Autorità per le Garanzie nelle Comunicazioni
Servizio Economico-Statistico

Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali (Delibera n. 423/17/CONS)

RAPPORTO TECNICO

Le strategie di disinformazione *online* e la filiera dei contenuti *fake*

https://www.agcom.it/tavolo-pluralismo-e-piattaforme-online

Indice

1. Introduzione 1
2. Definizioni: il perimetro 4
3. La filiera dei contenuti *fake online* 11
4. Una classificazione delle strategie di disinformazione *online* 18
5. Le strategie commerciali e i modelli di *business* della disinformazione *online* 19
   5.1 La disinformazione *online* che si finanzia attraverso la pubblicità 19
   5.2 Disinformazione e truffe *online* 32
   5.3 Disinformazione commerciale 36
6. Le strategie di disinformazione *online* di natura politico-ideologica 39
   6.1. Le strategie di disinformazione *online* di natura ideologica 41
   6.2. Le strategie di disinformazione *online* di natura politica 45
7. Osservazioni conclusive 50

1. Introduzione

La diffusione dei contenuti informativi attraverso internet ha comportato una **trasformazione** sia nei modelli di consumo delle notizie, sia nei processi di produzione e distribuzione dell'informazione di vario genere in relazione ai quali le piattaforme *online* assumo un ruolo sempre più rilevante[1].

Sono, pertanto, cambiati i **modelli di produzione** della notizia con riferimento, in particolare, alla tempistica con la quale viene creato il prodotto informativo e alla natura stessa dell'informazione tra notizia e intrattenimento. Il ridotto ciclo di produzione del bene informazione nel contesto digitale[2] implica una contrazione del tempo dedicato alla verifica dei fatti e all'attendibilità delle fonti e al controllo della qualità dei contenuti diffusi.

Inoltre, sono mutati il ***business model*** e la **struttura dei costi** del bene informazione, con effetti sulla qualità del prodotto informativo e sulle barriere all'ingresso nel settore. L'industria dell'informazione continua ad essere caratterizzata da elevati **costi fissi di produzione** della prima copia (rappresentati, essenzialmente, dai costi della redazione) la cui incidenza risulta maggiore per coloro che realizzano informazione primaria di qualità; d'altra parte, i **costi marginali** – costi di riproduzione e costi di distribuzione – si sono drasticamente ridotti (fino ad annullarsi). La digitalizzazione del prodotto informativo ha trasformato la notizia in un bene che può essere facilmente digitalizzato e riprodotto, distribuito e consumato anche in gruppo, spesso in modo gratuito, svincolato, pertanto, dal supporto fisico e dall'esigenza di stamparlo e recapitarlo materialmente al consumatore finale.

Inoltre, se da un lato l'utilizzo di modelli di *outsourcing* nella fase produttiva – perseguiti da alcuni editori presenti anche o esclusivamente *online*[3] – ha contribuito al contenimento dei costi delle redazioni, dall'altro lato l'accresciuta dipendenza degli editori provenienti dai mezzi tradizionali dalle piattaforme *online* ha influito sull'incidenza dei **costi di distribuzione** del prodotto informativo. Tenuto conto, infatti, della funzione di porta di accesso all'informazione svolta dalle piattaforme orizzontali (*search*, *social network*, portali, cfr. *infra*), e del rischio di esclusione da significativi flussi di traffico che esse sono in grado di veicolare, gli editori sono sostanzialmente indotti ad aderire alle piattaforme adeguando i propri contenuti (informativi) alle caratteristiche richieste dai servizi di queste ultime. Tale circostanza richiede, inoltre, agli editori attività di continuo aggiornamento tecnologico, attività che si rilevano assai onerose, con evidenti ricadute statiche e dinamiche sull'attività editoriale, sul livello di tecnologia del sistema, nonché sull'entità delle risorse economiche.

Si assiste, peraltro, ad una riduzione dei **costi d'entrata** nel settore e a un conseguente ampliamento degli operatori attivi dal lato dell'offerta che si affiancano agli editori provenienti dai mezzi tradizionali. Accanto alle agenzie di stampa che diffondono le notizie in forma strutturata con propri siti di informazione, in rete operano anche numerosi nuovi soggetti editoriali che forniscono servizi assai differenziati tra loro, tra cui: editori esclusivamente *online*, i portali che offrono una molteplicità di servizi *web* compresa una propria sezione di notizie, i *social network*, i motori di ricerca, fino ad arrivare a tutti quei soggetti (singoli giornalisti e più genericamente *influencer*) che in rete (con propri *blog*, pagine *social*) forniscono informazioni e commenti sull'attualità. Tali soggetti, pur non rientrando nell'ambito di una testata registrata e, quindi,

[1] Cfr, AGCOM, Rapporto sul Consumo di Informazione, febbraio 2018.

[2] Il ciclo di 24 ore di produzione della notizia proprio dei *business* dei quotidiani *offline* è sostituito da un processo di produzione di articoli e notizie che sono pubblicate e aggiornate numerose volte nel corso della giornata e appaiono in siti di informazione che monitorano reciprocamente i rispettivi contenuti. Per un'analisi circa la tempistica e i meccanismi di distribuzione dell'informazione *online* si veda J. Age, N. Herve, M.-L. Viaud, (2015), *The Production of Information in an Online World*, NetInstitute.org *working paper* nr 2015-05

[3] Tale modello è stato utilizzato, ad esempio da Huffington Post editore esclusivamente *online*. Cfr. Age, N. Herve, M.-L. Viaud, (2015), *The Production of Information in an Online World*, cit.

esulando spesso dai canoni della professione giornalistica, costituiscono altre fonti primarie di informazione *online* per i cittadini[4].

