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

ISSN 2704-7318
Fascicolo 1-2023
DOI: 10.32091/RIID0107
Articolo sottoposto a peer review

Il contrasto legislativo ai socialbot. Alcuni spunti per una riforma in Italia

Alessandro Tedeschi Toschi • Giampaolo Berni Ferretti

L'avvento di Internet, prima, e dei social media, poi, ha permesso uno scambio di contenuti con dei volumi ed una velocità in precedenza impensabili. All'interno di quelle che sono a tutti gli effetti delle "piazze digitali" operano anche i cosiddetti socialbot, 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 loro rapidità e precisione di reazione nei social network rendono questi agenti software pericolosamente utili per la diffusione di contenuti favorevoli od ostili. Il recente impiego dei socialbot all'interno del dibattito elettorale e politico di diverse nazioni e le sue conseguenze hanno posto l'urgenza che i legislatori statali intervengano contro questo fenomeno. Sull'esempio delle riforme adottate o proposte da diverse nazioni nel mondo, nonché della cornice d'azione predisposta dalle autorità europee, vengono presentati alcuni spunti per poter contrastare il cosiddetto "problema bot" declinato in questi particolari caratteri di propaganda politica.

Social network – Automazione – Bot – Propaganda – Disinformazione

1. I socialbot all'interno dei social media: natura e pericoli

Negli ultimi anni i social media sono diventati un terreno fertile per la propaganda ed il proselitismo di tutti coloro che ne hanno compreso le potenzialità e il funzionamento[1]. Ad attrarre l'attenzione verso questi strumenti di comunicazione hanno contribuito vari fattori: la facilità e la gratuità del loro utilizzo, il loro crescente impiego da parte delle persone come fonte di notizie[2], una forma di interazione non mediata ed istantanea con il pubblico che evita verifiche e contestualizzazioni da parte di professionisti e una architettura fisica consistente in una rete di server capaci di svolgere in autonomia miliardi di operazioni e di permettere la condivisione di contenuti senza controlli efficaci (gli algoritmi di verifica si sono dimostrati spesso incapaci di contrastare la diffusione delle cosiddette fake news)[3]. Insomma, i social media hanno portato a dei cambiamenti radicali nei modi e nei tempi di propagazione del proprio pensiero[4].

A. Tedeschi Toschi lavora nel settore privato. 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 il presidente dell'associazione culturale senza scopo di lucro "Milano Vapore".

Durante la campagna referendaria sulla permanenza della Gran Bretagna nell'Unione europea, il ruolo che queste piattaforme hanno avuto come catalizzatori e motori del dibattito all'interno della società è risultato particolarmente evidente[5]. A seguito dei controversi risultati di questa votazione è stato posto in evidenza come le discussioni e le interazioni che avvengono nei social network incidano sulle dinamiche sociali e perfino sulle votazioni politiche. È stato anche posto in luce come attività ben studiate ed organizzate di creazione e condivisione di contenuti sulle piattaforme siano in grado di creare l'illusione di un consenso su argomenti che in realtà non sono così popolari[6].

Il dibattito sviluppatosi online circa l'evento referendario d'oltremanica è stato anche il palcoscenico di un'estesa e strutturata operazione di propaganda sui social messa in atto da soggetti estranei al corpo elettorale britannico[7]: è stato riportato come oltre 150.000 account localizzati all'interno della Federazione Russa abbiano pubblicato su Twitter contenuti inerenti alla Brexit, diffondendo tra gli elettori britannici informazioni non veritiere e polarizzanti e dissimulandone la provenienza[8].

Anche la Francia ha subito attività di diffusione di contenuti falsi, fuorvianti o politicamente schierati compiute da soggetti ed organizzazioni riferibili alla Russia e ai suoi apparati di intelligence. Nel corso delle elezioni presidenziali del 2017 si è assistito ad una campagna ben organizzata ed estesa – soprannominata poi *Macron Leak* – delle cui principali operazioni sono stati accusati il Cremlino e il Servizio Informazioni delle forze armate russe (GRU)[9]. In questo contesto si è potuta notare anche una certa partecipazione di esponenti della *Alt-right* americana e della cosiddetta "Foreign legion" dell'altra candidata, Marine Le Pen[10].

Pure sull'altra sponda dell'Atlantico le attività di diffusione di contenti mistificatori e polarizzanti da parte di nazioni straniere sono state particolarmente intense. In merito alle elezioni presidenziali statunitensi del 2016 è stata rilevata un'intensa attività sui social media da parte di profili riconducibili a soggetti situati sul territorio russo. In particolare, è stato posto in evidenza come fin dal 2014 la Internet Research Agency (IRA), una compagnia privata strettamente legata al Cremlino[11], avesse il controllo di migliaia di profili su Twitter e di centinaia di account privati e pagine pubbliche su Facebook e come con essi sia stata in grado di raggiungere sui social, tramite la pubblicazione di contenuti, 126 milioni di cittadini americani[12].

Insomma, queste "piazze digitali" hanno consentito a soggetti in possesso dei capitali necessari e di un certo grado di conoscenza del loro funzionamento di raggiungere una platea di proporzioni altrimenti inarrivabili e di interferire nei processi democratici di altri paesi, ottenendo per sé o per le proprie nazioni vantaggi strategici, diplomatici e propagandistici[13].

Quella che a prima vista pare una schiera di fedelissimi seguaci profondamente convinti della validità delle opinioni espresse online può non essere effettivamente composta da esseri umani assiduamente presenti sulle piattaforme social. Può trattarsi, invece, dei cosiddetti "social bot", dei «programmi per computer progettati per utilizzare i social network simulando il modo in cui gli esseri umani comunicano e interagiscono insieme»[14]. Alcuni hanno anche coniato il termine "socialbot" (una parola unica), con cui viene indicata una sottocategoria dei social bot (due parole distinte) che comprende i software che, assumendo un'identità fasulla, si infiltrano in reti di utenti reali (ossia i network) e diffondono link e contenuti dannosi[15].

Nonostante il funzionamento dei social bot possa sembrare complesso, bisogna considerare come esso si fondi sulla ripetizione di azioni preimpostate dalla piattaforma stessa su cui operano (si pensi ai "Mi piace" e "Condividi" di Facebook, interazioni standardizzate e ripetitive per le quali esistono appositi e ben identificati pulsanti virtuali) e sulla raccolta dei contenuti caricati dagli utenti[16]. Risulta, quindi, semplice istruire questi programmi software ad individuare le azioni da compiere (i pulsanti virtuali da premere) e a compierle solo al verificarsi di determinate condizioni.

Bisogna anche tenere ben presente che tutti i bot (ampia categoria in cui rientrano social bot e socialbot) sono dei 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. Ciò significa che sullo sfondo di questa struttura digitale 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 in autonomia[17]. 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".

La comprensione delle modalità di creazione di un programma software autonomo in grado di agire su un social medium ci porta a dover prendere in considerazione anche il grado di coinvolgimento – e quindi di responsabilità – del loro creatore. A prima vista sembrerebbe ovvio attribuirgli la responsabilità di qualsiasi evento finisca col realizzarsi tramite l'impiego dei suoi prodotti[18], tuttavia, date alcune

peculiarità tecniche insite in qualsiasi tipo di programma software, l'attribuzione non può essere automatica. Infatti, sebbene sia il programmatore di un social bot a decidere quali azioni questo debba compiere, è ben possibile che non sia lui a indicare espressamente quali condizioni scatenino le azioni di questo agente software. Ciò deriva dal fatto che le sue attività su un social medium consistono nell'interazione con elementi esterni al suo codice sorgente e, quindi, esse possono essere le più disparate, dipendendo solamente dalle intenzioni del suo utilizzatore effettivo (le cui azioni ben possono esulare dal controllo del programmatore). La conseguenza è che sono coloro che impiegano attivamente i bot a dover essere ritenuti responsabili per le azioni compiute da questi strumenti digitali, in quanto sono loro ad orientarne le attività verso il raggiungimento di un determinato obiettivo. L'ulteriore conseguenza è che l'attribuzione di una qualche forma di responsabilità per gli eventi causati dai bot in capo al loro creatore debba avvenire solamente qualora egli li abbia effettivamente utilizzati oppure li abbia forniti al loro utilizzatore finale con la volontà di apportare un aiuto materiale alla realizzazione di un conosciuto progetto delittuoso (assumendo, così, una posizione assimilabile a quella dell'ausiliatore di un reato)[19].

Insomma, nel considerare i possibili effetti dell'utilizzo di un agente software bisogna tener ben presente che esso può essere strutturato secondo modi e forme che lo rendono un semplice strumento utile alla diffusione di qualsiasi tipo di contenuto. Di conseguenza, il loro uso per scopi illeciti deve essere valutato con particolare attenzione per poter giungere alla corretta attribuzione delle responsabilità per i reati compiuti con essi.

In merito ai social bot vi è anche da sottolineare come essi si giovino di diverse caratteristiche tipiche delle tecnologie digitali: la ripetibilità[20], la "leggerezza" per le capacità di calcolo dei computer, l'indifferenza per il luogo fisico dal quale sono connessi a Internet (che permette loro di rimanere al di fuori della giurisdizione delle nazioni oggetto delle campagne di disinformazione)[21], la possibilità di gestire qualsiasi profilo social (fattore estremamente utile nel caso in cui uno di questi venga sospeso o bandito), e la capacità di "comprendere" i contenuti pubblicati dagli utenti (tramite algoritmi di Natural Language Processing, detti NLP)[22]. La somma di tutte queste caratteristiche permette loro di essere impiegati per creare una schiera oceanica di fedelissimi seguaci artificiali che diano l'impressione di essere persone profondamente convinte della validità delle opinioni espresse. Questa particolare tecnologia di automazione delle interazioni sui social permette oggi di essere al centro dell'interesse di un vasto pubblico e di poterne orientare i "sentimenti" senza dover più ricorrere a schiere di dipendenti[23].

Il pericolo maggiore di questo processo è che i follower artificiali siano programmati per ripetere prontamente qualunque opinione, senza che questa sia effettivamente convincente, veritiera o anche solo in linea con i più basilari principi del rispetto della persona[24], e per dare ai meno esperti l'illusione che sia, invece, il grande pubblico a condividerla[25]. In altre parole, i socialbot possono essere usati – e sono già da tempo usati[26] – per distorcere le discussioni online su determinati temi, stimolare i seguaci di ideologie o personaggi pubblici e per generare false impressioni di popolarità[27]. Di questo fenomeno di accrescimento artefatto del sostegno hanno beneficiato anche personalità politiche attive sui social media, che si sono viste promuovere i propri contenuti pubblicati sulle piattaforme in modo molto esteso[28].

La diffusione artificialmente potenziata di contenuti al centro del dibattito può essere spiegata dal fatto che, come già mostrato da alcuni importanti studi[29], la formazione del pensiero politico degli individui è influenzata da più reti interpersonali di natura diversa, che essi tendono a prestare attenzione a ciò che sembra popolare e a fidarsi delle informazioni che circolano all'interno del loro contesto sociale[30], che essi spesso considerano come veritiere le opinioni che sono ampiamente diffuse[31] e che oggigiorno anche le cerchie virtuali di connessioni giocano un ruolo importante nella formazione delle convinzioni personali e delle opinioni politiche[32]. Così, l'uso dei socialbot è divenuto un valido strumento di sviamento del naturale orientamento delle discussioni online tra utenti ed un ancor più valido mezzo per indirizzare le persone attive sui social verso posizioni di nicchia, dando, però, l'illusione che queste siano popolari e condivisibili per via della loro intrinseca validità e non a causa di un'intensa e ben finanziata campagna propagandistica.

Di fatto, quindi, chi sappia creare e ben orchestrare queste folle di utenti artificiali ben potrebbe fingersi un semplice interprete dei sentimenti delle masse attive online, anche se in verità ne sta forzando gli orientamenti. Grazie ai socialbot, i soggetti in possesso delle giuste conoscenze possono diventare, per usare le parole di Ferrara e colleghi (2016)[33], dei burattinai che muovono a proprio piacimento i fili di migliaia di pupazzi digitali, liberi di orientarli nella direzione che più gli aggrada.

2. I tentativi di riforma legislativa nel mondo

Di fronte alle conseguenze delle influenze straniere esercitate sui social network, i governi di diverse na-

zioni hanno istituito commissioni di indagine per comprendere quali fossero state le dinamiche all'interno delle reti sociali online che avevano avuto un ascendente su questi risultati[34]. Le conclusioni a cui queste sono giunte sono state seguite dall'impegno di alcuni organi legislativi ad adottare delle riforme volte ad ostacolare le attività di disinformazione e di mistificazione avvenute sulle piattaforme digitali. Un breve spazio di intervento è stato dedicato anche al contrasto alle attività di diffusione potenziata di contenuti compiute dai socialbot con, tuttavia, un'incidenza che è stata messa in dubbio[35].

Tra le poche nazioni a reagire legislativamente alle attività di questi agenti software si sono distinti gli Stati Uniti d'America. Questi, però, nonostante la profondità degli effetti che la propaganda digitale aveva avuto sul loro tessuto sociale e politico, non sono ancora riusciti ad approvare una normativa nazionale che affronti il problema. Solo lo Stato della California è arrivato ad introdurre una normativa, il *Bolstering Online Transparency Act* del 2018 (S.B. 1001) (BOT Act)[36], che mira precisamente ad arginare l'influsso dei socialbot sulle dinamiche di interazione tra utenti e sulla formazione delle loro opinioni politiche. A livello federale, invece, sono state presentate diverse proposte di legge[37] senza che, però, sia stato possibile pervenire all'approvazione di un testo definitivo[38].

La piccola Repubblica di Singapore ha adottato un provvedimento di riforma al fine di contrastare la diffusione – e soprattutto la credibilità – di contenuti fuorvianti, mistificatori e falsi tramite Internet, i social media ed anche i socialbot[39]. La ferma risposta data dal legislatore della Città dei Leoni a questo nuovo fenomeno tramite l'approvazione del *Protection from Online Falsehoods and Manipulation Act* 2019 (POFMA)[40], però, è stata fortemente criticata per la durezza delle soluzioni adottate, la vaghezza di certe definizioni e per la facilità con cui essa potrebbe essere trasformata da strumento di tutela dell'opinione pubblica a mezzo di repressione del dissenso[41].

