facebook) di trovarsi all'estero.

[46]Vi sono anche degli strumenti utili per risalire all'indirizzo IP originale, quali *applet Java* e *Flash*, ma essi non sono in grado di garantire sempre il risultato richiesto, dato che esistono diverse applicazioni (che è possibile usare in combinazione tra loro) in grado di contrastare le attività di questi strumenti (si vedano, ad esempio, *Privoxy* e *Squid*).

[47]M. Luo, J.T. Hancock, D.M. Markowitz, *Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues*, in "Communication Research", 2020, in cui gli autori hanno messo in luce come su Facebook contenuti con un elevato numero di "mi piace" vengano percepiti come maggiormente credibili.

[48]Cfr. prima sezione penale della Corte di cassazione, sentenza del 2 gennaio 2017 n. 50; quinta sezione penale della Corte di cassazione, sentenze del 23 gennaio 2017 n. 8482, del 1° febbraio 2017 n. 4873 e del 6 settembre 2018 n. 40083.

[49]Su quest'argomento si veda il principio espressa dalla quinta sezione penale della Corte di cassazione, sentenza del 22 settembre 2004 n. 47452, nella quale si afferma che «la reputazione di una persona che per taluni aspetti sia già stata compromessa può divenire oggetto di ulteriori illecite lesioni in quanto elementi diffamatori aggiunti possono comportare una maggior diminuzione della reputazione nella considerazione dei consociati».

[50]Con riferimento a questa tipologia di attività pare opportuno fare attenzione a quanto detto dai giudici della quinta sezione penale della Corte di cassazione nella loro sentenza del 29 gennaio 2016, n. 3981. In essa gli Ermellini fanno riferimento alla condotta di condivisione delle critiche ad una persona offesa che «potrebbe assumere in astratto rilevanza penale soltanto qualora potesse affermarsi che con il proprio messaggio l'imputato aveva consapevolmente rafforzato la volontà dei suoi interlocutori di diffamare». In questa particolare sentenza è necessario compiere una fondamentale distinzione lessicale, in quanto i giudici nell'utilizzare il termine "condividere" intendono riferirsi alla partecipazione alla discussione concordando sulla criticabilità del soggetto considerato ma «senza ricorrere alle espressioni offensive utilizzate da altri, né dimostrando di volerle amplificare attraverso il proprio comportamento», mentre nel gergo delle attività sui *social media* tale verbo indica la ripubblicazione delle medesime espressioni offensive, che di fatto ne determina una volontaria amplificazione.

[51]Cfr. quinta sezione penale della Corte di cassazione, sentenza del 12 dicembre 2017 n. 55418. La decisione presa dalla Corte era relativa alla questione se la pubblicazione sul proprio profilo Facebook di video inneggianti allo Stato Islamico e l'apposizione di "mi piace" ad altri integrassero o meno il delitto di istigazione a delinquere previsto dall'articolo 414 del codice penale.

[52]Questi algoritmi di analisi e confronto con comportamenti "standard" non sono le uniche contromisure adottate per combattere i *bot*: spicca l'esperimento compiuto da Instagram dal 2019 in alcuni paesi di non mostrare il numero di "*mi piace*" alle fotografie e il numero di visualizzazioni dei video pubblicati sulla piattaforma.

[53]O. Varol, E. Ferrara, C.A. Davis et al., *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, cit.

[54]E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, cit.; J. Zhang, R. Zhang, Y. Zhang, G. Yan, *The Rise of Social Botnets: Attacks and Countermeasures*, in "IEEE Transactions on Dependable and Secure Computing", vol. 15, 2018, n. 6, p. 1068-1082. Si veda a questo proposito il progetto *InstaPy*, volto, a detta dei creatori, ad «automatizzare le tue interazioni sui social media per "coltivare" Likes, Commenti e Followers su Instagram»; il codice sorgente è in licenza gratuita GPL-3.0.

[55]Vi sono comunque degli algoritmi che si sono dimostrati efficaci nell'individuare i profili gestiti da bot. Si vedano, ad esempio, N. Guenon des Mesnards, T. Zaman, *Detecting Influence Campaigns in Social Networks Using the Ising Model*, 2018, e descritto in D. Walsh, *A new method for rooting out social media bots*, 2018.