L'abbassamento delle **barriere all'ingresso**, peraltro, ha consentito agli stessi utenti di partecipare alla produzione e ri-produzione di contenuti informativi e, insieme ad altri elementi caratterizzanti tali beni (fra cui, la visibilità del prodotto mediatico, l'interesse ad essere parte del sistema dei media, le finalità non esclusivamente economiche della produzione derivante dalla natura di *creative good* dello stesso), ha contribuito a rendere disponibili *online* una grande varietà di contenuti, rendendo il *web* un ambiente ricolmo di stimoli e *input* per gli utenti.

In questo nuovo contesto di riferimento, si assiste ad una contrazione dello spazio di esercizio del ruolo di intermediario svolto dall'editore tradizionale e ad un incremento della concorrenza per accaparrarsi il tempo di attenzione dell'utente per l'*infotainment online*. Nel dettaglio, la diffusione del nuovo mezzo di comunicazione ha comportato il passaggio da un modello di integrazione verticale delle diverse fasi della catena del valore, tipica dell'editoria *offline* al cui interno l'editore esercitava un controllo (diretto o indiretto), ad una separazione dei diversi stadi del processo produttivo all'interno dei quali spicca il ruolo delle piattaforme *online*.

In altri termini, si configura un processo di **disaggregazione, autoproduzione** e **disintermediazione** dell'offerta informativa tradizionale e di successiva riaggregazione e re-intermediazione da parte di fonti algoritmiche. I motori di ricerca e i *social network* sono spesso definiti "fonti algoritmiche" di informazione per richiamare la **personalizzazione algoritmica** resa possibile dalla quantità e qualità dei dati raccolti sugli individui che caratterizza i processi di generazione e divulgazione dei contenuti informativi disponibili attraverso le stesse. Gli algoritmi sottostanti al loro funzionamento, utilizzati in particolare, ma non solo, per filtrare le notizie disponibili e presentarle agli utenti secondo un ordine, spesso personalizzato, assumono un ruolo decisivo nel determinare le modalità di fruizione dell'informazione da parte degli utenti orientando significativamente il successo o meno in termini di *audience* di una notizia (o di un editore) rispetto ad un'altra e nel determinare le scelte di editori e giornalisti[5].

Tenuto conto del ruolo di *gatekeeper* per l'accesso alle informazioni e di luogo prediletto dagli utenti, le **fonti algoritmiche** rappresentano un punto d'approdo, imprescindibile per gli editori al fine di raggiungere i propri consumatori, che condiziona le strategie di distribuzione dei contenuti informativi attuate da questi ultimi; d'altra parte, nel medio-lungo periodo gli editori rischiano di perdere il contatto diretto con il pubblico e, quindi, la propria riconoscibilità a favore di quella dell'intermediario che ha reso possibile la fruizione dell'informazione *online*. Inoltre, l'editore, pur rimanendo responsabile dei contenuti prodotti, secondo il nuovo modello di distribuzione dell'informazione *online*, vede ridotta la possibilità di esercitare un controllo nella fase di diffusione degli stessi.

A questo si aggiunge che l'informazione *online* continua ad essere finanziata in modo prevalente attraverso la **pubblicità *online***[6] le cui caratteristiche e modelli distributivi di compravendita (soprattutto se basati su meccanismi automatici) offrono incentivi a tutti i *player* (piattaforma, produttore di notizie, inserzionista – sebbene con alcuni limiti) nella distribuzione di contenuti informativi – persino quelli di disinformazione – capaci di generare maggiore traffico.

I cambiamenti esaminati, oltre a poter inficiare sulla qualità dell'informazione *online*, contribuiscono a creare un ambiente favorevole per la produzione e distribuzione di *fake news* e/o di prodotti informativi di scarsa qualità; inoltre i meccanismi automatici di presentazione delle notizie all'interno delle piattaforme, unite alle azioni condotte dagli utenti, facilitano la propagazione in modo virale dei contenuti di disinformazione.

---

[4] In tale contesto di riferimento l'entrata di soggetti con opinioni e visioni estreme in grado di intercettare un'audience di nicchia sarebbe, inoltre, facilitata. Cfr. S. Mullainathan, A. Shleifer (2005), *The market for news*, American Economic Review, 2005, pp 1031-1053; M. Gentzkow, J.M. Shapiro, D. F. Stone (2016), *Media bias in the marketplace: Theory*. Chapter 14, Handbook of Media Economics, Vol. 1B, editors: Simon Anderson, J. Waldfogel, D. Stromberg.

[5] Cfr. Agcom (2018), *Big data*, Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera n. 217/17/CONS.

[6] Cfr. B. Martens, W. L. Aguiar, M. E. Gomez, F. Herrera Muller-Langer, *The digital transformation of news media and the rise of disinformation and fake news*, 2018, EUR - Scientific and Technical Research Reports, pp. 17-18.

Dal lato della domanda, il processo di digitalizzazione e la distribuzione dei contenuti attraverso internet ha condotto all'affermazione di nuovi **modelli di fruizione** dei media caratterizzati da fenomeni di *cross-medialità* e simultaneità negli usi dei mezzi di comunicazione per finalità informativa, che, sebbene abbiano comportato un aumento dell'accesso dei cittadini alle fonti informative, possono al contempo accrescere il rischio di un consumo superficiale e disattento del contenuto informativo.

Analogamente, un'altra tendenza che potrebbe amplificare il fenomeno della scarsa attenzione dedicata all'informazione è rappresentata dalla **frammentazione del consumo** di informazione (intesa sia come consumo di fonti di informazione presenti sui diversi mezzi, sia come consumo di fonti diverse all'interno dello stesso mezzo), diffusasi grazie all'ampliamento del panorama delle fonti informative disponibili e grazie allo spacchettamento del prodotto informativo e la riaggregazione dei contenuti che vengono fruiti in modo non lineare, sotto forma di frammenti veicolati in modi diversi sulle differenti piattaforme.