Nella Repubblica d'Irlanda, invece, è in discussione al *Dáil Éireann* (la camera bassa del Parlamento nazionale) l'*Online Advertising and Social Media (Transparency) Bill* 2017 (OASM Bill)[42]. Nel tentativo di edificare una propria difesa normativa, il legislatore dell'Isola di Smeraldo ha tratto evidente ispirazione dalla legge statale della California statunitense, cercando anche di superare le criticità in essa contenute. Al momento la proposta di legge, pur avendo il merito di aver riconosciuto la necessità di un intervento di tutela dei cittadini da indebite influenze sul processo di formazione delle loro opinioni, non è priva di rilevanti criticità[43].

3. Gli interventi comunitari contro la propaganda computazionale online

Di fronte alla sempre maggiore rilevanza dello spazio sociale digitale nell'evoluzione del dibattito politico e della tendenza all'estremizzazione, sono cresciuti anche all'interno delle istituzioni europee l'interesse e la preoccupazione per le attività di interferenza esterne e di proselitismo avvenute nei recenti processi elettorali degli Stati membri. L'Unione ha cominciato ad occuparsi del problema della diffusione su Internet di discorsi incitanti all'odio[44], giungendo anche ad approvare, di concerto con alcuni dei principali Social Network Provider, delle norme di autoregolamentazione che questi si sono impegnati a seguire per contrastare la diffusione di discorsi d'odio online (CCCDOO)[45].

In merito ai pericoli legati alla presenza dei socialbot, soprattutto per la conduzione di operazioni potenziate di propaganda computazionale, le istituzioni europee non hanno ancora elaborato dei provvedimenti specifici. Sono, comunque, rintracciabili diverse menzioni sparse all'interno di più documenti e comunicazioni relative al loro uso per amplificare artificialmente la diffusione di contenuti fuorvianti[46]. In particolare, è all'interno della comunicazione della Commissione sul contrasto alla disinformazione online che si trova il maggior numero di considerazioni circa il ruolo di amplificatori artificiali della diffusione di false informazioni giocato dai socialbot[47]. Identificato il potenziale di questi agenti software nel condurre operazioni di sviamento delle opinioni dei cittadini europei, era stato presentato anche l'impegno della Commissione alla formazione di un Codice di buone pratiche per le piattaforme online che comprendesse tra i propri obiettivi anche quello di «stabilire sistemi e norme chiari per i bot e fare in modo che la loro attività non possa essere confusa con l'interazione umana». Inoltre, sempre in un'ottica di limitazione degli effetti negativi delle campagne di propaganda computazionale, la comunicazione della Commissione aveva evidenziato come la maggior parte dei partecipanti alle consultazioni pubbliche relative al piano per l'istruzione digitale dei cittadini dell'Unione (adottato dalla Commissione nel gennaio del 2018)[48] ritenesse che i gestori delle piattaforme online dovessero adottare gli accorgimenti necessari ad «informare gli utenti quando i contenuti sono generati o diffusi da un bot».

Le istituzioni europee si sono brevemente espresse anche all'interno della recente proposta di regolamento relativo a un mercato unico dei servizi digitali (il cosiddetto *Digital Service Act* o DSA)[49] in merito ai pericoli rappresentati dalle attività dei socialbot. In particolare, al Considerando 57 della proposta vengono nominati tra i «rischi sistemici» l'uso di «bot e [...]

altri comportamenti automatizzati o parzialmente automatizzati che possono condurre alla rapida e ampia diffusione di informazioni». Questo perché essi possono essere impiegati per «la manipolazione intenzionale e spesso coordinata del servizio della piattaforma [social], con effetti prevedibili sulla salute pubblica, sul dibattito civico, sui processi elettorali, sulla sicurezza pubblica e sulla tutela dei minori». È da sottolineare come la presenza di agenti software sia vista come un rischio sistemico per l'Unione e non solo un pericolo secondario in quanto, come rilevato all'interno del precedente Considerando, «le piattaforme online di dimensioni molto grandi sono utilizzate in un modo che influenza fortemente la sicurezza online, la definizione del dibattito e dell'opinione pubblica nonché il commercio online»[50]. Inoltre, all'interno del Considerando 68 del DSA è stata anche espressamente sottolineata l'opportunità che i codici di condotta per le piattaforme online prendano in considerazione anche le «operazioni coordinate volte ad amplificare informazioni, compresa la disinformazione, come l'utilizzo di bot o account falsi per la creazione di informazioni false o fuorvianti», in quanto esse possono avere effetti negativi «sulla società e sulla democrazia, quali la disinformazione o le attività di manipolazione e abuso».

Sulla scorta delle considerazioni contenute all'interno della sua Comunicazione del 2018, la Commissione aveva anche provveduto nel corso del medesimo anno a redigere un Codice di buone pratiche per contrastare la disinformazione su Internet[51]. Successivamente, alla luce sia della sua Guida sul rafforzamento del Codice di condotta[52] che del DSA (che prevede espressamente che le autorità europee incoraggino l'elaborazione di codici di condotta), la Commissione ha provveduto a rafforzare il Codice di autoregolamentazione per i gestori delle piattaforme social online, chiamato oggi 2022 *Strengthened Code of Practice on Disinformation* (2022 SCPD)[53]. Esso contiene per i suoi sottoscrittori una serie di Impegni e di misure da adottare al fine di presentare ai loro utenti «informazioni trasparenti» in merito agli annunci che vedono sulle loro piattaforme. Tra gli impegni contenuti nel 2022 SCPD, il quattordicesimo include specificamente tra «comportamenti e pratiche manipolativi inammissibili» l'impiego di «bot» per amplificare artificialmente la diffusione o la percezione del sostegno pubblico a contenuti disinformativi. Al fine di dare attuazione concreta all'impegno di porre un limite a queste condotte, il codice comprende alcune misure che devono essere adottate. Queste consistono sostanzialmente nell'attribuzione ai sottoscrittori del 2022 SCPD dei compiti di istituire in autonomia delle politiche «in merito a comportamenti e pratiche manipolatori non consentiti tramite i propri servizi», di comunicarle e di spiegarle, di sviluppare metriche per stimare la penetrazione e l'impatto degli account falsi sugli utenti "autentici", di coordinarsi per redigere un resoconto esaustivo delle «tattiche, tecniche e procedure» (TTPs) impiegate per manipolare gli utenti e, infine, di collaborare a sviluppare elementi di base, obiettivi e parametri di riferimento comuni per le misure messe in atto per contrastare tali comportamenti e pratiche.

La presa di coscienza da parte delle autorità europee della pericolosità di pratiche di propaganda computazionale (tra le quali rientra l'impiego dei socialbot) ed i loro primi interventi di responsabilizzazione dei principali colossi del settore della fornitura di servizi di social networking sono particolarmente importanti. Questo perché forniscono un primo orientamento di indirizzo per l'identificazione di condotte che sono viste dagli organi dell'Unione come un pericolo per la tenuta degli ordinamenti democratici. Insomma, le disposizioni del 2022 SCPD e delle proposte attualmente discusse a livello comunitario sono un buon primo passo nella limitazione dell'uso di strumenti di disinformazione capaci di minacciare la tenuta degli ordinamenti democratici. Tuttavia, sembra opportuno presentare alcuni spunti utili per una regolamentazione nazionale maggiormente precisa e – soprattutto – che coinvolga con maggiore incidenza i principali Social Media Provider nella tutela del diritto dei cittadini a ricevere liberamente informazioni o idee senza influenze e condizionamenti (diritto che ben può essere minacciato in maniera surrettizia e indiretta grazie ad agenti software).

4. Spunti per una riforma della normativa italiana

Sebbene l'attenzione rivolta all'impiego dei socialbot per fini propagandistici all'interno del panorama italiano sia stata decisamente minore – ma non nulla[54] – rispetto a quella prestata nei contesti elettorali di Regno Unito e Stati Uniti, il loro utilizzo non può essere ignorato. Ciò perché anche all'interno delle reti italiane di utenti dei social media sono già attivi da molti anni questi agenti software autonomi (già nel 2012 era stata riferita una massiccia presenza di follower artificiali di politici italiani su Twitter e Facebook)[55].

Anche se la quantità di indagini che si concentrano specificamente sulla partecipazione di socialbot alle discussioni su temi politici italiani è molto limitata[56], la loro qualità li rende di notevole interesse. In particolare, le indagini compiute da Cresci e colleghi hanno evidenziato come già nel 2014 questi agenti software venissero impiegati da esponenti del mondo politico al fine di potenziare artificialmente la propria visibilità all'interno dei social network[57]. Questi autori hanno

ipotizzato che nel corso delle elezioni del 2014 per la carica a sindaco di Roma uno dei candidati si fosse rivolto ad una società di social media marketing per incrementare la propria celebrità online e che questa avesse impiegato quasi mille dei propri profili Twitter automatizzati per supportare e pubblicizzare le sue posizioni politiche[58]. In altre parole, la consultazione elettorale capitolina del 2014 è stata il palcoscenico di un'intensa attività ben studiata ed organizzata di accrescimento artefatto del sostegno mostrato online ad un candidato politico.

In ragione della realtà qui evidenziata vengono di seguito avanzate alcune riflessioni sugli ambiti che potrebbero essere oggetto di riforme per contrastare un fenomeno che ha già contagiato il contesto politico online italiano. Si tenga presente che le modifiche non dovrebbero limitarsi ad una semplice innovazione legislativa, dato che l'evoluzione delle tecnologie digitali e delle dinamiche sociali ad esse legate può essere di una tale rapidità da rendere inefficace, inopportuna o addirittura controproducente una semplice legge di contrasto ai socialbot. Più opportuno sarebbe un intervento a più livelli, ossia tramite non solo delle leggi ma anche dei regolamenti adottati da autorità amministrative indipendenti. Inoltre, non si può certo prescindere dalla considerazione dei progetti in corso di elaborazione a livello europeo ma non si può neanche limitarsi a ricopiarli senza ulteriori elaborazioni.

4.1. Una corretta definizione di social bot (due parole distinte) e socialbot (una parola unica)

Per poter giungere ad una riforma che sia in grado di rispondere alle sfide presentate dall'uso di strumenti di diffusione "potenziata" su Internet di contenuti disinformativi è necessario innanzitutto identificare correttamente l'oggetto stesso degli interventi. Infatti, non esiste al momento una definizione univoca di bot o di socialbot e ciò ha portato ad una confusione lessicale che ha avuto dei riflessi negativi sulla formazione di altre normative.

Per poter definire in modo esaustivo social bot e socialbot bisogna aver ben presente che essi fanno parte della più ampia categoria degli Internet bot (a sua volta una sottocategoria dei bot)[59]. Per quanto riguarda questa prima sottocategoria, riprendendo le più autorevoli dottrine sull'argomento[60], si può dire che essa include quei programmi software autonomi che si eseguono in continuazione e che sono in grado – senza alcun intervento umano diretto – di attivarsi al verificarsi di determinate condizioni e di individuare le azioni da compiere per portare a termine i compiti affidatigli, riconoscendo l'ambiente informatico online in cui operano e adattandosi ai cambiamenti di questo.

Nonostante il funzionamento degli Internet bot possa sembrare complicato, bisogna considerare come un sito Internet sia un insieme di pagine web tra loro collegate e salvate su un web server. L'attività di questa sottocategoria di bot può essere descritta come la loro interazione con gli elementi contenuti nelle pagine web, che sono ben marcati (i linguaggi in cui sono scritte queste pagine sono, infatti, chiamati "linguaggi di marcatura" o "linguaggi di markup") e, quindi, per essi facilmente riconoscibili. Di conseguenza, le interazioni tra un utente e un server, come il download di un particolare file, possono essere compiute facilmente anche da un Internet bot e in una frazione del tempo.

(A) Definizione di social bot (due parole distinte) – Tra gli Internet bot si può individuare una sottocategoria che comprende solamente quelli che sono progettati per operare sui social media e che vengono chiamati social bot. Una valida descrizione di questo insieme si ha nell'analisi delle social botnet compiuta da Abokhodair, Yoo, e McDonald, nella quale vengono fatti rientrare i «programmi per computer progettati per utilizzare i social network simulando il modo in cui gli esseri umani comunicano e interagiscono insieme»[61]. In altre parole, i social bot sono degli Internet bot progettati per agire all'interno dell'ambiente (informatico) di un sito Internet di un social medium riconoscendone gli elementi in esso inseriti (come i post ed i pulsanti virtuali con cui esprimere delle "reazioni" ai contenuti) ed interagendo autonomamente con essi con le medesime tempistiche e condotte tipiche dell'utente umano medio. Risulta così facile immaginare un social bot programmato per accedere al sito Internet di un social medium e apporre una "reazione" (come un "Mi piace" di Facebook) a tutti i contenuti che rispettano determinati parametri, ossia per individuare all'interno della piattaforma tutti gli elementi che hanno determinati attributi e per cliccare il pulsante virtuale che permette di mostrare l'interazione (con tutti i suoi significati di supporto) compiuta dall'account controllato[62].

(B) Definizione di socialbot (una parola unica) – Per la definizione di socialbot bisogna partire dalle parole di Abokhodair e colleghi e da quelle di Boshmaf e colleghi[63]. In merito alle parole di questi ultimi vi è da evidenziare come si preferisca in questa sede considerare solo la definizione inizialmente proposta di «programmi informatici che controllano gli account OSN [Online Social Network] e imitano gli utenti reali» e rigettare, invece, l'allargamento della definizione di questo termine avanzato successivamente e che recita che «un socialbot consiste di due principali componenti: un profilo su un OSN mirato (il volto) e

il software del socialbot (il cervello)»[64]. Pare opportuno – almeno a chi scrive – mantenere l'insieme delle definizioni proposte legato agli strumenti software di gestione e non allargarlo anche al profilo gestito. In altre parole, si preferisce proporre una nomenclatura con cui indicare i soli strumenti di controllo (i bot nelle loro varie declinazioni) e non anche gli oggetti controllati (i profili automatizzati tramite essi). Quindi, si propone di definire i socialbot (una parola unica) come appartenenti ad una sottocategoria dei social bot (due parole distinte) composta da programmi software autonomi progettati per agire su un social medium simulando il modo in cui un utente umano si comporta su di esso e che, dissimulando la propria natura ed identità, sono in grado di infiltrarsi nelle comunità online di utenti umani per diffondere contenuti ed interagire con essi. Sebbene si riconosca che l'opera di dissimulazione dell'essenza di cosa gestisca un account online venga compiuta anche attraverso l'uso di un "volto digitale", essa non può essere ricondotta a quest'unico elemento. Come ben evidenziato da Cresci e colleghi, l'opera di mimesi dei profili consiste non solo nel corredare i profili forniti ai socialbot di ritratti e descrizioni testuali credibili ma anche nell'assemblare il loro codice sorgente in modo da imitare le condotte medie tipiche di un essere umano. È quest'ultimo aspetto che risulta occupare una posizione di preminenza nelle procedure di determinazione della natura dell'amministratore di un account.