[56]Ad esempio, il contratto di licenza d'uso dei servizi di LinkedIn prevede, al punto 13 del paragrafo 8.2 ("Attività non consentite"), un esplicito divieto di «Utilizzare programmi bot o altri metodi automatizzati per accedere ai Servizi» ma i test effettuati con il codice di cui alla figura 1 (e con le sue successive versioni) per la redazione di questo articolo non hanno portato ad alcuna reazione da parte del gestore del medium.

[57]Le medesime considerazioni vengono esposte anche in R. Gorwa, D. Guilbeault, *op. cit.*

[58]Si veda *ex multis* G. Caldarelli, R. De Nicola, F. Del Vigna et al., *The role of bot squads in the political propaganda on Twitter*, in "Communication Physics", vol. 3, 2020, n. 81, 15 p.

[59]Cfr. ordinanza della Corte costituzionale del 26 giugno 2020, n. 132, in tema di portata della diffamazione operata tramite l'utilizzo di mezzi informatici.

[60]Cfr. la proposta di legge (*bill*), proposta dalla senatrice dello Stato della California D. Feinstein nel giugno del 2018.

[61]Che verrebbero introdotti nel Federal Election Campaign Act of 1971 con la nuova *Section 325* (a)(1)(A) e (B).

[62]Si vedano, ad esempio, L.H. Kahane, *Politicizing the Mask: Political, Economic and Demographic Factors Affecting Mask Wearing Behavior in the USA*, in "Eastern Economic Journal", vol. 47, 2021, p. 163-183; L. Aratani, *How did face masks become a political issue in America?*, in "The Guardian", June 29, 2020.

[63]*Sec.* 2 (3) del proposto *Bot Disclosure and Accountability Act of 2018*.

[64]La *Section 230* venne introdotta nel *Communications Decency Act* nel 1996 come diretta risposta alle decisioni assunte dai tribunali dello Stato di New York nei casi *Cubby, Inc. v. CompuServe, Inc.* e *Stratton Oakmont, Inc. v. Prodigy Services*, che avevano riconosciuto gli ISP responsabili per i contenuti diffamatori pubblicati dai loro utenti.

[65]B. Storm, *op. cit.*

[66]J. O'Brien, *Putting a Face to a (Screen) Name: The First Amendment Implications of Compelling ISPs to Reveal the Identities of Anonymous Internet Speakers in Online Defamation Cases*, in "Fordham Law Review", vol. 70, 2002, n. 6, p. 2745-2776; T.E. Lynch, *Good Samaritan or Defamation Defender? Amending the Communications Decency Act to Correct the Misnomer of Section 230 . . . Without Expanding ISP Liability*, in "Syracuse Science and Technology Law Reporter", vol. 19, 2008, 35p; B. Storm, *op. cit.*

[67]H.F. Kuehl, *op. cit.*

[68]Relativamente all'illiceità dei discorsi diffamatori di questo genere si veda, ad esempio, *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942), in cui la Corte Suprema degli Stati Uniti, richiamando anche la propria precedente decisone *Cantwell v. Connecticut*, 310 U.S. 296 (1940), ha dichiarato che «il ricorso a epiteti o abusi personali non è in alcun senso [una forma di] comunicazione di informazioni o opinioni salvaguardata dalla Costituzione, e la sua punizione come atto criminale non solleverebbe alcun dubbio».

[69]Si vedano, ad esempio, *McIntyre v. Ohio Elections Commission*, 514 U.S. 334 (1995); *Buckley v. American Constitutional Law Foundation*, 525 U.S. (1999); *Citizens United v. Federal Election Commission*, 130 S. Ct. 876 (2010).

[70]*McIntyre v. Ohio Elections Commission*, 514 U.S. 347 (1995). La Corte distrettuale della Virginia, tra le tante, ha successivamente esplicitato la natura "non assoluta" del diritto di esprimersi anonimamente, dichiarando anche in *In re Subpoena duces tecum to America Online, inc.*, No. 40570 (Va. Cir. Ct. Jan. 31, 2000) che «coloro che subiscono danni a causa di comunicazioni illecite o perseguibili su Internet dovrebbero essere in grado di cercare un risarcimento appropriato impedendo ai trasgressori di nascondersi dietro ad un scudo illusorio di presunti diritti del Primo Emendamento».

[71]Si veda, ad esempio, la decisione della Corte Suprema americana *Bivens v. Six Unknown Named Agents of the Federal Bureau of Narcotics*, 403 U.S. 388 (1971).