La minore attenzione e il minor grado di approfondimento che scaturisce dai nuovi modelli di consumo dell'informazione rende il cittadino maggiormente esposto al **pericolo di disinformazione**, di confusione fra notizie reali e false (*fake*), nonché al rischio che la frammentazione del consumo possa tradursi in una **frammentazione sociale**, generando fenomeni di polarizzazione ideologica volti a minare la coesione sociale.

In questo contesto di riferimento, sia dal lato dell'offerta sia dal lato della domanda, si realizzano condizioni che, come si vedrà nel rapporto, favoriscono l'emergere perfino di un ***business* economicamente sostenibile dei contenuti falsi o *fake***. Il termine 'fake' individua un concetto un po' più complesso della nozione di notizia falsa, in quanto rappresenta anche una manipolazione di una notizia vera, presentando in modo artefatto elementi fattuali veri assieme a suggestioni e manipolazioni che ne possono alterare il contenuto. Le distorsioni dell'informazione *online*, infatti, nelle molteplici forme che assumono, si inseriscono all'interno del mercato dell'informazione *online* come prodotti informativi, talora dando luogo a mercati paralleli, in cui vengono scambiati beni e servizi per la creazione, produzione e diffusione di contenuti *fake*. Pertanto, il fenomeno ha una sua rilevanza economica in senso stretto, in quanto attira e produce risorse economiche, ha al centro la concorrenza per l'attenzione degli utenti e la loro profilazione in base ai dati rivelati dalla fruizione *online* e, più in generale, ha un rilievo strategico, non solo dal punto di vista economico ma anche ideologico, politico e sociale.

In tale quadro, e nell'ottica di individuare strumenti di contrasto ai fenomeni di disinformazione *online*, è utile preliminarmente ripercorrere il processo attraverso cui viene ideato, prodotto e distribuito il contenuto *fake* come strumento di disinformazione. Una strategia di disinformazione non si esaurisce in un singolo contenuto *fake*, ma si basa su una precisa, ripetuta e sistematica manipolazione di contenuti veicolati e alimentati ai fini della loro diffusione, anche per il tramite di falsi *account*, a specifici gruppi di utenti opportunamente profilati a tal fine. Nell'ambito dei lavori del Tavolo Tecnico è, infatti, emersa l'opportunità di ricostruire la filiera dei contenuti *fake* e il modo in cui questi vengono valorizzati (monetariamente o meno) all'interno di strategie di varia natura.

L'analisi effettuata nel rapporto, quindi, propone *in primis* un **impianto definitorio** sul perimetro relativo ai disturbi dell'informazione *online*, e successivamente un quadro delle principali attività, dell'organizzazione, delle tecnologie e dei flussi di risorse utilizzate per la creazione, produzione e distribuzione dei contenuti *fake*, dedicando un'attenzione particolare alle **strategie di disinformazione *online***. Queste ultime, infatti, sono caratterizzate da una struttura organizzata che si pone obiettivi, di natura economica e non, sia di breve-medio, sia di lungo periodo. In particolare, si è rinvenuta l'opportunità di effettuare un esame ad ampio spettro di tali strategie, ricomprendendo tanto quelle che si fondano su motivazioni di ordine economico, quanto quelle che si basano su obiettivi ideologico-politici, volti ad alimentare l'auto-selezione di gruppi di utenti e la loro polarizzazione.

Dall'analisi e descrizione delle strategie e dei modelli di *business* emergenti, con l'ausilio di alcuni *case study*, il documento intende mettere in luce le **criticità** da affrontare per il contrasto della disinformazione *online* con l'obiettivo di coadiuvare le attività dei gruppi di lavoro del Tavolo per l'individuazione delle adeguate soluzioni tecniche, di mercato e di autoregolamentazione del fenomeno.

Tenuto conto della complessità del fenomeno della disseminazione della disinformazione *online*, la presentazione di *case studies*, operata nei paragrafi successivi, illustrerà alcuni **archetipi generali**, utili a individuare gli elementi tipici e salienti dei fenomeni di disinformazione *online*, tra i quali la polarizzazione e l'*hate speech* indotto e alimentato da false notizie, nella consapevolezza che tali casi non possono essere esaustivi rispetto alle vastità della tipologia delle distorsioni dell'informazione *online* e alla varietà delle strategie e delle tecnologie disponibili caratterizzanti i processi di creazione, produzione, distribuzione dei contenuti *fake* e della relativa valorizzazione.

2. Definizioni: il perimetro

Il dibattito sulla disinformazione *online* è emerso in modo rilevante a livello internazionale nel 2016, dapprima durante la campagna elettorale per il referendum del Regno Unito sull'uscita dall'Unione Europea (cd. Brexit) e, ancor più, a seguito delle elezioni presidenziali USA, con il diffondersi di diversi studi sulla propagazione di notizie false durante la relativa campagna elettorale.

A partire da novembre 2016, studi e indagini hanno dimostrato la forte circolazione di contenuti cd. *fake* sulle principali piattaforme *online* (in special modo Facebook)[7], la presenza, sugli stessi canali, di *account* che diffondevano informazioni false e tendenziose[8], ovvero di inserzioni pubblicitarie, su temi di campagna elettorale, i cui spazi risultavano acquistati da *account* falsi di origine russa[9]. Successivamente, ricerche accademiche scientificamente fondate e basate sull'utilizzo di grandi masse di dati hanno evidenziato la portata in termini quantitativi della diffusione di *fake news*[10], ribadendo l'attenzione della comunità scientifica verso un fenomeno evidentemente patologico rispetto ai principi di correttezza dell'informazione.