Inoltre si è visto come i contenuti diffusi sui social media possano essere sia favorevoli che ostili a determinate persone, eventi od opinioni (in particolare, è stato notato come i socialbot vengano impiegati per promuovere il loro "padrone" o "committente" e le loro opinioni e, all'opposto, per screditare persone od orientamenti a loro invisi)[65] ma, data la varietà e la creatività che può connotare i messaggi veicolati da questi agenti software, è meglio evitare di dare specifiche colorazioni o valutazioni.

Si riconosce di star qui proponendo una definizione dissonante rispetto a quella adottata in un'importante legislazione: il BOT Act della California statunitense. Nella normativa dello Stato d'Oro, infatti, il termine «bot» viene utilizzato per indicare «un account online automatizzato in cui tutte o sostanzialmente tutte le azioni o i post di tale account non sono il risultato [dell'azione] di una persona»[66] ma questa scelta – ed anche altre – non sono esenti da critiche[67].

4.2. Ambito di applicazione di una normativa di contrasto ai socialbot

Individuato l'oggetto di una riforma del nostro ordinamento, bisogna concentrarsi su coloro che dovrebbero essere destinatari di oneri e doveri inerenti al contrasto alle attività dei socialbot. Infatti, per attuare una strategia di argine degli effetti di distorsione delle naturali dinamiche di evoluzione dei dibattiti online causate da questi agenti software, è necessario individuare le categorie di soggetti direttamente o indirettamente implicati nelle campagne di propaganda computazionale.

(A) Titolari di profili su piattaforme di social networking – Innanzitutto, coloro verso cui indirizzare dei doveri sono i titolari di profili attivi su queste infrastrutture digitali. Sono, infatti, loro i soggetti che hanno la possibilità di affidare la gestione di uno o più account a questi agenti software, di dissimulare la loro natura tramite i più vari accorgimenti[68] e di programmarli per pubblicare in modo automatizzato contenuti, condividere quelli pubblicati da altri o apporvi reazioni. Insomma, i primi destinatari di disposizioni di limitazione dell'uso di socialbot dovrebbero essere coloro che potrebbero usarli – e già li usano – per portare avanti attività di propaganda e disinformazione.

(B) Social Media Provider – Non si può disconoscere la necessità di attribuire anche ai fornitori di servizi di social networking delle responsabilità ulteriori rispetto a quelle di cui attualmente sono gravati. In particolare, essi dovrebbero essere caricati di diversi obblighi di tutela delle discussioni pubbliche online da indebite interferenze. Questo perché i Social Media Provider (SMP) sono stati in grado di trarre ingenti profitti dalle attività di potenziamento artificiale della diffusione di contenuti – soprattutto quelli polarizzanti e sensazionalistici – e dalle interazioni che gli utenti hanno avuto con essi. Questi guadagni non incentivano (almeno sul breve periodo) a contrastare in autonomia queste forme "inautentiche" di propagazione di contenuti. Anzi, essi sono una fonte di guadagno aggiuntiva e alcuni dei principali gestori sono già stati accusati in passato di aver anteposto i profitti al benessere dei loro utenti[69]. La questione economica fornisce una valida giustificazione all'attribuzione in capo a queste imprese del compito di vigilare sull'evoluzione delle interazioni che avvengono tra i loro utenti e di proteggerle. Infatti, i principali SMP riportano da anni fatturati e proventi che ben sarebbero in grado di sostenere ulteriori strutture interne di gestione e regolazione delle loro piattaforme[70].

In merito all'individuazione precisa dei soggetti a cui attribuire queste responsabilità, la soluzione più adeguata sembra essere quella adottata dal legislatore della California statunitense, che ha previsto delle soglie minime di accessi medi annuali da parte degli utenti per determinare quali amministratori di servizi

online siano soggetti alle disposizioni del BOT Act. Questo indice di "popolarità" minimo è stato ripreso anche dalla OASM Bill irlandese. Anche la Commissione europea ha adottato all'interno del DSA[71] una «soglia operativa riguardante i prestatori di servizi che rientrano nell'ambito di applicazione» di alcuni obblighi «in materia di dovere di diligenza per taluni servizi intermediari». Nello specifico, in tale proposta viene delineata la figura delle «piattaforme online di dimensioni molto grandi» (in inglese *Very Large Online Platforms* o VLOPs), ossia quei servizi della società dell'informazione di hosting di contenuti che hanno «un ampio raggio d'azione nell'Unione, stimato attualmente a oltre 45 milioni di destinatari dei servizi»[72].

Anche una normativa italiana dovrebbe presentare dei limiti di attribuzione di responsabilità a dei soggetti di grandi dimensioni, ossia dovrebbe sancire degli obblighi solamente ai SMP le cui piattaforme abbiano un certo numero medio di utenti o di visitatori unici in un determinato lasso di tempo. La determinazione dell'esatto numero dovrebbe, però, essere indicata all'interno di un decreto o di un regolamento, in quanto essi sono degli strumenti normativi che meglio rispondono alle necessità di rapido aggiornamento tipiche della regolamentazione delle tecnologie digitali. Un numero preciso di utenti che debba essere considerato valido per qualsiasi tipologia di sito web (incluso quello di un social medium) non terrebbe conto delle loro possibili peculiarità ed evoluzioni e rischierebbe, così, di creare degli squilibri nell'attuazione della tutela degli utenti e delle incongruenze nei trattamenti dei diversi SMP. Molto più opportuno sarebbe la definizione di diverse classi di "Piattaforme Online" e l'indicazione per ciascuna di esse di limiti di accessi personalizzati.

Il confinamento dell'ambito di applicazione della riforma solo a SMP di grandi dimensioni dipende dalla considerazione del fatto che a qualsiasi attribuzione di maggiori responsabilità di tutela – e, quindi, di maggiori obblighi di controllo – corrisponde l'adozione di strumenti e procedure onerosi che solo alcuni operatori dei settori informatici sarebbero capaci di adottare[73]. Si riconosce che una previsione di questo genere possa diminuire l'argine alla disinformazione online ma in questo contesto permane anche la necessità di non di rafforzare ulteriormente la struttura oligopolistica del mercato dei servizi per il social networking.

4.3. Un obbligo di vigilanza delle attività sui social media

Il primo e più importante ambito di responsabilizzazione dei principali fornitori di servizi di social networking è quello della vigilanza su ciò che avviene sulle piattaforme. Questa attività dovrebbe avvenire seguendo alcune direttrici.

(A) Adozione di strumenti di rilevazione della presenza di socialbot – L'obbligo di vigilanza dovrebbe innanzitutto tradursi nell'imposizione dell'adozione di strumenti di rilevazione della presenza di socialbot sui social media. Questa strada era stata intrapresa dal *Bot Disclosure and Accountability Act* 2018 (BDA Act, detto anche *Feinstein Bill*), alla cui Sezione 4(b)(3) era stato disposto che un SMP dovesse adottare «un processo per identificare, valutare e verificare se l'attività di qualsiasi utente del sito web di un social medium è condotta da un programma o un processo software automatizzato volto a impersonare o replicare l'attività umana online». Similmente, anche il 2022 SCPD europeo prevede che i suoi firmatari adottino misure di monitoraggio e rilevamento delle TTPs di manipolazione sulle piattaforme social (tra le quali è espressamente incluso l'impiego di «bot» per amplificare artificialmente la diffusione o la percezione del sostegno pubblico a contenuti disinformativi)[74]. Anche il DSA contiene la considerazione che, «per attenuare con diligenza i rischi sistemici» legati alla popolarità e all'utilizzo delle piattaforme online, i VLOPs dovrebbero attuare misure per il «rafforzamento dei processi interni o della vigilanza sulle loro attività, in particolare per quanto riguarda il rilevamento dei rischi sistemici»[75]. Sulla scorta di questa valutazione, il progetto europeo imporrebbe ai gestori delle piattaforme più grandi di individuare, analizzare e valutare – almeno una volta all'anno – gli «eventuali rischi sistemici significativi derivanti dal funzionamento e dall'uso dei loro servizi», tra i quali «la manipolazione intenzionale del servizio, anche mediante un uso non autentico o uno sfruttamento automatizzato del servizio»[76]. Inoltre, pure il nostro legislatore nazionale, nel più ampio ambito della prevenzione della manipolazione dell'informazione online e della garanzia della trasparenza sul web, aveva provato ad introdurre in capo ai «gestori delle piattaforme informatiche» un dovere di «costante monitoraggio» di quanto accade sui loro social media[77].

Una via nazionale di contrasto a questi agenti software dovrebbe, quindi, essere lastricata da una legge che imponga ai principali Social Media Provider (pSMP), ossia coloro le cui piattaforme raggiungano determinati numeri di visitatori, di adottare delle procedure e degli strumenti idonei a verificare se l'attività di qualsiasi profilo attivo sui loro media sia eseguita da un socialbot. Una previsione di questo genere dovrebbe anche essere accompagnata da un'armonizzazione con le disposizioni del decreto legislativo 9 aprile 2003, n. 70, che prevedono l'esclusione di alcuni

obblighi e responsabilità in capo ai prestatori di un servizio della società dell'informazione (categoria all'interno della quale facilmente si possono far rientrare i pSMP). Tale armonizzazione trarrebbe ispirazione dal rapporto intercorrente tra l'esenzione da «obblighi generali di sorveglianza» e l'attribuzione di «obblighi supplementari» in capo ai soli soggetti che soddisfano determinati requisiti previsti dal DSA. In altre parole, riprendendo il lessico del DSA, bisognerebbe prevedere che solo i pSMP debbano sottostare ad un obbligo di individuare, analizzare e valutare i rischi sistemici significativi derivanti dall'uso del loro servizio.

Ulteriori specificazioni in merito a quali procedure e strumenti dovrebbero adottare i pSMP sarebbero necessariamente da delegare a fonti secondarie del diritto. Questo perché la loro maggiore rapidità di formazione e di modifica le rende più idonee a regolare tecnologie in rapida evoluzione. Inoltre, dato anche che ciascun social medium può avere caratteristiche estremamente particolari, è più opportuno che una disposizione di legge sia ampia e generica e che siano decreti o regolamenti a identificare le varie sottocategorie e le relative disposizioni speciali per ciascuna di esse.

Bisogna anche riconoscere che, come ben messo in evidenza dalla letteratura sull'argomento[78], gli agenti software subiscono un affinamento ininterrotto delle loro routine interne di mimica dei comportamenti umani[79]. La più evidente criticità è che, allo stato dell'arte, l'efficacia degli strumenti di rilevazione utilizzati dai ricercatori dipende in gran parte dal grado di sofisticatezza dei socialbot analizzati: quelli maggiormente capaci di imitare le azioni compiute mediamente dagli umani sono in grado di passare il vaglio di questi mezzi di controllo[80]. È necessario, quindi, che gli obblighi per i pSMP non si concretizzino in elenchi di procedure, strumenti o indici da utilizzare per rilevare la presenza di profili gestiti da programmi software autonomi[81].

(B) Cooperazione, coordinamento e informazione – Come appena detto, l'evoluzione dei questi agenti software è di una tale intensità che impegni costanti ma separati di ciascun singolo gestore di piattaforme social potrebbero non riuscire a tenere il passo con l'innovazione degli strumenti che verrebbe loro chiesto di scovare. Di conseguenza, si dovrebbe riprendere lo spirito del 2022 SCPD, in particolare quello dell'impegno 14 e delle successive misure di attuazione, che impongono di «concordare e pubblicare un elenco e una terminologia delle TTPs impiegate dagli attori malintenzionati, che dovrebbe essere aggiornato su base annuale e consistere in una intesa comune sui comportamenti e sulle pratiche manipolative non consentite sui loro servizi [che sia] sempre aggiornata». L'imposizione di una collaborazione tra pSMP avrebbe anche il vantaggio di spingere ad una condivisione dei fenomeni rilevati e delle soluzioni, favorendo una maggior preparazione nei confronti di TTPs non ancora sperimentate da tutti.

Dovrebbe anche essere attribuito ai pSMP l'onere di presentare regolarmente rapporti sulle politiche da loro adottare per adempiere agli obblighi di legge e sulle ragioni della loro adozione (come stabilito anche dalle misure 14.1 – 14.3 del 2022 SCPD), indirizzandoli specificamente alle autorità competenti alla tutela del pluralismo e della libertà d'informazione dei cittadini (che nel nostro contesto sarebbe l'Autorità per le Garanzie nelle Comunicazioni).

4.4. Un obbligo di predisporre strumenti di segnalazione di sospette attività di socialbot

Oltre all'obbligo di rilevazione dell'impiego di socialbot, dovrebbe anche essere imposto ai SMP di permettere agli utenti di contribuire alla tutela del dibattito pubblico online. Dopotutto, il nostro legislatore ha già cercato di introdurre una forma di coinvolgimento degli utenti nelle attività di monitoraggio che si vorrebbero imporre ai gestori delle piattaforme social[82]. Similmente, il proposto DSA prevede l'obbligo per alcuni prestatori di servizi digitali di predisporre «meccanismi per consentire a qualsiasi persona o ente di notificare» a loro la sussistenza di situazioni contrarie alle normative[83]. In particolare, la proposta della Commissione europea evidenzia come sia importante che «tutti i prestatori di servizi di hosting, indipendentemente dalle loro dimensioni, predispongano meccanismi di notifica»[84].

(A) Procedure e strumenti di segnalazione – Un coinvolgimento del pubblico nell'identificazione di profili gestiti da socialbot dovrebbe essere attentamente ponderato e non essere visto come uno strumento risolutivo, dato che vi è il rischio che tali strumenti vengano utilizzati in modo arbitrario o abusivo. Si deve, quindi, riconoscere anche la necessità che processi di segnalazione della presenza di agenti software autonomi sui social siano calibrati con attenzione[85]. Infatti, tanto maggiori sono la facilità e la velocità con cui è possibile per gli utenti segnalare la sospetta gestione di un profilo da parte di un socialbot, tanto più facilmente questa procedura potrebbe essere più nociva che utile: dato che tutti i siti Internet dei social media sono composti da elementi digitali preimpostati e standardizzati (che qualsiasi programma software autonomo potrebbe essere addestrato a riconoscere e utilizzare), si potrebbe arrivare alla paradossale situazione in cui uno o più socialbot potrebbero segnalare in modo coordinato il profilo di una persona reale che

lo usa in modo legittimo ma che sia invisa al gestore di una squadra di questi agenti software. Inoltre, è sempre possibile che uno o più utenti sfruttino questa procedura per far sospendere il profilo legittimamente usato da un'altra persona.