[72]*Fairfax Media Publications Pty Ltd v. Voller*, *Nationwide News Pty Limited v. Voller* e *Australian News Channel Pty Ltd v. Voller* [2021] HCA 27.

[73]Emanato in attuazione della direttiva 2000/31/CE, relativa a taluni aspetti giuridici dei servizi della società dell'informazione nel mercato interno, con particolare riferimento al commercio elettronico.

[74]Il comma 1 dell'articolo 16 del d.lgs. 70/2003 sancisce che «Nella prestazione di un servizio della società dell'informazione, consistente nella memorizzazione di informazioni fornite da un destinatario del servizio, il prestatore non è responsabile delle informazioni memorizzate a richiesta di un destinatario del servizio . . . ».

[75]Il comma 1 dell'articolo 17 del d.lgs. 70/2003 sancisce che «nella prestazione dei servizi di cui agli articoli 14, 15 e 16, il prestatore non è assoggettato ad un obbligo generale di sorveglianza sulle informazioni che trasmette o memorizza, né ad un obbligo generale di ricercare attivamente fatti o circostanze che indichino la presenza di attività illecite».

[76]Si vedano, ad esempio, Seconda Sezione del Tribunale Napoli Nord, sentenza del 4 novembre 2016 e Sezione Specializzata d'Impresa della Corte di appello di Milano, sentenza del 7 gennaio 2015, citate in G. Magri, *False Recensioni di Beni o Prodotti in Internet e Responsabilità Civile*, in "Revista Brasileira de Direito Civil", vol. 20, 2019, n. 2, pp. 113-135.

[77]Tale figura è stata delineata in Tribunale di Milano, sentenza del 07 giugno 2011, n. 7680.

[78]S. Sica, *Responsabilità del provider: per una soluzione "equilibrata" del problema*, in "Il Corriere Giuridico", 2013, n. 4, pp. 505-510; G. Magri, *op. cit.*; C. D'Urso, *I profili informatici nella valutazione della responsabilità dell'Hosting Provider*, in "Rivista Italiana di Informatica e Diritto", 2021, n. 1, pp. 79-88, in cui l'autore cita anche Tribunale di Roma, sentenza del 12 luglio 2019, n. 14757, nella causa R.G. n. 62326/2015, *RTI c. Dailymotion*.

[79]F. Manjoo, K. Roose, *How to Fix Facebook? We Asked 9 Experts*, in "The New York Times", 2017.

[80]Adottato in attuazione della direttiva UE n. 2015/849 e del Regolamento UE n. 2015/847.

[81]In particolare, la soluzione adottata dalla Corte Suprema dello Stato dell'Illinois – la cosiddetta *Rule* 224 – è da tempo oggetto di attenti studi e ritenuta da alcuni come una delle migliori. Si veda, ad esempio, B. Storm, *op. cit.*, che definisce quella della Corte dell'Illinois come «la migliore delle soluzioni adottate dalle corti americane».

[82]Si veda, ad esempio, la sentenza del caso *Ashton v. Kentucky*, 384 U.S. 195 (1966), nella quale i giudici americani ritennero che la maggior parte delle normative penali statali sulla diffamazione (*libel statutes*) fossero troppo vaghe per superare un vaglio di costituzionalità, finendo così col porre le basi per una loro invalidazione.

[83]A.J. Wagner, A.L. Fargo, *Criminal Libel in the Land of the First Amendment*, The international Press Institute (IPI), 2015, 39 p.

[84]47 U.S.C. § 551(c)(2)(B).

[85]T.E. Lynch, *op. cit.*; H.F. Kuehl, *op. cit.*

[86]Un'altra considerata molto valida è quella sviluppata dalla Corte Suprema dello Stato dell'Illinois, la cosiddetta *Rule* 224.

[87]*Doe v. Cahill*, 884 A.2d 451, 462-65 (Del. 2005).

[88]T.E. Lynch, *op. cit.*

[89]Si vedano *ex multis* C.B. Vincent, *Cybersmear II: Blogging and the Corporate Rematch Against John Doe Version 2.006*, in "Delaware Journal of Corporate Law", vol. 31, 2006, n. 3, p. 987-1009; *Best Western International v. Doe*, No. CV-06-1537-PHX-DGC, 2006 WL 2091695 (D. Ariz. 2006); *McMann v. Doe*, 460 F.Supp. 2d 259 (D. Mass. 2006); *In re Does* 1-10, 242 S.W.3d 805 (Tex. App. 2007).