Non a caso, a fine 2016, l'Oxford Dictionary ha designato **post-truth** (post-verità), *"an adjective defined as 'relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief'*[11]*"*, come parola dell'anno.

Più che dal termine post-verità, il dibattito è stato però segnato soprattutto dall'espressione **fake news**, utilizzata indistintamente per indicare una vasta gamma di concetti, con diversi livelli di criticità, tra cui:

- notizie provenienti da fonti che non hanno effettuato un'opportuna attività professionale di verifica delle fonti;
- notizie satiriche che, talvolta, lette fuori contesto, possono essere percepite dai cittadini come reali;
- notizie provenienti da fonti che alimentano teorie complottistiche o cospirazioniste;
- notizie provenienti da fonti specializzate in gossip o *rumors* e pseudoscienza, caratterizzate dal ricorso ad annunci (di eventi di attualità o scoperte scientifiche) non verificati;
- *hate news* provenienti da fonti che promuovono razzismo, misoginia, omofobia, e altre forme di discriminazione;

[7] Cfr. Silverman C., *This Analysis Shows How Fake Election News Stories Outperformed Real News On Facebook*, 16 novembre 2016, www.buzzfeed.com; Shavit N., *Data on Facebook's fake news problem*, 29 novembre 2016, www.jumpshot.com; Pagella Politica per AGI, *Referendum e fact checking: la notizia più condivisa è una bufala*, 2 dicembre 2016, https://tinyurl.com/ydy44526

[8] Cfr. Silverman C. et al., *Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate*, 20 ottobre 2016, www.buzzfeed.com; Albright J., *The #Election2016 Micro-Propaganda Machine*, novembre 2016, www.medium.com

[9] Negli Stati Uniti, il tema è stato discusso al Senato ed è stato alla base della proposta legislativa *Honest Ads Act*. Attualmente, le interferenze russe nella campagna elettorale presidenziale statunitense del 2016 sono oggetto di un'indagine speciale affidata al Direttore FBI Robert Mueller. Anche Facebook ha pubblicato un documento sulle interferenze straniere nella stessa campagna elettorale: Weedon J., Nuland W., Stamos A. (2017), *Information Operations on Facebook*. Per maggiori informazioni, si veda *infra*, par. 6.2.

[10] Cfr., fra gli altri, Alcott H., Gentzkow, M. (2017), "Social Media and Fake News in the 2016 Election", *Journal of Economic Perspectives*, 31 (2), pp. 211–236; Vosoughi S., Roy D., Aral S. (2018), "The spread of true and false news online", *Science*, 359(6380), pp. 1146-1151.

[11] https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016

- notizie (del tutto o parzialmente) corrette che però utilizzano una titolazione sensazionalistica (a scopi di *clickbaiting* – cattura dei *click*);
- notizie provenienti da fonti che forniscono, in maniera tendenziosa, informazioni a supporto di specifici punti di vista e orientamenti politici[12];
- infine, contenuti che imitano le notizie nella forma ma non nel processo organizzativo e nello scopo[13].

Alla luce di quanto premesso, appare evidente come l'espressione *fake news* – anche in virtù del suo recente utilizzo[14] in maniera eccessivamente inclusiva e generica[15] – risulti poco adatta a cogliere le molteplici sfumature dei problemi dell'informazione *online* e a delimitare in maniera operativa gli aspetti di interesse del Tavolo Tecnico, che infatti si è concentrato sin dall'inizio sul concetto di disinformazione *online*.

In tal senso, si ritiene piuttosto che i contenuti oggetto di specifica rilevanza nell'ambito del Tavolo Tecnico siano in realtà le notizie, anche completamente inventate, fabbricate e diffuse (in particolar modo attraverso le piattaforme *online*) allo scopo di ingannare il pubblico e manipolarne l'orientamento, attraverso il ricorso a stati emotivi, per motivi ideologici, politici o di vantaggio economico[16]. La rilevanza del tema della disinformazione *online*, infatti, è soprattutto legata all'entità delle ricadute negative che il fenomeno può generare per la formazione dell'opinione pubblica e, quindi, dal punto di vista sociale e politico. Un tema chiaramente di forte interesse per l'Autorità, che vede nel pluralismo e nella correttezza dell'informazione i suoi valori fondativi, e che a tal fine ha concentrato la propria attività di indagine anche sull'evoluzione dell'offerta di informazione tramite le piattaforme *online* (cfr. l'Indagine conoscitiva in corso su "*Piattaforme digitali e sistema dell'informazione*").

Tenendo presente questo angolo di visuale, e con l'obiettivo di giungere a fornire delle definizioni che rispondano alle esigenze operative del Tavolo Tecnico e siano sufficientemente adattabili ad eventuali cambiamenti che dovessero rendersi necessari per l'evolversi del contesto di riferimento, si ritiene opportuno compiere una preliminare ricognizione di tutti i termini e le nozioni che finora hanno caratterizzato la tassonomia del dibattito sui fenomeni patologici del sistema informativo *online*.

Nella discussione in ambito internazionale, volendo astrarsi dal concetto di "*fake news*", si è soliti riferirsi all'insieme di tali fenomeni anche usando espressioni come "*disordini*", "*disturbi*", "*distorsioni*", "*fallimenti*", "*inquinamento*" *dell'informazione* veicolata su internet. Si tratta, in ogni caso, di un insieme ampio e variegato, che comprende fenomeni con caratteristiche, finalità, intenzionalità e grado di problematicità sotto il profilo della tutela del pluralismo informativo anche molto differenti tra loro.

Al riguardo, i più recenti tentativi di classificazione[17] hanno condotto all'individuazione di alcune macro-categorie di fenomeni.