La normativa suggerita dovrebbe anche prevedere l'introduzione di adeguate misure e limitazioni provvisorie alle attività del profilo segnalato[86]. In particolare, potrebbero essere introdotte alcune soluzioni "contenitive" del profilo, quali la sua demarcazione, la limitazione delle interazioni che si possono compiere con esso oppure la sua sospensione.

La demarcazione del profilo e dei contenuti pubblicati costituirebbe una forma più blanda di limitazione delle sue attività ma renderebbe, comunque, immediatamente più consapevoli gli utenti che venissero in contatto con esso della possibilità di essere vittime di campagne dissimulate di disinformazione o propaganda. Per avere efficacia, detta demarcazione dovrebbe essere chiara e cospicua – per riprendere la terminologia utilizzata nel BDA Act e nel BOT Act statunitensi e nel POFMA singaporiano – ed apparire non solo nello "spazio personale" online del profilo ma anche all'interno di ciascun contenuto da esso pubblicato. A tutto ciò aggiungendo anche delle limitazioni ai conteggi delle interazioni che questo profilo continuerebbe a compiere una volta marcato (ad esempio, il suo "Mi piace" ad un contenuto su Facebook non dovrebbe essere sommato nei conteggi mostrati ad altri utenti).

La limitazione delle interazioni del profilo social sulla piattaforma (che potrebbe essere sommata alla demarcazione), finirebbe col rendere l'account segnalato un semplice strumento di consultazione passivo di un archivio online e potrebbe sembrare una misura eccessiva. Tuttavia, impedirebbe proprio uno dei principali compiti demandati ai socialbot, ossia di incrementare artificialmente la popolarità mostrata di determinati contenuti attraverso le interazioni con essi.

(B) Procedura di difesa del profilo segnalato – Non si può ignorare la possibilità che la procedura di segnalazione venga utilizzata in modo abusivo e contro soggetti che usano in modo lecito i propri profili social. Di conseguenza, dovrebbe essere prevista anche una procedura di reclamo contro la segnalazione (e le misure di limitazione) attivabile dal titolare del profilo indicato. A questo proposito il primo esempio a cui potersi riferire è la Sezione 4 (c)(6) del BDA Act, il quale avrebbe imposto a ciascun Social Media Provider di istituire ed implementare «un processo che dia a un utente umano del sito web del social medium l'opportunità di dimostrare che l'attività online dell'utente è conforme alla policy» adottata dalla piattaforma in merito all'automazione dei profili. Un secondo esempio è quello del DSA, il quale al proprio articolo 17 impone che «le piattaforme online forniscono ai destinatari del servizio [...] l'accesso a un sistema interno di gestione dei reclami efficace, che consenta di presentare per via elettronica e gratuitamente reclami contro le [...] decisioni adottate dalla piattaforma online» e che detti sistemi siano «di facile accesso e uso e affinché consentano e agevolino la presentazione di reclami sufficientemente precisi e adeguatamente motivati». Inoltre, dovrebbe essere imposto ai pSMP di informare prontamente, in modo chiaro e con gli strumenti più opportuni il destinatario della segnalazione[87].

Ulteriori specificazioni in merito agli obiettivi e ai criteri che i passaggi di tale procedimento dovrebbero rispettare andrebbero inserite all'interno di fonti di rango secondario. Ad ogni modo, la prima di tali fasi dovrebbe consistere in una comunicazione chiara, completa e facilmente comprensibile al titolare del profilo segnalato. La comunicazione dovrebbe anche contenere un link che permetta di avviare la procedura di "difesa" del profilo e di spiegazione della genuinità delle proprie condotte[88].

Tra le fasi del procedimento di reclamo dovrebbe esserne prevista anche una di controllo della natura del soggetto che ha attivato la procedura (questo perché una procedura standardizzata si esporrebbe irrimediabilmente al rischio di essere svolta da bot appositamente creati). Inoltre, in un momento successivo al controllo, dovrebbe esserci uno spazio in cui il reclamante debba spiegare – per usare le parole della Commissione europea – in modo preciso ed adeguatamente motivato che le attività del suo profilo sono state compiute da un essere umano. Sebbene tale dimostrazione possa apparire come una *probatio diabolica*, si deve tenere conto che l'aspetto rilevante di questa fase del procedimento non è dare una argomentazione particolarmente convincente, quanto piuttosto che sia un essere umano a fornirla. Ciò perché, anche se una strutturazione in tale senso della procedura risulta piuttosto fragile, ad esempio perché vulnerabile a strategie di utilizzo ibrido dell'account (che in questi casi viene chiamato *account cyborg*)[89], essa permetterebbe comunque di costituire un impedimento al prosperare delle botnet (dette anche *bot farm*), ossia dei sistemi in cui un amministratore umano (detto *botherder*) si limita ad orientare le attività di vari socialbot tramite un programma di direzione centrale (detto *botmaster*) e non istruisce direttamente a ciascuno di essi. Le attività di reclamo per ciascuno dei profili segnalati risulterebbero per il *botherder* molto dispendiose in termini di tempo e potrebbero non essere notate per un certo periodo. In aggiunta, la previsione di una procedura di reclamo

contro i provvedimenti intrapresi dai pSMP è anche uno strumento di indispensabile garanzia degli utenti da processi decisionali altrimenti troppo opachi e da possibili errori degli strumenti e dei modelli di organizzazione e controllo implementati, nonché di tutela dai possibili abusi di queste procedure.

(C) Revisione della segnalazione da parte di un incaricato – A conclusione della procedura, dovrebbe essere imposto ai pSMP di disporre un controllo di revisione da parte di un proprio incaricato (umano) sia del profilo segnalato sia della validità della segnalazione stessa[90]. La procedura di revisione da parte di un soggetto incaricato di questa sorta di contenzioso dovrebbe avvenire solo nel caso di reclamo da parte del titolare del profilo stesso. Ciò deriva da ragioni di opportunità e di contenimento dei costi, nonché di carattere logico: se il soggetto a cui viene notificata la segnalazione non si adopera per contestarne la veridicità, vuol dire che non ha interesse a intraprendere tale procedura oppure che è un agente software incapace di portarla a termine. Infine, la legge dovrebbe disporre che il gestore della piattaforma social preveda anche delle forme di limitazione dell'utilizzo di questo sistema di segnalazione al fine di evitare che esso venga abusato per ostacolare le attività di un profilo in realtà legittimo[91].

A ulteriore tutela da abusi, la riforma qui suggerita dovrebbe prevedere anche delle "categorie protette" di profili per certe figure professionali come quella del giornalista. Questo perché, altrimenti, ci sarebbe il concreto rischio che professionisti qualificati, qualora pubblichino dei contenuti inerenti a personaggi pubblici o a questioni controverse, finiscano con l'essere immediatamente segnalati dai loro detrattori. In altre parole, in assenza di restrizioni all'utilizzo della procedura di segnalazione dei profili contro persone che rivestono ruoli rilevanti in una società democratica, questo iter potrebbe finire con l'ostacolare il diritto di cronaca e con lo spostare il fulcro del cosiddetto "problema bot" dalla condivisione di contenuti illeciti all'ostruzionismo contro contenuti legittimi.

Ad ogni modo, l'individuazione dei parametri di valutazione della natura di cosa stia utilizzando l'account segnalato dovrebbe essere compiuta con estrema cautela, dal momento che già oggi l'efficacia degli strumenti di rilevazione sviluppati ed utilizzati dai ricercatori in questo campo non è ottimale e dipende in gran parte dal grado di sofisticatezza dei socialbot analizzati. Alcuni studi hanno esplicitato quali caratteri delle attività di un account dovrebbero essere presi in considerazione per svolgere una valutazione che possa avere un certo grado di efficacia[92] ma non è possibile cristallizzare un elenco ben definito di parametri. Risulta, quindi, particolarmente importante che una normativa non indichi in forma cogente ed esclusiva quali strumenti o indici debbano essere utilizzati.

4.5. Un obbligo di adozione di politiche attive di contrasto alle attività dei socialbot

Al fine di contrastare efficacemente le attività compiute dai socialbot, è opportuno imporre ai pSMP non solo di monitorarli ma anche di agire attivamente contro il loro impiego per la diffusione di disinformazione. La necessità di attribuire ai gestori delle piattaforme un obbligo di combattere direttamente la presenza e le opere dei socialbot discende da ragioni già accennate. La principale è il fatto che essi sono stati in grado di trarre ingenti profitti dalle attività di potenziamento artificiale della diffusione di contenuti, soprattutto quelli polarizzanti e approssimativi.

L'urgenza di ostacolare le attività di questi agenti software ha già portato il legislatore federale statunitense a proporre l'attribuzione in capo ai SMP di un obbligo di intervento. Infatti, alla Sezione 4(c)(4) del BDA Act era stato proposto che ciascuno di essi dovesse istituire ed implementare «un processo in base al quale il social media provider adotterà ragionevoli azioni preventive e correttive per mitigare i tentativi di un utente di utilizzare un programma o un processo software automatizzato volto a impersonare o replicare un'attività umana online». La Sezione 4(c)(5) aveva anche previsto che ciascun SMP dovesse istituire ed implementare «un processo in base al quale il social media provider rimuoverà post, immagini o ogni altra attività online di un utente o di un profilo che facciano uso di un programma o un processo software automatizzato volto a impersonare o replicare un'attività umana online».

Nel contesto europeo il coinvolgimento attivo dei VLOPs nel contrasto ai socialbot trova – in termini più generici – una propria istituzione in più di una disposizione. Infatti, l'articolo 27 del DSA dispone che «le piattaforme online di dimensioni molto grandi adottano misure di attenuazione ragionevoli, proporzionate ed efficaci, adattate ai rischi sistemici specifici individuati» e il Considerando 57 include questi agenti software autonomi tra i rischi sistemici insiti nell'uso dei social media. In particolare, l'articolo 27 del DSA prevede che i VLOPs adeguino i propri «sistemi di moderazione dei contenuti o di raccomandazione», ossia che rendano efficaci ed adeguate le misure che incidono sulla capacità dei destinatari dei servizi di hosting forniti dalle piattaforme di grandi dimensioni di operare su di esse (potendo così anche cessare o sospendere l'account di questi destinatari)[93]. Il 2022 SCPD, poi, prevede che i suoi sottoscrittori implementino «politiche chiare in merito a comportamenti e pratiche

manipolatori non consentiti sui loro servizi»[94]. In altre parole, questo codice di autoregolamentazione impone ai SMP che vi abbiano aderito di intervenire attivamente contro l'uso dei socialbot finalizzato ad influenzare indebitamente e surrettiziamente la formazione delle opinioni degli utenti.

Anche il disegno di legge presentato in Italia al Senato per prevenire la manipolazione dell'informazione online avrebbe introdotto in capo ai gestori delle piattaforme social – assieme all'obbligo di monitorare costantemente le attività compiute su di esse «con particolare riguardo ai contenuti verso i quali gli utenti manifestano un'attenzione diffusa e improvvisa»[95] – la responsabilità di intervenire prontamente per sanzionare tutte le condotte compiute in violazione delle sue previsioni[96]. Sebbene questa proposta di riforma avesse come obiettivo quello di impedire la diffusione di contenuti disinformativi e non solo quello di contrastare l'impiego di socialbot, si può notare come il nostro legislatore abbia già avvertito la necessità di coinvolgere maggiormente i SMP nell'impedire che i loro servizi vengano sfruttati al fine di distorcere il dibattito online.

L'imposizione di agire attivamente per contrastare i socialbot trova una propria giustificazione anche nelle attente considerazioni e negli approfonditi studi già menzionati in precedenza[97]. Secondo questi, infatti, gli agenti software malevoli vengono largamente impiegati per generare false impressioni di popolarità di determinate opinioni – spesso estremiste – finendo con l'esacerbare la conflittualità nelle discussioni online e con l'avere riflessi profondi anche sulle dinamiche della società offline.

Quindi, in termini equiparabili a quanto proposto al Campidoglio statunitense e a Palazzo Berlaymont, anche una normativa italiana dovrebbe imporre ai pSMP di adoperarsi attivamente per contrastare la diffusione potenziata di contenuti e l'incremento artificiale dei loro parametri di popolarità (primo fra tutti il numero di reazioni positive o di supporto) sulle piattaforme che gestiscono. Ciò anche per dare migliore attuazione alle disposizioni del proposto regolamento europeo di adozione di «misure di attenuazione» delle attività di diffusione potenziata di contenuti compiute dai socialbot. Nello specifico, queste misure di attenuazione degli effetti delle azioni degli agenti software autonomi dovrebbero imporre la cancellazione di tutti i contenuti prodotti dal profilo incriminato (come post, status o commenti) e anche tutte quelle attività di ri-condivisione dei contenuti di terzi tramite detto profilo (come le attività di retweet su Twitter) e di apposizione di reazioni di approvazione dei contenuti di terzi (come il "Mi piace" di Facebook).

Una previsione di questo genere potrebbe apparire eccessiva, in quanto con essa non viene dato alcun risalto a quali contenuti siano stati effettivamente diffusi in modo potenziato tra gli utenti, tuttavia essa evita il problema di determinare esattamente quali azioni del profilo considerato siano state effettivamente compiute da un socialbot e quali, invece, da un utente umano che ne abbia ripreso anche solo provvisoriamente il controllo. Inoltre, dando rilievo solamente agli strumenti utilizzati per la gestione del profilo viene superata un'eventuale grave criticità insita in una normativa relativa a questo tema: quella della valutazione dei contenuti pubblicati o supportati tramite i socialbot. Infatti, andando ad eliminare dei contenuti sulla base delle modalità con cui sono stati diffusi e non della validità, dell'oggettività o dell'accuratezza degli stessi, si evita di sottoporre la questione a valutazioni che implicano una certa discrezionalità e, quindi, passibili di accuse di arbitrarietà o partigianeria. Per usare la terminologia propria del diritto statunitense, la normativa auspicata dovrebbe avere, sì, un contegno molto rigido ma, comunque, essere «content-neutral»[98].