[90]T.E. Lynch, *op. cit.*

[91]M.R. Allegri, *Alcune considerazioni sulla responsabilità degli intermediari digitali, e particolarmente dei social network provider, per i contenuti prodotti dagli utenti*, in "Informatica e Diritto", vol. 26, 2017, n. 1-2, pp. 69-112.

[92]*Sec.* 4 (c)(3) del proposto *Bot Disclosure and Accountability Act of 2018*.

[93]Si vedano, a titolo esemplificativo, N. Abokhodair, D. Yoo, D.W. McDonald, *op. cit.*; E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, cit.; E. Ferrara, *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election*, cit.; G. Caldarelli, R. De Nicola, F. Del Vigna et al., *op. cit.*

[94]Si veda, ad esempio, come il ben dichiarato modello di business di Facebook ponga al centro la registrazione delle preferenze degli utenti al fine di poter indirizzare loro degli annunci pubblicitari "personalizzati". La società stessa lo descrive con accuratezza in diverse proprie pagine web e lo stesso fondatore lo ha anche spiegato (con la famosa frase *Senator, we run ads*) di fronte ad una commissione del Senato degli Stati Uniti nel 2018. Si veda inoltre, sempre a titolo esemplificativo, quanto riportato in J. Koetsier, *Facebook Wants You To Want Personalized Ads. It's Not Going Well*, March 1, 2021.

[95]Si noti come in E. Ferrara, O. Varol, C. Davis et al., *The Rise of Social Bots*, cit., vengano presentati diversi sistemi per il rilevamento di *social bot* e che l'Osservatorio sui Social Media dell'Indiana University presenti sul proprio sito Internet diversi strumenti di rilevazione e contrasto dei bot e della disinformazione su Internet e sui social media.

[96]Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica, a norma dell'articolo 11 della legge 29 settembre 2000, n. 300.

[97]Sentenza del 12 luglio 2011, C-324/09, *L'Oréal SA e altri c. eBay International AG e altri*.

[98]Sentenza del 24 novembre 2011, C-70/10, *Scarlet Extended SA c. Société belge des auteurs, compositeurs et éditeurs SCRL (Sabam)*.

[99]Sentenza dell'11 settembre 2014, C291/13, *Sotiris Papasavvas c. O Fileleftheros Dimosia Etaireia Ltd e altri*, nella quale è stato ribadito che «Dal considerando 42 della direttiva 2000/31 risulta peraltro che le deroghe alla responsabilità previste da tale direttiva riguardano esclusivamente i casi in cui l'attività di prestatore di servizi della società dell'informazione sia di ordine meramente tecnico, automatico e passivo» e sono state citate le conclusioni sul precedente caso *L'Oréal SA c. eBay*.

[100]G. Magri, *op. cit.*

[101]Sulla sola piattaforma di sviluppo GitHub sono presenti oltre 16.000 diversi progetti (*repositories*) liberamente accessibili di algoritmi di analisi del linguaggio naturale; cfr. un elenco di tali progetti.

[102]B. Zarouali, T. Dobber, G. De Pauw, C. de Vreese, *Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media*, in "Communication Research", 2020.

[103]Si vedano *ex multis* C.D. Corley, D.J. Cook, A.R. Mikler, K.P. Singh, *Text and Structural Data Mining of Influenza Mentions in Web and Social Media*, in "International Journal of Environmental Research and Public Health", vol. 7, 2010, n. 2, p. 596-615; G. Paltoglou, M. Thelwall, *Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media*, in "ACM Transactions on Intelligent Systems

and Technology", vol. 3, 2012, n.4, 19 p.; A. Ceron, L. Curini, S.M. Iacus, *iSA: a fast, scalable and accurate algorithm for sentiment analysis of social media content*, in "Information Sciences", vol. 367-368, 2016, p. 105-124; B. Zarouali, T. Dobber, G. De Pauw, C. de Vreese, *op. cit.*

[104] D. Susser, B. Roessler, H. Nissenbaum, *Technology, autonomy, and manipulation*, in "Internet Policy Review", vol. 8, 2019, n. 2, p. 22 p.