Nel dettaglio, quando si fa riferimento a un tipo di contenuti informativi non veritieri o inaccurati non creati con un intento doloso ma comunque atti ad essere recepiti dagli utenti come notizie su fatti reali, si parla di **misinformation**. In questo caso, l'informazione diffusa risulta scorretta per leggerezza, per un'errata comprensione dei fatti, per una mancata verifica delle fonti o anche volutamente ma con l'intento di scherzare o deridere. In questa categoria, rientrano generalmente la *satira/parodia* (contenuti informativi volutamente

[12] Rielaborazione basata su Zimdars M. (2016), *False, Misleading, Clickbait-y, and Satirical "News" Sources*.
[13] Lazer D.M.J. *et al*. (2018), "The science of fake news", *Science*, 359 (6380), pp. 1094-1096.
[14] Peraltro l'espressione *fake news* è stata originariamente impiegata anche in relazione ai media tradizionali (cfr. Center for Media and Democracy (2016), *Fake TV News: Widespread and Undisclosed*).
[15] Cfr., tra gli altri, Tandoc Jr. E. C., Lim Z. W., Ling R. (2017), "Defining 'Fake News': A Typology of Scholarly Definitions", *Digital Journalism*, 5 (7), pp. 1-17.
[16] Cfr. Pennycook G., Cannon D., Rand D.G. (2017), *Prior exposure increases perceived accuracy of fake news*, *mimeo*. Sulla diffusione di notizie false per motivi ideologici cfr. Lewis R., Marwick A., "Taking the Red Pill: Ideological Motivations for Spreading Online Disinformation", in Schudson, M. et al. (2017), *Understanding and Addressing the Disinformation Ecosystem*; sulla diffusione di notizie false per motivi di vantaggio economico, in particolare tramite flussi pubblicitari *online*, cfr. Tambini D. (2017), *Fake News: Public Policy Responses*.
[17] Cfr. Rapporto del Consiglio d'Europa, *Information disorder: Toward an interdisciplinary framework for research and policy making*, (2017), commissionato a First Draft e Shorenstein Center on Media, Politics and Public Policy/Harvard Kennedy School e realizzato da Claire Wardle e Hossein Derakhshan, https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c ; Wardle C. (2017), "Fake news. It's Complicated", *First Draft News*, https://firstdraftnews.org/fake-news-complicated/; Jack C. (2017), *Lexicon Of Lies. Terms for Problematic Information*, https://datasociety.net.

falsi e caricaturali ma passibili di essere assimilati come reali, divulgati per prendere in giro aspetti o personaggi della vita contemporanea), i *contenuti fuorvianti* (l'uso fuorviante di informazioni per inquadrare un argomento o una persona) e le *false connessioni* (laddove i titoli, le immagini, o le didascalie utilizzate non supportano il contenuto informativo).

Diversamente, quando i contenuti informativi sono fondati su fatti reali (molte volte a carattere privato) ma contestualizzati in modo da poter essere anche virali e divulgati con il preciso intento di danneggiare una persona, un'organizzazione o un Paese, o affermare/screditare una tesi, si usa il termine **malinformation**. Esempi di questo tipo di disturbi dell'informazione possono essere le *fughe di notizie* volte a favorire episodi di incitamento all'odio (*hate speech*) e molestie (*online harassment*), o l'*amplificazione di notizie* (anche satiriche) fondate su fatti reali per affermare/screditare una determinata tesi.

Nel caso in cui la manipolazione dei contenuti informativi veicolati *online* è caratterizzata sia da falsità, sia dall'intento doloso, si fa riferimento alla nozione di **disinformation**. In questa categoria sono incluse tutte quelle informazioni false (ma suscettibili di essere recepite come vere), deliberatamente create per danneggiare, anche grazie all'impatto emotivo, una persona, un gruppo sociale, un'organizzazione o un Paese, o affermare/screditare una tesi, e consapevolmente diffuse per scopi politici, ideologici o commerciali (incluso il *clickbaiting*), quasi sempre attraverso piattaforme *online* che tendono ad aumentarne la propagazione massiva. Si tratta, infatti, di contenuti contraddistinti da viralità, ossia dall'attitudine, in base all'argomento trattato, a trasferire stati emotivi e percezioni su larga scala. Nello specifico, sono riconducibili a questa tipologia fenomeni quali *false contestualizzazioni* (che si verificano quando contenuti veritieri sono condivisi con false informazioni di contesto), *contenuti veicolati da false fonti* (contenuti divulgati da fonti false che impersonano fonti autentiche), *contenuti creati in maniera artificiosa* (contenuti totalmente falsi e infondati creati per ingannare e/o danneggiare), *notizie manipolate* (informazioni o immagini veritiere manipolate in modo volutamente ingannevole).

Alle categorie sopra esposte, una parte degli studi esistenti in materia affianca ulteriori tipi. Tra questi si annoverano **information operations** e **propaganda**. Più precisamente, il termine *information operations*, coniato in ambito militare, designa un uso strategico di risorse tecnologiche, operative/militari e psicologiche per minare le capacità informazionali di soggetti/organizzazioni (politici, economici, statuali) rivali e sostenere soggetti/organizzazioni politicamente, ideologicamente, economicamente affini. Il termine *propaganda*, invece, quando utilizzato nell'ambito dei disturbi dell'informazione *online*, viene associato alle presentazioni selettive di informazioni, alle cornici persuasive e al ricorso ad appelli emotivi da parte di attori politici, economici o statuali.

Stanti gli aspetti differenziali appena illustrati tra le varie categorie di contenuti e fenomeni ritenuti critici dal punto di vista informativo, è comunque possibile identificare dei tratti salienti che accomunano le stesse.