Tale considerazione permette di includere pure l'ipotesi in cui il titolare umano del profilo ne affidi la gestione a un socialbot ma che questo venga immediatamente identificato dagli strumenti e dai modelli di controllo predisposti dal pSMP. Infatti, in questo modo si prescinde dalla rilevanza della causazione di un danno nella pratica impossibile da dimostrare (come può essere quello di aver effettivamente causato un'indebita modificazione delle opinioni di una persona tramite l'impiego di uno o più socialbot che le abbiano dato la falsa impressione che determinate opinioni siano più condivise che nella realtà), per ammantare, invece, di antigiuridicità una mera condotta rilevabile anche attraverso sistemi automatici di controllo. Di conseguenza, l'obbligo per un pSMP di intervenire per limitare o sospendere le attività di un profilo social e per limitare la diffusione o cancellare i contenuti da questo pubblicati dipenderebbe da una mera condotta realizzata dai suoi utenti, liberandolo dalla necessità di svolgere qualsiasi forma di valutazione degli effetti delle attività di diffusione e supporto compiuta da tali agenti software o della liceità dei contenuti.

4.6. Un obbligo di rivelazione dell'impiego di social bot (due parole distinte)

È importante sottolineare come una normativa di contrasto ai socialbot (una parola unica) non dovrebbe essere delineata in modo da vietare qualsiasi forma di automazione delle attività che si possono svolgere

all'interno di un social network. Questo perché i programmi software per l'automazione di alcune azioni su Internet hanno natura essenzialmente strumentale, che ne permette l'impiego sia per il compimento di attività lecite che illecite[99]. In altre parole, dato che i programmi appartenenti all'ampia categoria degli Internet bot non sono intrinsecamente atti ad offendere e che spesso, al contrario, possono risultare utili[100], sarebbe sbagliato adottare un divieto generale di utilizzo di soluzioni software per automatizzare alcune azioni o passaggi da compiere su Internet o anche solo sui social media.

Per di più, a voler essere particolarmente puntigliosi, si può anche argomentare che qualsiasi azione compiuta su Internet sia sempre il risultato di una serie più o meno lunga di passaggi automatizzati tramite l'impiego di software e che, quindi, un divieto generale di utilizzare soluzioni di automazione digitale di alcune o di tutte le attività su Internet porterebbe con sé delle notevoli incertezze interpretative.

La soluzione più opportuna sarebbe di strutturare un divieto di utilizzo di socialbot (una parola unica) regolamentando prima l'uso della più generale categoria dei social bot (due parole distinte). Questo perché, come detto, nella sottocategoria dei socialbot rientrano programmi software autonomi volutamente programmati e camuffati al fine di nascondere la propria natura artificiale. Al contrario, la più ampia categoria dei social bot identifica più genericamente quei «programmi per computer progettati per utilizzare i social network simulando il modo in cui gli esseri umani comunicano e interagiscono insieme»[101], includendovi anche tutti quelli di cui non viene dissimulata la vera natura.

Insomma in merito ai social bot (due parole distinte) la normativa qui auspicata dovrebbe evitare di porre dei divieti di utilizzo, limitandosi piuttosto a imporre che i profili gestiti da programmi software autonomi presentino una chiara indicazione della natura digitale dei loro amministratori. Una delineazione di un obbligo in questi termini trova un valido esempio nelle disposizioni del *California Business and Professions Code* inserite dal BOT Act. Al Paragrafo 17941 del Codice, infatti, viene sancito che tutte le proibizioni imposte circa l'uso di un programma software autonomo per la gestione di un profilo online che interagisca su Internet con delle persone cadono se l'utilizzatore «rivela di essere un bot». Viene anche specificato, però, come non sia sufficiente una qualsiasi forma di comunicazione al destinatario ma che l'informazione sia «chiara, evidente e ragionevolmente progettata per informare le persone con cui il bot comunica o interagisce che si tratta di un bot». In altre parole, l'obbligo di disclosure sancito prevede anche il rispetto del principio di adeguatezza in merito alle informazioni date. Sebbene su questo aspetto la normativa statale della California sia particolarmente sintetica, vi è da notare come durante il dibattito sulla sua approvazione vennero identificati come indici interpretativi i principi già fissati nell'ambito della pubblicità online[102] da parte della Federal Trade Commission (la stessa legge, dopotutto, presenta una certa identità lessicale tra le proprie disposizioni ed i regolamenti dell'autorità federale)[103]. Richiamandosi a queste prescrizioni è, così, possibile ricavare tutta una serie di fattori da considerare per poter giudicare se la rivelazione dell'identità dell'agente software sia stata effettivamente «chiara, evidente e ragionevolmente progettata».

La stessa considerazione della natura dei social bot (due parole distinte) come neutra e strumentale traspare dal POFMA. Anche in questa normativa non viene proibito qualsiasi impiego di strumenti software per l'automazione di attività su Internet ma solo un loro utilizzo che sia idoneo a porre in pericolo l'ordine pubblico di Singapore, influenzarne gli esiti elettorali o incitare sentimenti discriminatori tra differenti gruppi di persone. Sebbene tale normativa differisca profondamente dalla proposta qui avanzata per quanto riguarda l'individuazione delle condotte affette da antigiuridicità, anch'essa pone alla propria base il principio secondo cui non devono essere vietate tutte le forme di utilizzo di strumenti software autonomi ma solo quelle che possano indurre in errore i destinatari.

Inoltre, una previsione di questo genere non sarebbe nuova nemmeno all'interno del panorama legislativo europeo. Infatti, nella proposta di regolamento comunitario che stabilisce regole armonizzate sull'intelligenza artificiale (la legge sull'intelligenza artificiale)[104] vengono proposti – invero solo «per taluni sistemi specifici di IA» – degli obblighi minimi di trasparenza, in particolare quando vengono utilizzati le chatbot o i deep fake.

(A) Segnalazione dell'utilizzo di un social bot – Sembra, quindi, opportuno non vietare qualsiasi utilizzo di strumenti software autonomi per lo svolgimento automatizzato di attività sui social media. Piuttosto, dovrebbe essere vietato solo il loro utilizzo dissimulato – che li qualificherebbe come socialbot – e dovrebbero essere disposti degli obblighi di «chiara, evidente e ragionevolmente progettata» pubblicità del loro impiego. Questo perché, si ribadisce, l'utilizzo di strumenti informatici per il compimento di determinate attività su Internet può apportare notevoli benefici e dovrebbero essere solo certe modalità con cui vengono impiegati ad integrare un illecito. Un utilizzo trasparente e di facile comprensione di programmi software

autonomi aggiuntivi rispetto a quelli impiegati dai SMP per fornire i propri servizi (tra i quali rientrano sostanzialmente i social bot) permetterebbe ai destinatari delle loro interazioni di avere una visione più accurata e cristallina delle dinamiche di propagazione dei contenuti che verrebbero loro mostrati e potrebbero, quindi, avere una comprensione maggiore della reale popolarità delle opinioni veicolate, soppesandone con maggiore consapevolezza il loro valore.

Il dovere di trasparenza dovrebbe essere imposto anche agli utenti dei social media, a differenza di quanto proposto dalla Sezione 4(c) del BDA Act statunitense. L'indicazione di un obbligo anche a questi destinatari – anziché ai soli fornitori dei servizi di social networking – deriva dalla considerazione dell'opportunità di responsabilizzare non solo i fornitori di un servizio ma anche i suoi utilizzatori, imponendo così anche ad essi il rispetto di una normativa e non il semplice adeguamento a delle condizioni contrattuali di utilizzo.

Inoltre, seguendo l'esempio della Sezione 4(d) del BDA Act, la quale afferma che «nulla in questa sezione deve essere interpretato in modo da richiedere a qualsiasi social media provider di consentire [l'uso di] un programma o un processo software automatizzato inteso a impersonare o replicare l'attività umana online», dovrebbe essere permesso a tutti i SMP di vietare espressamente l'uso di diverse tipologie di social bot a qualsiasi fine ed indipendentemente da qualsiasi obbligo di rivelazione.

(B) Predisposizione di strumenti per segnalare l'uso di un social bot – A contrafforte della previsione alla lettera (A) dovrebbe essere previsto un obbligo in capo a tutti i SMP di dotare le proprie piattaforme social degli opportuni strumenti e procedure per permettere ai loro utenti di segnalare il fatto che uno o più profili di cui sono titolari sono gestiti da social bot. Riprendendo, così, la lettera del Paragrafo 4(c)(2) del BDA Act, il quale impone ai SMP il dovere di istituire ed attuare per ciascuna delle piattaforme possedute o gestite «un processo che consenta ad un utente del sito Internet di un social medium di fornire un avviso chiaro e cospicuo a qualsiasi altra persona o utente».

(C) Demarcazione del profilo gestito da un social bot – Il rispetto degli obblighi di segnalazione dell'impiego di un agente software per la gestione di un account dovrebbe – si spera – rendere maggiormente consapevoli i destinatari delle dinamiche della diffusione potenziata di contenuti e spingerli ad interrogarsi sulla reale popolarità delle opinioni veicolate. In questo modo si dovrebbero poter contrastare quei meccanismi innati negli esseri umani di fiducia nella veridicità delle informazioni che circolano all'interno del loro contesto sociale e di propensione alla condivisione di quanto percepito come ampiamente apprezzato. Per ottenere questo risultato, dovrebbe essere previsto anche l'obbligo di apporre sui profili gestiti da dei social bot un simbolo o una qualche altra forma di demarcazione che indichi chiaramente la natura digitale di cosa sta gestendo il profilo.

Per la specificazione dei caratteri della demarcazione sembra possibile prendere ad esempio la Sezione 17941 (b) del *California Business and Professions Code*, che riprende gli indici fissati nella regolazione della pubblicità online dalla Federal Trade Commission statunitense. Questa Sezione del Codice californiano sancisce che la comunicazione deve essere «chiara, evidente e ragionevolmente progettata per informare le persone con cui il bot comunica o interagisce». I principi formulati dall'autorità federale del commercio, invece, considerano: (i) il posizionamento e rilievo dell'informativa e quanto sia vicina alla relativa rivendicazione; (ii) se la divulgazione sia inevitabile; (iii) se altre parti dell'annuncio distraggano dall'informativa; (iv) se la divulgazione debba essere ripetuta per assicurarsi che sia vista; e (v) se il linguaggio sia comprensibile[105]. Inoltre, vi sono già state diverse pronunce delle autorità giurisdizionali americane che permettono un'interpretazione sufficientemente metodica dell'adeguatezza delle soluzioni adottate[106].

L'inserimento di questo genere di precisazioni seguirebbe le considerazioni espresse dalle istituzioni europee nell'ambito delle interazioni tra persone e sistemi software che «possono comportare rischi specifici di impersonificazione o inganno». Infatti, in ragione di questo rischio è stato suggerito dagli organi di Bruxelles che l'uso di questo genere di programmi software autonomi dovrebbe «essere, in determinate circostanze, soggetto a specifici obblighi di trasparenza»[107].

Vi è da ammettere un problema che potrebbe sorgere dagli obblighi di segnalazione dell'impiego di social bot e di demarcazione dell'account: i limiti temporali di tale etichettatura. Infatti, se un profilo non venisse mai automatizzato dal suo titolare o, all'opposto, venisse affidato in via definitiva ad un agente software, sarebbe ovvio in quali casi esso debba essere contrassegnato. Tuttavia, possono presentarsi non pochi problemi per quanto riguarda i cosiddetti *account cyborg*. Sorgerebbe l'interrogativo circa i limiti temporali di applicazione delle previsioni, ossia se il profilo di un utente che faccia uso di un social bot debba mostrare un chiaro e cospicuo avviso dell'uso del programma solo quando sia effettivamente sotto il suo controllo oppure quando la sua gestione automatizza sia preminente rispetto quella compiuta dall'essere umano. Le possibili soluzioni sono diverse: l'etichettatura del profilo (i) solo quand'è effettivamente sotto il controllo di un social bot, (ii) quando la gestione

automatizza dell'account sia preminente oppure (iii) in modo permanente nel caso sia stato gestito da un agente software anche solo una volta. La più opportuna è, forse, una combinazione delle prime due, ossia l'etichettatura di un profilo come «Automatizzato» quando esso sia effettivamente dato in gestione ad un social bot – situazione che il suo titolare dovrebbe diligentemente notificare – e la permanenza di tale etichettatura, qualora per la maggior parte del tempo l'account sia sotto il controllo dell'agente software.

Infine, si deve sottolineare come l'obbligo dovrebbe valere per tutti i SMP e non solo per quelli che verrebbero identificati come pSMP. Questa estensione dei doveri a prescindere dal numero medio di utenti o di visitatori unici deriva dal fatto che l'attuazione di un sistema di etichettatura dei profili dati in gestione ad un social bot non risulterebbe particolarmente onerosa per il gestore di una piattaforma e, quindi, sarebbe implementabile anche da quelli di piccole dimensioni. Sebbene un'estensione di una disposizione di tal genere a tutti i fornitori di servizi di social networking possa sembrare inefficace per contrastare attivamente le attività di propaganda computazionale portate avanti grazie all'impiego di socialbot, essa avrebbe comunque dei (limitati) vantaggi. Il più importante è che gli utenti avrebbero l'opportunità di venire a conoscenza più facilmente dell'esistenza di profili che non sono la "facciata" dietro cui opera un'altra persona e, quindi, sarebbero – si spera – resi maggiormente consapevoli del fatto che la diffusione di opinioni su Internet potrebbero non essere dovute alla loro validità o effettiva condivisione ma solo ai meccanismi di propagazione basati sull'utilizzo di agenti software.

4.7. Un obbligo di adeguata verifica della clientela per il compimento di alcune azioni

Un'ultima prescrizione per i pSMP che si potrebbe introdurre gravita attorno al concetto di "adeguata verifica della clientela". Questa nozione – che si riconosce poter essere foriera di criticità – riprende lo spirito delle disposizioni contenute all'interno dell'articolo 18 del decreto legislativo 21 novembre 2007, n. 231 (il cosiddetto "decreto antiriciclaggio") ed estese anche ai gestori delle piattaforme online di scambio di valute digitali, i cosiddetti *Digital Currency Exchanges* (DCE). Si può immaginare di tradurre l'obbligo di adeguata verifica in una serie di disposizioni simili per i gestori delle piattaforme social più diffuse. Delle disposizioni di questo genere non permetterebbero solo di contrastare direttamente le attività di propaganda computazionale ma anche di rendere più facilmente identificabile – e, quindi, perseguibile – l'autore di una violazione o di altri reati (come la diffamazione online).