[105] Si veda a questo riguardo anche D. Das, *Social Media Sentiment Analysis using Machine Learning – Part I*, Sept. 6, 2019 e Id., *Social Media Sentiment Analysis using Machine Learning – Part II*, Sept. 22, 2019, in cui l'autore svolge un'accurata descrizione (corredata anche di stralci del codice sorgente del programma utilizzato) di come sia possibile creare un algoritmo che compia un'analisi dei contenuti pubblicati da vari utenti su Twitter e di come grazie al *Machine Learning* sia possibile anche «rilevare il sentimento associato a un particolare tweet e categorizzarlo come negativo o positivo».

[106] Tribunale di Roma, sentenza del 07 gennaio 2015, n. 29, Corte di appello di Roma, sentenza del 29 aprile 2017, n. 2883, e diciassettesima sezione civile del Tribunale di Roma, sentenza del 10 gennaio 2019, n. 693. Quest'ultima, in particolare, ha operato un'attenta analisi della giurisprudenza dei giudici dell'Unione e di quella nazionale.

[107] *Trkulja v. Google Inc. LLC* [2012] VSC 533.

[108] M.R. Allegri, *op. cit.*

[109] L'implementazione di tale funzione è stata annunciata dalla società proprietaria del servizio il 16 agosto 2021; cfr. l'Annuncio in questione.

[110] H.S. Al-Khalifa, R.M. Al-Eidan, *An experimental system for measuring the credibility of news content in Twitter*, in "International Journal of Web Information Systems", vol. 7, 2011, n. 2, p. 130-151.

[111] A. Hern, *Twitter hides Donald Trump tweet for 'glorifying violence'*, in "The Guardian", May 29, 2020.

[112] A questo riguardo, uno studio condotto dalla Chilling Effects Clearinghouse (un progetto congiunto della Electronic Frontier Foundation e di diverse scuole di legge, tra cui Harvard, Stanford e Berkeley) ha analizzato le lettere di diffida inviate a Google ai sensi della *Section 512* del DMCA rivelando che oltre la metà degli avvisi sono stati inviati da aziende che prendevano di mira possibili concorrenti.

[113] Si veda a questo proposito sezione specializzata d'impresa della Corte di appello di Milano, sentenza del 7 gennaio 2015.

[114] Corte di cassazione, sentenza del 19 marzo 2019, n. 7708.

[115] Si veda a questo riguardo M. Stella, E. Ferrara, M. De Domenico, *Bots increase exposure to negative and inflammatory content in online social systems*, in "National Academy of Sciences of the United States of America", vol. 115, 2018, n. 49, p. 12435-12440; T. Benson, *Trolls and bots are flooding social media with disinformation encouraging states to end quarantine*, in "Insider", Apr. 24, 2020 e B. Strick, F. Syavira, *Papua unrest: Social media bots 'skewing the narrative'*, in "BBC News", Oct. 11, 2019.

* * *

Social media, artificial profiles, and reputation protection. How the advent of social bot for managing social media profiles can pose a serious threat to people's reputations and what might be the responses to this danger

Abstract: The use of social media has allowed anyone to spread their thoughts with a speed and a capillarity that eclipse those of the large traditional press companies. Within these "virtual squares" also operate the so-called "social bot", programs that, once provided with the login credentials of an account, are capable of managing it independently, but giving the impression of being a real person. Their rapidity and precision of reaction on social networks make these "digital tools" dangerously useful for the dissemination of insinuations, slanders and offenses with a breadth and speed that can irreparably destroy the reputation of anyone who is targeted by their administrators. The current legal system for the protection of victims of defamation, despite the recent jurisprudential rulings on the online dissemination of falsehoods and the extension of the use of remedies, such as preventive seizure to social networks, still risks being ineffective when confronting a well-concerted use of these "socialbot". Not even the countermeasures introduced by the owners of social media have been able to effectively stem this phenomenon. Thus, it is necessary to rethink the protection of reputation extending it to the prevention of improper use of IT tools that are now within everyone's reach. Therefore, various hypotheses are proposed in this article for solving the problem, including the extension to the owners of social media of the obligation of "customer due diligence" already existing for financial operators or the establishment of an obligation for social media owners to set up an internal reporting system with also a mark visible to the other users of the content deemed defamatory.

Keywords: Defamation – Social networks – Automation – Bot