In primo luogo, tutte le categorie precedenti prevedono la distribuzione dei contenuti su internet. In proposito, vale sottolineare come le caratteristiche del mezzo – delle piattaforme *online* in particolar modo – contribuiscano a favorire la diffusione delle notizie e innescare fenomeni di "viralizzazione" dei contenuti informativi. Infatti, per mezzo delle piattaforme *online*, i meccanismi di personalizzazione automatica, da un lato, e le azioni di condivisione di contenuti compiute dagli utenti, dall'altro, facilitano la proliferazione di notizie false o distorte, la diffusione in modo veloce e la propagazione virale di tali contenuti. A ciò, peraltro, si aggiunga la crescente tecnica, peculiare soltanto di internet, di divulgazione dei predetti contenuti anche attraverso i cd. *bot* (*account* automatizzati che impersonano gli umani), con il preciso scopo di originare, amplificare e accelerare la diffusione di determinate informazioni, creando artatamente il cd. *snowball effect*.

In secondo luogo, per tutte le categorie precedentemente richiamate, le distorsioni informative si concretizzano nell'arco di un processo articolato in quattro fasi: creazione, produzione, distribuzione e valorizzazione dei contenuti informativi. Nella prima fase, viene creato il messaggio che si intende diffondere; nella seconda fase, il messaggio viene trasformato in un prodotto informativo *online*; nella terza fase, il medesimo contenuto viene pubblicato e diffuso tra gli utenti, tenendo presente che dal momento in cui il prodotto diviene pubblico può essere rielaborato (divenendo un nuovo prodotto) e rilanciato anche dagli utenti stessi; nella quarta i contenuti

vengono valorizzati producendo o meno un guadagno monetario più o meno immediato raggiungendo lo scopo desiderato (cfr. *infra*, par. 3).

Infine, in tutte le categorie descritte, è possibile individuare una componente soggettiva e una componente oggettiva del fenomeno, a seconda che si considerino i soggetti coinvolti (la fonte da cui proviene il messaggio – ossia, i creatori, produttori e distributori – e i destinatari del messaggio), ovvero l'oggetto del messaggio informativo divulgato. Sotto il profilo soggettivo, le fonti da cui provengono i messaggi informativi possono essere organizzazioni editoriali e non, soggetti politici, gruppi di utenti, ecc., includendo anche, come anticipato, i *bot*. Le motivazioni alla base della creazione e diffusione di determinati messaggi informativi possono essere molteplici e riguardare la sfera economica (massimizzare i profitti attraverso la raccolta pubblicitaria), politica (screditare una parte politica, favorirne un'altra), ideologica (affermare una data ideologia o punto di vista). Conseguentemente, in base alla tipologia e alle motivazioni dei soggetti da cui provengono i messaggi, possono variare i soggetti individuati quali destinatari dei contenuti informativi. La scelta del *target*, in ogni caso, non prescinde dalla consapevolezza che ciascun utente possieda la propria rete di legami forti e deboli, *online* e *offline*, con altri individui, e che ciascun utente possa compiere diverse azioni informative (da cliccare sul *link* di una notizia, ad esprimere una reazione rispetto alla stessa, condividerla, commentarla, fino a partecipare a una discussione sulla notizia e postare proprie immagini, foto e video in merito all'argomento). Sotto il profilo oggettivo, un messaggio informativo si contraddistingue, fra le altre cose, per il formato (testo, audio, video, …), la durata (i messaggi possono essere concepiti per circolare ed esplicare i propri effetti per un lungo periodo, ovvero solo per un lasso temporale più breve o appena per un momento), il grado di falsità/scorrettezza/manipolazione, ovvero l'intenzionalità di arrecare un pregiudizio ovvero di generare *hate speech* e *hate harm*.

Tra l'altro, la distinzione tra i due piani, soggettivo e oggettivo, può riflettersi nella scelta dell'approccio metodologico seguito nello studio della materia. In proposito, si riscontra, da una parte, la presenza di lavori che indagano le distorsioni dell'informazione *online* ponendo l'accento sulle fonti da cui provengono i contenuti informativi: in questo caso si tende a riferire l'analisi a una selezione di fonti (siti e pagine *web*) identificati (generalmente, da organizzazioni indipendenti) come divulgatori di notizie false e ingannevoli[18]. D'altra parte, si rileva la presenza di studi che pongono l'attenzione sul contenuto veicolato, per cui solitamente si riferisce l'analisi a una selezione di messaggi/notizie classificate come false da organizzazioni indipendenti[19]. Si rinvengono, altresì, studi che indagano la materia adottando un approccio metodologico misto, ossia riferendo l'analisi tanto alle fonti da cui vengono divulgate le notizie quanto ai contenuti delle notizie stesse[20].

Sulla base delle considerazioni che precedono, è possibile quindi sintetizzare come segue alcuni concetti principali e definire in maniera puntuale l'oggetto di studio del Tavolo Tecnico:

[18] Cfr., tra gli altri, Del Vicario M., Bessi A., Zollo F., Petroni F., Scala A., Caldarelli G., Stanley H. E., Quattrociocchi W. (2016), "The Spreading of Misinformation Online", *Proceedings of the National Academy of Science* 113(3); Fletcher R., Cornia A., Graves L., Nielsen R. K. (2018), *Measuring the reach of "fake news" and online disinformation in Europe*.
[19] Cfr., ad esempio, Vosoughi S., Roy D., Aral S. (2018), "The spread of true and false news online", cit.
[20] Cfr., ad esempio, Bessi A., Coletto M., Davidescu G. A., Scala A., Caldarelli G., Quattrociocchi W. (2015), "Science vs Conspiracy: Collective Narratives in the Age of Misinformation", *PLoS ONE* 10(2).