Dato che l'ambito considerato non è quello delle transazioni economiche ma quello dell'espressione delle proprie opinioni, andrebbero adottati diversi accorgimenti che evitino di trasformare una previsione di tutela dell'opinione pubblica in uno strumento di repressione del dissenso. Infatti, fino a quando le legittime espressioni di scontento vengono diffuse sui social media all'interno di un contesto democratico, dove viene tutelato in modo sostanziale il diritto di critica, la formazione di un database che identifichi tutti coloro che pubblicano le proprie rimostranze non comporterebbe particolari problemi. Tuttavia, nel caso in cui ai SMP sia imposto di schedare i cittadini che vivono in un contesto autoritario e repressivo, una norma che privi le persone della tutela offerta dall'anonimato su Internet potrebbe avere dei risvolti drammatici. Quindi, l'adeguato controllo degli utenti dovrebbe essere configurato in termini poco restrittivi per quanto riguarda l'accesso ai servizi offerti dai pSMP.

Più opportuno sarebbe prevedere che l'ingresso alle piattaforme social non sia subordinato al completamento di una procedura di adeguata verifica ma solo che tale adempimento sia necessario per poter compiere determinate tipologie di azioni suscettibili di essere sfruttate per potenziare la diffusione di contenuti e rendere false impressioni di popolarità. In questa maniera (che si riconosce criticabile) verrebbe comunque garantita la possibilità di esprimere le proprie opinioni senza timori di ripercussioni e, allo stesso tempo, sarebbe possibile arginare – ma non interrompere – le condotte automatizzate di sviamento del dibattito pubblico. Un possibile equilibrio tra la libertà di espressione e la tutela dell'opinione pubblica – che potrebbe risultare più sbilanciato di quanto immaginato – potrebbe trovarsi imponendo ai pSMP le seguenti soluzioni.

(A) Mancato conteggio delle azioni dei profili non verificati – Il mancato conteggio delle interazioni con i contenuti di terzi, come la loro ri-condivisione, la loro visualizzazione[108] o l'apposizione di reazioni, compiute da profili non verificati potrebbe essere una valida soluzione per contrastare l'accrescimento artificiale del sostegno o del biasimo a determinati contenuti. Infatti, questa previsione eviterebbe di mostrare degli indici di apprezzamento (come i "Mi piace" di Facebook) artificialmente gonfiati.

Vi è da ammettere che questa ipotesi non impedirebbe ad una persona di creare un profilo verificato e di affidarlo poi ad un socialbot ma si limiterebbe solo a permettere la facile identificazione di un colpevole.

(B) Procedure complesse per interagire con i contenuti sui social media – Altro possibile strumento

per arginare le attività di propagazione e supporto di contenuti da parte dei socialbot potrebbe essere quello di imporre ai profili non verificati lo svolgimento di una procedura più complessa (rispetto a quella riservata agli account verificati) per il compimento di attività sui social media. L'applicazione pratica potrebbe declinarsi in varie forme ma si può immaginare, ad esempio, la risoluzione di test reCAPTCHA.

Andrebbe anche stabilito che i profili per i quali non sia stata svolta la procedura di adeguata verifica dovrebbero essere sottoposti periodicamente a queste procedure di controllo. Una disposizione di questo genere potrebbe essere male accolta dai pSMP, dato che renderebbe meno accessibili i loro servizi agli utenti e potrebbe spingerli verso piattaforme più piccole e non gravate da tali obblighi. La conseguenza potrebbe essere quella di una "corsa al rialzo" riguardo alla frequenza o al numero di azioni dopo le quali svolgere una qualche forma di verifica. Al fine di mitigare le possibili titubanze dei pSMP, si potrebbe considerare l'obbligo di svolgere queste procedure solo per coloro che stiano interagendo con i contenuti di terzi (ossia ri-condividerli o apporvi delle reazioni) che reindirizzino ad elementi esterni alla piattaforma (come le pagine di un altro sito web). Quest'ultima ipotesi avrebbe anche il vantaggio di spingere le persone che non hanno compiuto la procedura di adeguata verifica a prestare maggiore attenzione ai contenuti che stanno cercando di condividere (azione spesso compiuta più sulla base del titolo che del contenuto)[109].

5. Conclusioni

In conclusione, l'avvento dei social media ha portato ad un allargamento dei dibattiti che non sarebbe stato altrimenti possibile. Tuttavia, le architetture alla base di questo sistema di libero scambio di idee hanno tralasciato alcuni aspetti di sicurezza che oggi costituiscono delle vere e proprie falle che permettono a chiunque (umano e non) di mostrarsi con le caratteristiche che più gli aggradano e che, di conseguenza, permettono di non far trasparire su Internet la propria effettiva natura e di distorcere la realtà dei fatti e la popolarità delle opinioni tramite strumenti relativamente semplici come i socialbot.

Gli agenti software al centro di questa disamina sono stati sfruttati negli ultimi anni per potenziare la diffusione di contenuti mistificatori e polarizzanti con finalità soprattutto di propaganda politica. Sebbene il preciso grado di efficacia dell'utilizzo di socialbot non sia stato identificato, in molti hanno convenuto che essi abbiano avuto una certa influenza nel trasmettere alle persone false impressioni di popolarità di determinate opinioni, spesso estremiste, finendo con l'esacerbare la conflittualità nelle discussioni online e con l'avere riflessi profondi anche sulle dinamiche della società offline.

Di fronte a queste tattiche di incremento artefatto del sostegno mostrato online a certe persone od opinioni diverse nazioni hanno adottato o stanno cercando di adottare delle normative con cui poter contrastare l'impiego dei socialbot. Queste normative non sembrano aver centrato appieno l'obiettivo dei legislatori ma sono comunque un primo passo. Anche le Istituzioni comunitarie hanno cominciato ad occuparsi del problema della diffusione su Internet dei discorsi incitanti all'odio, giungendo già ad approvare il CCCDOO, che comprende tra i propri obiettivi quello di «stabilire sistemi e norme chiari per i bot e fare in modo che la loro attività non possa essere confusa con l'interazione umana». Anche il recentissimo DSA ha ripreso – tra le altre – la questione dell'uso di «bot e di altri comportamenti automatizzati o parzialmente automatizzati» per portare avanti operazioni che possono avere effetti negativi sulla società e sulla democrazia.

Tra gli Stati che si sono preoccupati di istituire strumenti di contrasto all'utilizzo di strumenti software autonomi per distorcere il dibattito pubblico online non figura ancora l'Italia, nonostante sia stato accertato che la nostra politica nazionale è già stata teatro di attività di proselitismo questo genere. Si è voluto, quindi, proporre degli spunti su come il cosiddetto "problema bot" declinato nei caratteri della propaganda politica possa essere risolto. Le previsioni qui proposte – che si riconoscono tacciabili di arbitrarietà e di inefficacia – hanno preso largamente spunto da alcune delle normative proposte in altre nazioni, tenendo anche conto della cornice comunitaria, e si sono concentrate principalmente sull'attribuzione ai gestori dei social media di una serie di obblighi di intervento e prevenzione. Non sono mancate proposte di attribuzione di responsabilità in capo ai cittadini per un uso "scorretto" degli agenti software all'interno dei social network, secondo un orientamento maggiormente improntato alla trasparenza che alla limitazione di utilizzo.

In ultima analisi, per riprendere le parole di Hines[110], la proposta di riforma qui esposta «richiede solo che il burattino riconosca i suoi fili, non che indichi il suo burattinaio» ma sarebbe un valido intervento di contrasto agli effetti prodotti dalle attività di propaganda potenziata. Ciò, in particolare, per quanto riguarda l'imposizione di un obbligo di rivelazione dell'utilizzo di agenti software perché tale confessione spingerebbe gli altri utenti ad interrogarsi con maggiore attenzione sull'effettiva popolarità dei contenuti pubblicati sulle piattaforme e, quindi,

promuoverebbe una maggiore riflessione sul valore delle opinioni veicolate.

Ringraziamenti

La realizzazione di questo articolo è stata possibile anche grazie alla cortese disponibilità del dott. Stefano Cresci (ricercatore dell'Istituto di Informatica e Telematica del CNR).

Note

[1]Si vedano, *ex multis*, 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, pp. 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, pp. 757-774; B. McLaughlin, T. Macafee, *Becoming a Presidential Candidate: Social Media Following and Politician Identification*, in "Mass Communication and Society", vol. 22, 2019, n. 5, pp. 584-603.

[2]Si veda, *ex multis*, S.A. Myers, A. Sharma, P. Gupta, J. Lin, *Information Network or Social Network? The Structure of the Twitter Follow Graph*, in "Proceedings of WWW '14: 23rd International World Wide Web Conference (Seoul, Korea), April 2014, pp. 493-498.

[3]Si veda, a titolo esemplificativo, K. Shu, A. Sliva, S. Wang et al., *Fake News Detection on Social Media: A Data Mining Perspective*, in "ACM SIGKDD explorations newsletter", vol. 19, 2017, n. 1, pp. 22-36.

[4]Come sapientemente rilevato in M. Bassanini, G.E. Vigevani, *Primi appunti su* fake news *e dintorni*, in "mediaLAWS", vol. 1, 2017, pp. 11-22, grazie alla facilità e alla velocità di utilizzo di Internet e delle piattaforme online, oggi «ciascun utente, un tempo mero ricettore di informazioni, può divenirne oggi produttore, pur non disponendo del bagaglio di competenza e di esperienza che dovrebbe essere proprio di un giornalista professionista».

[5]Si vedano, *ex multis*, L. Luceri, A. Deb, S. Giordano, E. Ferrara, *Evolution of bot and human behavior during elections*, in "First Monday", vol. 24, 2019, n. 9; N. Grinberg, K. Joseph, L. Friedland et al., *Fake news on Twitter during the 2016 US presidential election*, in "Science", vol. 363, 2019, n. 6425, pp. 374-378; Y. Gorodnichenko, T. Pham, O. Talavera, *Social media, sentiment and public opinions: Evidence from #Brexit and #USElection*, in "European Economic Review", vol. 136, 2021.

[6]Si vedano, *ex multis*, Y. Theocharis, M.E. Roberts, P. Barberá, J.A. Tucker, *From Liberation to Turmoil: Social Media and Democracy*, in "Journal of democracy", vol. 28, 2017, n. 4, pp. 46-59; S. Sanovich, D. Stukal, J.A. Tucker, *Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia*, in "Comparative Politics", vol. 50, 2018, n. 3, pp. 435-482; J.A. Tucker, A. Guess, P. Barbera et al., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, 2018.

[7]P.N. Howard, B. Kollanyi, *Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum*, 2016; M. Del Vicario, F. Zollo, G. Caldarelli et al., *Mapping social dynamics on Facebook: The Brexit debate*, in "Social Networks", vol. 50, 2017, pp. 6-16; UK House of Commons, *Disinformation and 'fake news': Final Report*, HC1791, 2019.

[8]C. Kriel, A. Pavliuc, *Reverse engineering Russian Internet Research Agency tactics through network analysis*, in "Defence Strategic Communication", vol. 6, 2019, pp. 199-227; D. Wolchover, A. Robinson, *Is Brexit a Russia-backed Coup?*, in "New Law Journal", 2020; J. Horder, *Online Free Speech and the Suppression of False Political Claims*, in "Journal of International and Comparative Law", vol. 8, 2021, pp. 15-52.

[9]E. Ferrara, *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election*, in "First Monday", vol. 22, 2017, n. 8; J.B.J. Vilmer, A. Escorcia, M. Guillaume, J. Herrera, *Information Manipulation: A Challenge for Our Democracies*, rapporto del Centre d'Analyse, de Prévision et de Stratégie (CAPS) e dell'Institut de Recherche Stratégique de l'Ecole Militaire della Repubblica francese, Parigi, agosto 2018, 210 p., in particolare a pp. 106-111.

[10]*Ivi*, in particolare a pp. 107-108.

[11]U.S. Office of the Director of National Intelligence (ODNI), *Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution*, 6 January 2017; J.B.J. Vilmer, A. Escorcia, M. Guillaume, J. Herrera, *Information Manipulation*, cit.; R.S. Mueller III, *Report On The Investigation Into Russian Interference in the 2016 Presidential Election*, March 2019.

[12]Cfr. *Open Hearing: Social Media Influence in the 2016 U.S. Election*, U.S. Government Publishing Office, 2018. È disponibile la trascrizione delle testimonianze; J.B.J. Vilmer, A. Escorcia, M. Guillaume, J. Herrera, *Information Manipulation*, cit., in particolare a pp. 84-85; C. Kriel, A. Pavliuc, *Reverse engineering*, cit., pp. 199-227.

[13]Si vedano a titolo esemplificativo 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", 2015, pp. 839-851; D.M. Beskow, K.M. Carley, *Characterization and comparison of Russian and Chinese disinformation campaigns*, in Shu K., Wang S., Lee D., Liu H. (eds.), "Disinformation, misinformation, and fake news in social media", Springer, 2020, pp. 63-81.

[14]N. Abokhodair, D. Yoo, D.W. McDonald, *Dissecting a social botnet*, cit., pp. 839-851.

[15]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", 2011, pp. 93-102.

[16]E. Ferrara, O. Varol, C. Davis et al., *The rise of social bots*, in "Communications of the ACM", vol. 59, 2016, n. 7, pp. 96-104; D. Guilbeault, *Automation, algorithms, and politics. Growing bot security: An ecological view of bot agency*, in "International Journal of Communication", vol. 10, 2016, n. 19, pp. 5003-5021.

[17]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.

[18]Si prenda ad esempio la conclusione a cui è giunta la Corte Suprema dello Stato australiano di Victoria nella propria decisione *Trkulja v. Google Inc. & Anor* [2012] VSC 533. In essa 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. Si veda anche S.M. Benjamin, *Debate: Algorithms and Speech*, in "University Of Pennsylvania Law Review", vol. 161, 2013, n. 6, pp. 1445-1494; F. Pasquale, *Toward a Fourth Law of Robo-*

tics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society, in "Ohio State Law Journal", vol. 78, 2017, pp. 1243-1255.