	DEFINIZIONE
FAKE NEWS	Termine spesso utilizzato in maniera ampia e generica per indicare indistintamente una vasta gamma di disturbi dell'informazione. Può essere utilizzato per indicare notizie completamente inventate, create artificiosamente, anche aventi carattere sensazionalistico e di puro *clickbaiting*.
MIS-INFORMAZIONE ONLINE	Categoria di contenuti informativi divulgati su Internet non veritieri o riportati in modo inaccurato, suscettibili di essere recepiti come reali, ma non creati con un intento doloso.
MALA-INFORMAZIONE ONLINE	Categoria di contenuti informativi fondati su fatti reali (anche a carattere privato) divulgati su Internet e contestualizzati in modo da poter essere anche virali e veicolare un messaggio con il preciso intento di danneggiare una persona, un'organizzazione o un Paese, o affermare/screditare una tesi.
DISINFORMAZIONE ONLINE	Categoria di contenuti informativi, anche sponsorizzati, artatamente creati in modo da risultare verosimili, contraddistinti non solo dalla falsità dei fatti, ma anche dalla loro contagiosità, nonché dall'intento doloso di pubblicazione e diffusione. Il contenuto viene costruito attorno a un messaggio con la precisa intenzione di danneggiare una persona, un'organizzazione o un Paese, o affermare/screditare una tesi, ingannando il pubblico. La diffusione dolosa di questi contenuti informativi può avvenire per finalità politico/ideologiche o per motivi economici (attrazione del maggior numero possibile di *click* e monetizzazione attraverso la raccolta pubblicitaria). Tali contenuti vengono diffusi in modo massivo attraverso le piattaforme *online*. Gli stessi possono essere di vario formato (testo, audio, video, ecc.) e riguardare diversi tipi di argomenti e tematiche, inclusi quelli di specifico interesse istituzionale per l'Autorità, ossia rientranti nell'ambito delle cd. *hard news* (a titolo esemplificativo: politica, cronaca, attualità, economia, scienza, sanità, ambiente e territorio, governo e pubblica amministrazione, ecc.).

In sostanza, anche dalle definizioni che precedono emerge come gli elementi salienti da considerare al fine di operare una classificazione delle varie distorsioni dell'informazione *online* siano riconducibili alle fasi di produzione dei contenuti informativi (falsità dei contenuti; contagiosità degli stessi; intento doloso sottostante alla loro creazione; motivazione politico/ideologica o economica di chi li crea per poi diffonderli), diffusione degli stessi (in maniera massiva) e impatto per il pluralismo informativo (generazione di effetti sulla formazione dell'opinione pubblica), ossia:

A. <u>PRODUZIONE DEI CONTENUTI INFORMATIVI</u>

- **Falsità dei contenuti (*componente oggettiva*)**: diffusione di contenuti falsi, infondati, manipolati o riportati in maniera non veritiera, creati ad arte in modo da risultare verosimili nel contesto mediatico.
- **Contagiosità (*componente oggettiva*)**: attitudine dei contenuti informativi, in base all'argomento trattato, dibattuto e di particolare interesse, e al linguaggio utilizzato, a trasferire stati emotivi e percezioni tra gli utenti, ovvero a condizionare il comportamento dei riceventi (agentività). In tal senso, sulla base della letteratura scientifica più recente[21], possono essere considerati indicatori di contagiosità: il livello di polarizzazione, anche endogena, dell'argomento trattato (un argomento può essere considerato polarizzante quando divisivo, ossia in grado di creare o accentuare la separazione degli individui in gruppi distinti)[22] e il taglio (positivo/negativo, favorevole/sfavorevole) conferito al contenuto informativo.

[21] Cfr. Del Vicario, M., Quattrociocchi, W., Scala, A., & Zollo, F. (2018), "Polarization and Fake News: Early Warning of Potential Misinformation Targets", *arXiv preprint arXiv:1802.01400*

[22] Più in generale, per polarizzazione ideologica si intende il risultato individuale di quel processo sociale di separazione e frammentazione della popolazione in gruppi distinti, separati e non comunicanti tra loro su tematiche divisive. Cfr. Sunstein, C. R. (2002), "The Law of Group Polarization", *Journal of Political Philosophy*, 10(2), pp. 175–195; Idem (2017), *#Republic. Divided Democracy in the Age of Social Media*, Princeton University Press. Il grado di polarizzazione di un contenuto può essere chiaramente associato alla sua capacità di generare fenomeni di *echo chamber* o *confirmation bias* in specifici gruppi di utenti (cfr. Quattrociocchi, W., Scala, A. & Sunstein, C. (2016), "Echo Chambers on Facebook", *working paper*).

- **Intento doloso (*componente soggettiva*)**: diffusione di contenuti falsi, costruiti attorno ad un messaggio con il preciso scopo di ingannare il pubblico per arrecare intenzionalmente danno a un soggetto (ad esempio: una persona fisica, un gruppo di persone, un'organizzazione o un'azienda), o affermare/screditare una tesi.
- **Motivazione politico/ideologica o economica (*componente soggettiva*)**: finalità di affermazione di una determinata ideologia o di sostegno ad un determinato punto di vista/opinione politica o di massimizzazione dei profitti attraverso la vendita di inserzioni pubblicitarie.