[19] In merito alla figura del complice "ausiliatore" si veda quanto illustrato in G. Fiandaca, E. Musco, *Diritto penale parte generale*, Zanichelli, 2009, a p. 504, e le recenti precisazioni compiute in Sezione quinta penale della Corte di Cassazione, sentenza del 9 novembre 2021, n. 8973.

[20] B. Mønsted, P. Sapieżyński, E. Ferrara, S. Lehmann, *Evidence of complex contagion of information in social media: An experiment using Twitter bots*, in "PloS one", vol. 12, 2017, n. 9; O. Boichak, S. Jackson, J. Hemsley, S. Tanupabrungsun, *Automated Diffusion? Bots and Their Influence During the 2016 U.S. Presidential Election*, in G. Chowdhury, J. McLeod, V. Gillet, P. Willett (eds.), "Transforming Digital Worlds. Proceedings of 13th International Conference, iConference 2018, March 25-28, 2018, Springer, 2018, pp. 17-26; C. Shao, G.L. Ciampaglia, O. Varol et al., *The spread of low-credibility content by social bots*, in "Nature communications", vol. 9, 2018, n. 1, pp. 1-9.

[21] Come perspicacemente rilevato anche dal deputato Seán Kyne nel corso del dibattito del 13 dicembre 2017 sull'approvazione della *Online Advertising and Social Media (Transparency) Bill* 2017 della Repubblica d'Irlanda.

[22] Per un'introduzione al natural language processing si veda, *ex multis*, P.M. Nadkarni, L. Ohno-Machado, W.W. Chapman, *Natural language processing: an introduction*, in "Journal of the American Medical Informatics Association", vol. 18, 2011, n. 5, pp. 544-551.

[23] S. Cresci, *A decade of social bot detection*, in "Communications of the ACM", vol. 63, 2020, n. 10, pp. 72-83.

[24] Si veda a questo particolare proposito l'analisi sul rischio di lesioni irreparabili della reputazione amplificate dall'uso di socialbot compiuta in A. Tedeschi Toschi, G. Berni Ferretti, *Social media, profili artificiali e tutela della reputazione*, in questa Rivista, 2021, n. 2, pp. 107-130.

[25] E. Ferrara, O. Varol, C. Davis et al., *The rise of social bots*, cit., pp. 96-104; S.C. Woolley, *Automating power: Social bot interference in global politics*, in "First Monday", vol. 21, 2016; D. Guilbeault, *Automation, algorithms, and politics*, cit.

[26] Si vedano, *ex multis*, A. Bessi, E. Ferrara, *Social bots distort the 2016 US Presidential election online discussion*, in "First Monday", vol. 21, 2016, n. 7; E. Treré, *The Dark Side of Digital Politics: Understanding the Algorithmic Manufacturing of Consent and the Hindering of Online Dissidence*, in "IDS Bulletin", vol. 47, 2016, n. 1, pp. 127-138; E. Ferrara, *Disinformation and Social Bot Operations*, cit.

[27] P.T. Metaxas, E. Mustafaraj, *From Obscurity to Prominence in Minutes: Political Speech and Real-Time Search*, in "Proceedings of the WebSci10. Extending the Frontiers of Society Online", April 26-27th, 2010; J. Ratkiewicz, M.D. Conover, M. Meiss et al., *Detecting and tracking political abuse in social media*, in "Proceedings of the International AAAI Conference on Web and Social Media", vol. 5, 2011, n. 1, pp. 297-304; C. Wagner, S. Mitter, C. Körner, M. Strohmaier, *When Social Bots Attack: Modeling Susceptibility of Users in Online Social Networks*, in M. Rowe, M. Stankovic, A.-S. Dadzie (eds.), "Proceedings of the WWW'12 Workshop on 'Making Sense of Microposts'", 2012, pp. 41-48; A. Bessi, E. Ferrara, *Social bots*, cit.; P.N. Howard, B. Kollanyi, *Bots, #Strongerin, and #Brexit*, cit.; S.C. Woolley, *Automating power*, cit.; F. Giglietto, N. Righetti, G. Marino, *Understanding Coordinated and Inauthentic Link Sharing Behavior on Facebook in the Run-up to 2018 General Election and 2019 European Election in Italy*, 2019; K. Stricklin, M.K. McBride, *Social Media Bots: Laws, Regulations, and Platform Policies*, in "CNA Information Memorandum", 2020, in particolare a pp. 22-23.

[28] Si vedano, *ex multis*, S. Shorey, P.N. Howard, *Automation, Big Data and Politics: A Research Review*, in "International Journal of Communication", vol. 10, 2016, n. 24, pp. 5032-5055; A. Bessi, E. Ferrara, *Social bots*, cit.; B. Kollanyi, P.N. Howard, S.C. Woolley, *Bots and Automation over Twitter during the First U.S. Presidential Debate*, in "Comprop data memo", vol. 1, 2016, pp. 1-4; D.M. Beskow, K.M. Carley, *Agent Based Simulation of Bot Disinformation Maneuvers in Twitter*, in "Winter simulation conference (WSC)", 2019, pp. 750-761.

[29] P.F. Lazarsfeld, B. Berelson, H. Gaudet, *The People's Choice. How the Voter Makes Up His Mind in a Presidential Campaign*, Columbia University Press, 1944; B. Berelson, P.F. Lazarsfeld, W.N. McPhee, *Voting: a study of opinion formation in a presidential campaign*, University of Chicago Press, 1954; E. Katz, P.F. Lazarsfeld, *Personal influence: The part played by people in the flow of mass communications*, Routledge, 1955.

[30] Y. Jun, R. Meng, G.V. Johar, *Perceived social presence reduces fact-checking*, in "Proceedings of the National Academy of Sciences", vol. 114, 2017, n. 23, pp. 5976-5981; K.C. Yang, O. Varol, C.A. Davis et al., *Arming the public with artificial intelligence to counter social bots*, in "Human Behavior and Emerging Technologies", vol. 1, 2019, n. 1, pp. 48-61; H. Wolters, K. Stricklin, N. Carey, M.K. McBride, *The Psychology of (Dis)information: A Primer on Key Psychological Mechanisms*, in "CNA Research Memorandum", 2021, pp. 1-74.

[31] 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", vol. 49, 2022, n. 2, pp. 171-195, in cui gli autori hanno messo in luce come su Facebook contenuti con un elevato numero di "Mi piace" vengano percepiti come maggiormente credibili.

[32] D. Centola, *The Spread of Behavior in an Online Social Network Experiment*, in "Science", vol. 329, 2010, n. 5996, pp. 1194-1197; R.M. Bond, C.J. Fariss, J.J. Jones et al., *A 61-million-person experiment in social influence and political mobilization*, in "Nature", vol. 489, 2012, n. 7415, pp. 295-298; L. Muchnik, S. Aral, S.J. Taylor, *Social influence bias: A randomized experiment*, in "Science", vol. 341, 2013, n. 6146, pp. 647-651; C. Becatti, G. Caldarelli, R. Lambiotte, F. Saracco, *Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections*, in "Palgrave Communications", vol. 5, 2019, n. 1, pp. 1-16.

[33] E. Ferrara, O. Varol, C. Davis et al., *The rise of social bots*, cit., p. 103.

[34] Si vedano, ad esempio, le conclusioni della *U.S. Senate Select Committee on Intelligence* sulle interferenze della Federazione Russa nelle elezioni presidenziali americane del 2016. Per quanto riguarda l'Italia, si veda il dossier predisposto per la Commissione parlamentare di inchiesta sulla diffusione massiva di informazioni false, in cui vengono menzionati i «BOT» come «principali strumenti di creazione e diffusione di disinformazione».

[35] S. Woolley, N. Monaco, *Amplify the Party, Suppress the Opposition: Social Media, Bots, and Electoral Fraud*, in "Georgetown Law Technology Review", vol. 4, 2020, pp. 447-461.

[36] V. il testo.

[37] Dette proposte di legge federale sono il *Bot Disclosure and Accountability Act* 2018 (S. 3127), il *Bot Disclosure and Accountability Act* 2019 (H.R. 4536) e il *Bots Research Act* 2019 (H.R. 2860).

[38]Per una disamina delle normative proposte od approvate all'interno dell'ordinamento degli Stati Uniti d'America per contrastare le attività dei socialbot si veda A. Tedeschi Toschi, G. Berni Ferretti, *Il contrasto legislativo ai socialbot e le soluzioni avanzate negli Stati Uniti d'America a livello federale e statale*, in "Diritto di Internet", vol. 3, 2022, in particolare a pp. 461-481.

[39]Per una disamina della normativa della Repubblica di Singapore, chiamata *Protection from Online Falsehoods and Manipulation Act*, si veda A. Tedeschi Toschi, G. Berni Ferretti, *Il contrasto legislativo ai socialbot e le soluzioni avanzate nella Repubblica di Singapore e nella Repubblica d'Irlanda*, in "mediaLAWS", vol. 3, 2022, pp. 352-385 in particolare il paragrafo 2.1.

[40]V. il testo della legge di Singapore.

[41]Si vedano, *ex multis*, International Commission of Jurists, *Singapore: ICJ calls on government not to adopt online regulation bill in current form*, 12 aprile 2019 e Human Rights Watch, *Singapore: Reject Sweeping 'Fake News' Bill Proposed Law Would Excessively Restrict Online Freedom of Speech*, 3 aprile 2019. Anche in dottrina è stato evidenziato il potenziale effetto di questa legge di indurre all'auto-censura. Si vedano, *ex multis*, S. Chen, C.W. Chia, *Singapore's latest efforts at regulating online hate speech*, in "Research Collection School Of Law", vol. 6, 2019, pp. 1-18; H. Lee, T. Lee, *From contempt of court to fake news: public legitimisation and governance in mediated Singapore*, in "Media International Australia", vol. 173, 2019, n. 1, pp. 81-92; R.K. Helm, H. Nasu, *Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation*, in "Human Rights Law Review", vol. 21, 2021, n. 2, pp. 302-328; J.Y. Tay, *No news is good news, but "fake news" is bad news: A comparative analysis of Singapore's and Australia's measures to combat misinformation on social media*, in "Singapore Academy of Law Journal", vol. 33, 2021, n. 2, pp. 600-624.

[42]V. il testo della proposta di legge irlandese.

[43]Per una disamina della normativa proposta nella Repubblica d'Irlanda che affronta anche i socialbot si veda A. Tedeschi Toschi, G. Berni Ferretti, *Il contrasto legislativo ai socialbot*, cit., pp. 352-385, in particolare il paragrafo 2.2.

[44]Già in passato l'Unione europea aveva fornito una prima definizione di cosa sia un discorso incitante all'odio all'interno della Decisione-quadro 2008/913/GAI del Consiglio, del 28 novembre 2008, *sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale.*

[45]V. il *Codice di Condotta sul Contrasto ai Discorsi d'Odio Online Illegali* (CCCDOO).

[46]COM(2018) 236, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Contrastare la disinformazione online: un approccio europeo*; SWD(2018) 408; SWD(2020) 348; SWD(2021) 355.

[47]COM(2018) 236, in particolare al punto 2.2.ii.

[48]COM(2018) 22 final, Comunicazione della Commissione sul piano d'azione per l'istruzione digitale.

[49]COM(2020) 825, Proposta di regolamento del Parlamento europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali).

[50]Considerando 56 della COM(2020) 825.

[51]European Commission, *2018 Code of Practice on Disinformation*.

[52]COM(2021) 262.

[53]European Commission, *2022 Strengthened Code of Practice on Disinformation*.

[54]Si veda il disegno di legge al Senato della XVIII legislatura, n. 1900, per la creazione di una Commissione di indagini sulle attività di diffusione di informazioni e contenuti illegali, falsi, non verificati e ingannevoli su media tradizionali e reti sociali telematiche.

[55]A. Vogt, *Hot or bot? Italian professor casts doubt on politician's Twitter popularity*, in "The Guardian", 22 luglio 2012.

[56]Si vedano gli studi riportati in R. Bracciale, A. Martella, C. Visentin, *From Super-Participants to Super-Echoed. Participation in the 2018 Italian Electoral Twittersphere*, in "Partecipazione e Conflitto", vol. 11, 2018, n. 2, pp. 361-393; F. Giglietto, N. Righetti, G. Marino, *Understanding coordinated and inauthentic link*, cit.; F. Giglietto, N. Righetti, L. Rossi, G. Marino, *It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections*, in "Information, Communication & Society", vol. 23, 2020, n. 6, pp. 867-891; F. Pierri, A. Artoni, S. Ceri, *Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections*, in "PloS one", vol. 15, 2020, n. 1.

[57]S. Cresci, R. Di Pietro, M. Petrocchi et al., *DNA-inspired online behavioural modelling and its application to spambot detection*, in "IEEE Intelligent Systems", vol. 31, 2016, n. 5, pp. 58-64; Id., *The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race*, in "Proceedings of the 26th international conference on world wide web companion", 2017, pp. 963-972; Id., *Exploiting digital DNA for the analysis of similarities in Twitter behaviours*, in "2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)", 2017, pp. 686-695; Id., *Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling*, in "IEEE Transactions on Dependable and Secure Computing", vol. 15, 2018, n. 4, pp. 561-576.

[58]Id., *The paradigm-shift*, cit., pp. 963-972; Id., *Social fingerprinting*, cit., pp. 561-576.

[59]Per una definizione approfondita di bot si vedano 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, pp. 21-35; M. Tsvetkova, R. García-Gavilanes, L. Floridi, T. Yasseri, *Even good bots fight: The case of Wikipedia*, in "PLoS one", vol. 12, 2017, n. 2.

[60]Id., *Even good bots fight*, cit., che riprendono F.C. Cheong, *Internet agents: spiders, wanderers, brokers, and bots*, Thousand Oaks, 1993, 413 p.

[61]N. Abokhodair, D. Yoo, D.W. McDonald, *Dissecting a social botnet*, cit., pp. 839-851.

[62]Ad esempio, un social bot programmato per agire sul sito Internet del social medium Twitter potrebbe essere programmato perché, una volta compiuto l'accesso alla piattaforma, ricerchi tutti gli elementi della classe "tweet" che tra gli attributi presentino il contenimento di un determinato hashtag – come #Cnr_Igsg – e per ciascuno di essi identifichi il tasto virtuale di retweet e lo clicchi.

[63]Y. Boshmaf, I. Muslukhov, K. Beznosov, M. Ripeanu, *The socialbot network*, cit., pp. 93-102, in particolare a p. 93.

[64]*Ivi*, p. 96.

[65]Si vedano, *ex multis*, P.T. Metaxas, E. Mustafaraj, *From obscurity to prominence*, cit.; J. Ratkiewicz, M. Conover, M. Meiss et al., *Truthy: Mapping the spread of astroturf in microblog streams*, in "WWW '11: Proceedings of the 20th International Conference Companion on World Wide Web", 2011, pp. 249-252; C. Wagner, S. Mitter, C. Körner, M. Strohmaier, *When Social Bots Attack*, cit., pp. 41-48; A. Bessi, E. Ferrara, *Social bots*, cit.; B. Kollanyi, P.N. Howard, S.C. Woolley, *Bots and Automation over Twitter* , cit., pp. 1-4; S. Shorey, P.N. Howard, *Automation, Big Data and Politics*, cit., pp. 5032-5055; M.T. Bastos, D. Mercea, *The Brexit Botnet and User-Generated Hyperpartisan*

News, in "Social Science Computer Review", vol. 37, 2019, n. 1, pp. 38-54; S. Cresci, *A decade of social bot detection*, cit., pp. 72-83.

[66]Paragrafo 17940(a) del Codice degli Affari e delle Professioni della California (*California Business and Professions Code*), introdotto nel 2018 dal BOT Act.

[67]M. Lamo, R. Calo, *Regulating Bot Speech*, in "UCLA Law Review", vol. 66, 2019, pp. 988-1028; S. Weissmann, *How Not to Regulate Social Media*, in "The New Atlantis", vol. 58, 2019, pp. 58-64; B. Casey, M.A. Lemley, *You Might Be a Robot*, in "Cornell Law Review", vol. 105, 2020, pp. 287-362; B. Stricke, *People v. Robots: A Roadmap for Enforcing California's New Online Bot Disclosure Act*, in "Vanderbilt Journal of Entertainment & Technology Law", vol. 22, 2020, pp. 838-894.

[68]S. Cresci, R. Di Pietro, M. Petrocchi et al., *DNA-inspired online behavioural*, cit., pp. 58-64, in particolare a p. 2; S. Cresci, R. Di Pietro, M. Petrocchi et al., *Social fingerprinting*, cit., in particolare a p. 4. In essi gli autori hanno riportato come i profili gestiti dai socialbot da loro studiati fossero stati corredati di informazioni personali dettagliate, quali foto del profilo (rubate), brevi biografie e possedessero una fitta rete di contatti (followers e following) genuini.

[69]Si veda la dichiarazione di F. Haugen, ex manager di Facebook, rilasciata innanzi alla Commissione del Senato degli Stati Uniti d'America su Commercio, Scienza e Trasporti e nella quale ha affermato che «i vertici dell'azienda conoscono i modi per rendere Facebook e Instagram più sicuri e non vogliono apportare le modifiche necessarie perché hanno anteposto i loro immensi profitti alle persone», di aver visto che «Facebook ha ripetutamente incontrato conflitti tra i propri profitti e la nostra sicurezza. Facebook ha costantemente risolto quei conflitti a favore dei propri profitti».

[70]Ad esempio, nel 2021 la Meta Platforms, Inc., società proprietaria di Facebook e Instagram, ha riportato un fatturato di 117 miliardi di dollari mentre la ByteDance Ltd., proprietaria di TikTok, ne ha riportato uno di 58 miliardi di dollari.

[71]COM(2020) 825.

[72]Relazione alla Proposta di regolamento del Parlamento europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali), COM(2020) 825. Questo valore viene ripreso anche all'interno dell'art. 25 della proposta stessa, rimandando alla futura adozione di atti delegati da parte della Commissione per il calcolo del numero medio mensile di destinatari attivi del servizio nell'Unione (parr. 1 e 3).

[73]Si pensi, ad esempio, ai blog basati sul software open-source WordPress che certamente sono degli spazi digitali sui quali condividere contenuti e sui quali è anche possibile che questi interagiscano tra loro condividendo opinioni e valutazioni. Essi sono, però, spesso gestiti direttamente dai loro "autori" e sono forniti in forma gratuita e libera da una fondazione no-profit. Vi è da rilevare come il DSA proponga, al Considerando 43, di non applicare le proprie disposizioni «alle microimprese e alle piccole imprese [...] a meno che [...] esse non soddisfino i criteri per qualificarsi come piattaforme online di dimensioni molto grandi».

[74]Impegno 14. e misure 14.1. e 14.2. del 2022 SCPD.

[75]Art. 27, let. c, COM(2020) 825.

[76]Art. 26, COM(2020) 825. In merito a questa previsione, che riecheggia quello di un obbligo generale di sorveglianza (espressamente escluso dall'art. 7 della proposta di regolamento), vi è da tenere a mente che, ai sensi dell'art. 25, essa è inclusa nel novero degli obblighi supplementari applicabili solo alle VLOPs.

[77]L'art. 7, comma 1, del DDL S. 2688 – XVII Leg. (il cosiddetto «DDL Gambaro») recitava, infatti, che «i gestori delle piattaforme informatiche sono tenuti ad effettuare un costante monitoraggio dei contenuti diffusi attraverso le stesse, con particolare riguardo ai contenuti verso i quali gli utenti manifestano un'attenzione diffusa e improvvisa, per valutarne l'attendibilità e la veridicità». Il testo completo del disegno di legge è disponibile sul sito del Senato. Sebbene l'oggetto di regolamentazione di questo disegno di legge fosse diverso dalla normativa qui ipotizzata, esso mostra come il nostro legislatore abbia già proposto di attribuire ai Social Media Provider maggiori obblighi di controllo su quanto accada sulle loro piattaforme.

[78]Si vedano *ex multis* L. Alvisi, A. Clement, A. Epasto et al., *Sok: The evolution of sybil defense via social networks*, in "2013 IEEE symposium on security and privacy", 2013, pp. 382-396; S. Cresci, R. Di Pietro, M. Petrocchi et al., *The paradigm-shift*, cit., pp. 963-972; S. Cresci, R. Di Pietro, M. Petrocchi et al., *Social fingerprinting*, cit., pp. 561-576; L. Luceri, A. Deb, S. Giordano, E. Ferrara, *Evolution of bot*, cit.; R.J. Schuchard, A.T. Crooks, *Insights into elections: An ensemble bot detection coverage framework applied to the 2018 U.S. midterm elections*, in "PLoS one", vol. 16, 2021, n. 1.

[79]C. Yang, R. Harkreader, G. Gu, *Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers*, in "IEEE Transactions on Information Forensics and Security", vol. 8, 2013, n. 8, pp. 1280-1293; K.C. Yang, O. Varol, C.A. Davis et al., *Arming the public*, cit., pp. 48-61; S. Cresci, *A decade of social bot detection*, cit., pp. 72-83; M.K. McBride, Z. Gold, K. Stricklin, *Social Media Bots: Implications for Special Operations Forces*, in "CNA Research Memorandum", 2020, pp. 1-98.

[80]Si vedano, ad esempio, le conclusioni raggiunte in S. Cresci, R. Di Pietro, M. Petrocchi et al., *The paradigm-shift*, cit., pp. 963-972.

[81]All'opposto, la Sezione 40(4) del POFMA della Repubblica di Singapore indica esplicitamente un serie di fattori utili a «determinare se un account online [...] è controllato da un bot». Inoltre, in fondo all'elenco dei fattori di cui il Ministro competente deve tenere conto nella propria analisi viene anche affermato che è possibile sottoporre a valutazione «qualsiasi altro fattore che il Ministro [competente] consideri rilevante».

[82]Si veda l'art. 7, comma 4, del DDL S. 2688 – XVII Leg., che affermava che i gestori delle piattaforme informatiche «nella loro azione di monitoraggio, devono avvalersi anche delle segnalazioni degli utenti effettuate attraverso appositi strumenti accessibili dalla piattaforma medesima».

[83]Articolo 14, COM(2020) 825.

[84]Considerando 40, COM(2020) 825.

[85]Ad esempio, l'art. 20, par. 2, COM(2020) 825, impone che le piattaforme online debbano sospendere «per un periodo di tempo ragionevole il trattamento delle notifiche e dei reclami presentati [...] da persone, enti o reclamanti che con frequenza presentano notifiche o reclami manifestamente infondati». Il successivo par. 3 indica alcuni parametri da tenere in considerazione per valutare se i soggetti considerati abbiano utilizzato in modo abusivo le procedure costituite dal gestore della piattaforma online.

[86]Similmente, anche l'art. 20, par. 1, COM(2020) 825 impone che le piattaforme online debbano sospendere per un periodo di tempo ragionevole la prestazione dei loro servizi ai destinatari del servizio che con frequenza violino le disposizioni della proposta di regolamento (ossia, pubblichino «contenuti manifestamente illegali»).

[87]Sempre secondo le declinazioni del caso, anche l'art. 15, par. 1, DSA contiene un obbligo per il prestatore di servizi digitali di informare il destinatario della segnalazione in merito ad essa ed alle proprie decisioni al riguardo.

[88]Il principio secondo il quale un utente può "difendere" il proprio profilo da provvedimenti di limitazione o blocco dovuti all'illecito uso di socialbot è prevista anche all'interno del POFMA.

[89]Z. Chu, S. Gianvecchio, H. Wang, S. Jajodia, *Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?*, in "IEEE Transactions on dependable and secure computing", vol. 9, 2012, n. 6, pp. 811-824; S. Bradshaw, P.N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*, in "Copyright, Fair Use, Scholarly Communication, etc.", vol. 207, 2019; S. Cresci, *A decade of social bot detection*, cit., pp. 72-83.

[90]Similmente, anche l'art. 17, COM(2020) 825 impone che le piattaforme online costituiscano un sistema interno di gestione dei reclami che sia «tempestivo, diligente e obiettivo». Inoltre, il par. 5 di detto articolo dispone anche che le decisioni «non siano prese avvalendosi esclusivamente di strumenti automatizzati».

[91]Una preoccupazione simile viene espressa anche al Considerando 47, COM(2020) 825, in cui viene anche sottolineato che «presentare con frequenza notifiche o reclami manifestamente infondati [...] mina la fiducia e lede i diritti e gli interessi legittimi delle parti interessate».

[92]Si veda, ad esempio, S. Cresci, R. Di Pietro, M. Petrocchi et al., *Fame for sale: efficient detection of fake Twitter followers*, in "Decision Support Systems", vol. 80, 2015, pp. 56-71.

[93]In merito al concetto di "moderazione dei contenuti" si veda l'art. 2, lett. p, COM(2020) 825.

[94]Misura 14.1., *2022 Strengthened Code of Practice Disinformation*.

[95]Art. 7, comma 1, del DDL S. 2688 – XVII Leg.

[96]Art. 7, commi 2 e 3, del DDL S. 2688 – XVII Leg. Si noti come il comma 3 avrebbe previsto la comminazione di una sanzione per il mancato intervento di rimozione di «notizie false, esagerate o tendenziose che riguardino dati o fatti manifestamente infondati o falsi» secondo una logica simile a quella del POFMA.

[97]Si veda quanto affermato *supra* al paragrafo 1, in particolare alle note 29, 30 e 31.

[98]Si veda, ad esempio, la sentenza della Corte Suprema *Ward v. Rock Against Racism*, 491 U.S. 781 (1989).

[99]Questa osservazione viene avanzata anche in J. Horder, *Online Free Speech*, cit., pp. 15-52.

[100]R. Gorwa, D. Guilbeault, *Unpacking the Social Media Bot: A Typology to Guide Research and Policy*, in "Policy & Internet", vol. 12, 2020, n. 2, pp. 225-248; G. Caldarelli, R. De Nicola, F. Del Vigna et al., *The role of bot squads in the political propaganda on Twitter*, in "Communications Physics", vol. 3, 2020, n. 1, pp. 1-15.

[101]N. Abokhodair, D. Yoo, D.W. McDonald, *Dissecting a social botnet*, cit., pp. 839-851.

[102]Si veda la discussione del *Committee on Arts, Entertainment, Sports, Tourism, and Internet Media* del Senato della California del 21 giugno 2018.

[103]Caratteri di chiarezza ed evidenza della comunicazione sanciti dalla legge della California sono indicati anche da diversi regolamenti e linee-guida della Federal Trade Commission (per le quali le informazioni devono essere presentate «clearly and conspicuously»). Si vedano, ad esempio, le linee-guida del 2013.

[104]COM(2021) 206 final, proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione.

[105]Committee on Arts, Entertainment, Sports, Tourism, and Internet Media, *SB 1001 Bill Analysis 1*, 2018.

[106]Si vedano, ad esempio, le decisioni Be In, *Inc. v. Google, Inc.*, No. 12-CV-03373-LHK, 2013 WL 5568706 (N.D. Cal. Oct. 9, 2013); *Nguyen v. Barnes & Noble Inc.*, 763 F.3d 1171 (9th Cir. 2014); *Norcia v. Samsung Telecomms. Am.*, LLC, 845 F.3d 1279 (9th Cir. 2017).

[107]Considerando 70, COM(2021) 206 final.

[108]Così dando risposta alle osservazioni in S. Weissmann, *How Not to Regulate Social Media*, cit., pp. 58-64, in merito al conteggio falsato delle visualizzazioni su YouTube.

[109]B.D. Horne, S, Adali, *This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News*, in "Eleventh international AAAI conference on web and social media", maggio 2017. Si veda l'intervista di Bloomberg Technology a Frances Haugen, che ha dichiarato che cliccare su un link prima di condividerlo riduce la disinformazione del 10-15%.

[110]M. Hines, *I Smell a Bot: California's S.B. 1001, Free Speech, and the Future of Bot Regulation*, in "Houston Law Review", vol. 42, 2019, pp. 405-435.

* * *

The legislative fight against socialbots. Some ideas for a reform in Italy

Abstract: The advent of the internet, first, and social media, then, allowed an exchange of contents with volumes and speeds previously unthinkable. Within what are to all intents and purposes 'digital town squares' are also active the so-called socialbots. These are programs that, once provided with the login credentials of an account, can manage it autonomously while giving the impression of being a real person. Their speed and precision of reaction in social networks make these software agents dangerously useful for the dissemination of favourable or hostile content. The recent use of socialbot within the electoral and political debate in several nations and its consequences have raised the urgency for state legislators to intervene against this phenomenon. Following the example of the reforms adopted or proposed by various nations around the world, as well as the framework for action set up by the European authorities, several ideas are presented to counter the so-called "bot problem" connected to political propaganda.

Keywords: Social network – Automation – Bot – Propaganda – Disinformation