B. DIFFUSIONE DEI CONTENUTI INFORMATIVI

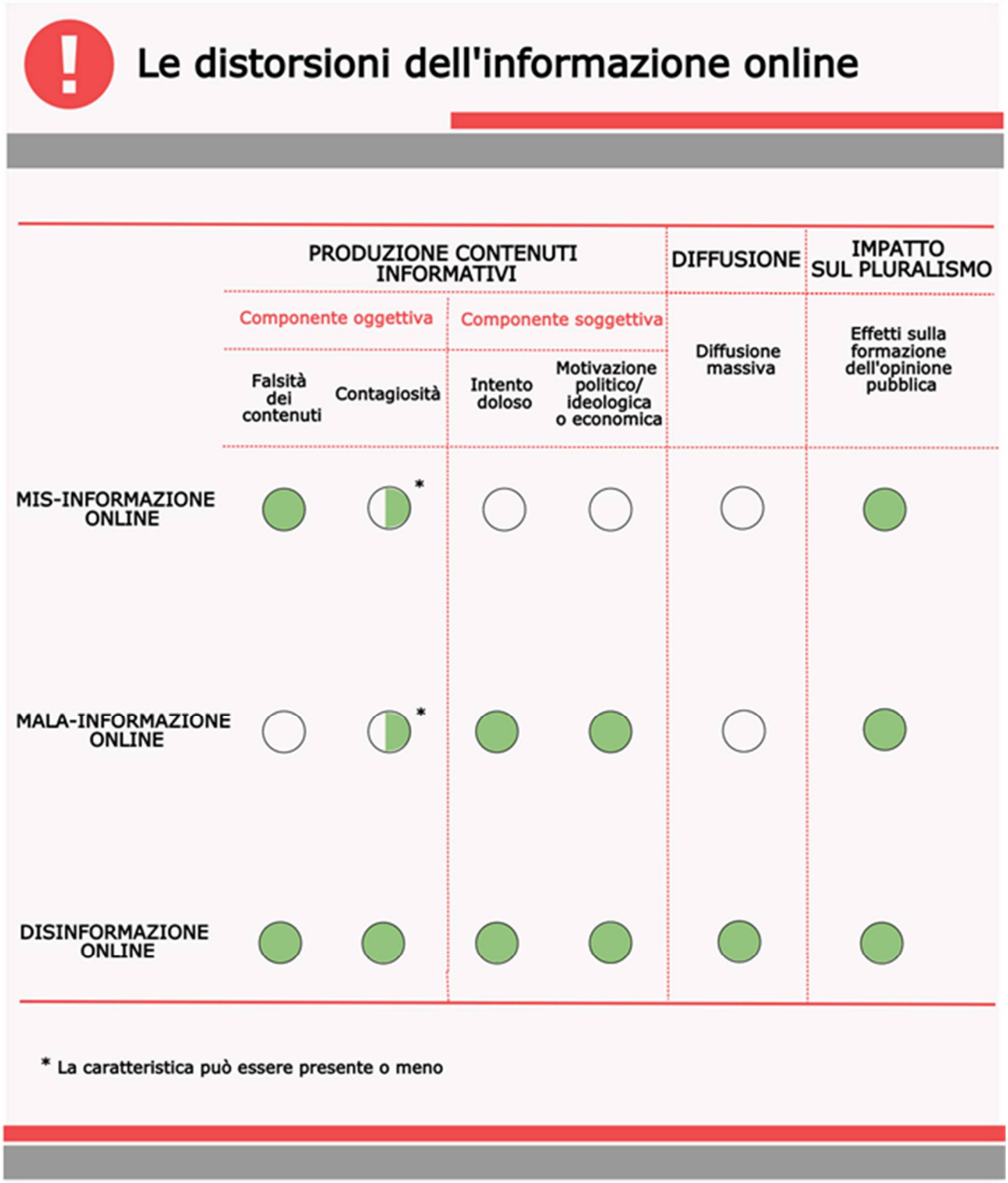
- **Diffusione massiva**: diffusione per mezzo delle piattaforme *online* di contenuti falsi, fuorvianti, ecc., anche attraverso *spamming* o tecniche e sistemi automatici (es. *bot*), atte a rendere virale la propagazione.

C. IMPATTO PER IL PLURALISMO INFORMATIVO

- **Effetti sulla formazione dell'opinione pubblica e sull'agenda politica**: contenuti informativi atti ad incidere sulla formazione dell'opinione pubblica e sulle priorità del dibattito socio-politico, riguardando in particolare tematiche quali politica, cronaca e attualità, scienza e salute, governo e pubblica amministrazione, territorio e ambiente, economia e finanza.

Nel dettaglio, la presenza di tutti i sei elementi sopra elencati identifica disturbi dell'informazione ascrivibili ai casi di disinformazione *online* (**Figura 1**). Diversamente, nei casi di mis-informazione *online* e mala-informazione *online*, i disturbi informativi sono caratterizzati dalla presenza di una combinazione dei predetti elementi, per cui, rispetto alla disinformazione *online*, possono mancare la falsità stessa dei contenuti, l'intento doloso, o la diffusione in maniera massiva. In merito alla potenziale contagiosità dei contenuti, vale rilevare come la stessa possa essere presente o meno nei disturbi di mis-informazione e mala-informazione, a seconda che si tratti o meno di contenuti inerenti a un argomento dibattuto e polarizzante:

Figura 1 – Le distorsioni dell'informazione *online*



In conclusione, i disturbi dell'informazione *online* riconducibili ai fenomeni di mis-informazione, mala-informazione e disinformazione come sopra definiti, in quanto atti ad incidere sulla formazione dell'opinione pubblica, costituiscono tutti materia di studio del Tavolo Tecnico.

3. La filiera dei contenuti *fake online*

Nonostante le distorsioni dell'informazione siano antiche quanto l'umanità, nell'era digitale esistono delle condizioni di contesto peculiari che rendono la disinformazione – in particolar modo – un fenomeno tendenzialmente **pervasivo**, che si dimostra più **efficace** e più **difficile da individuare** rispetto al passato, poiché cambiano gli strumenti di diffusione, la sua velocità di propagazione, la capacità di comprendere il fenomeno da parte di coloro che passivamente ne diventano amplificatori, nonché la consapevolezza dei meccanismi cognitivi sottostanti .

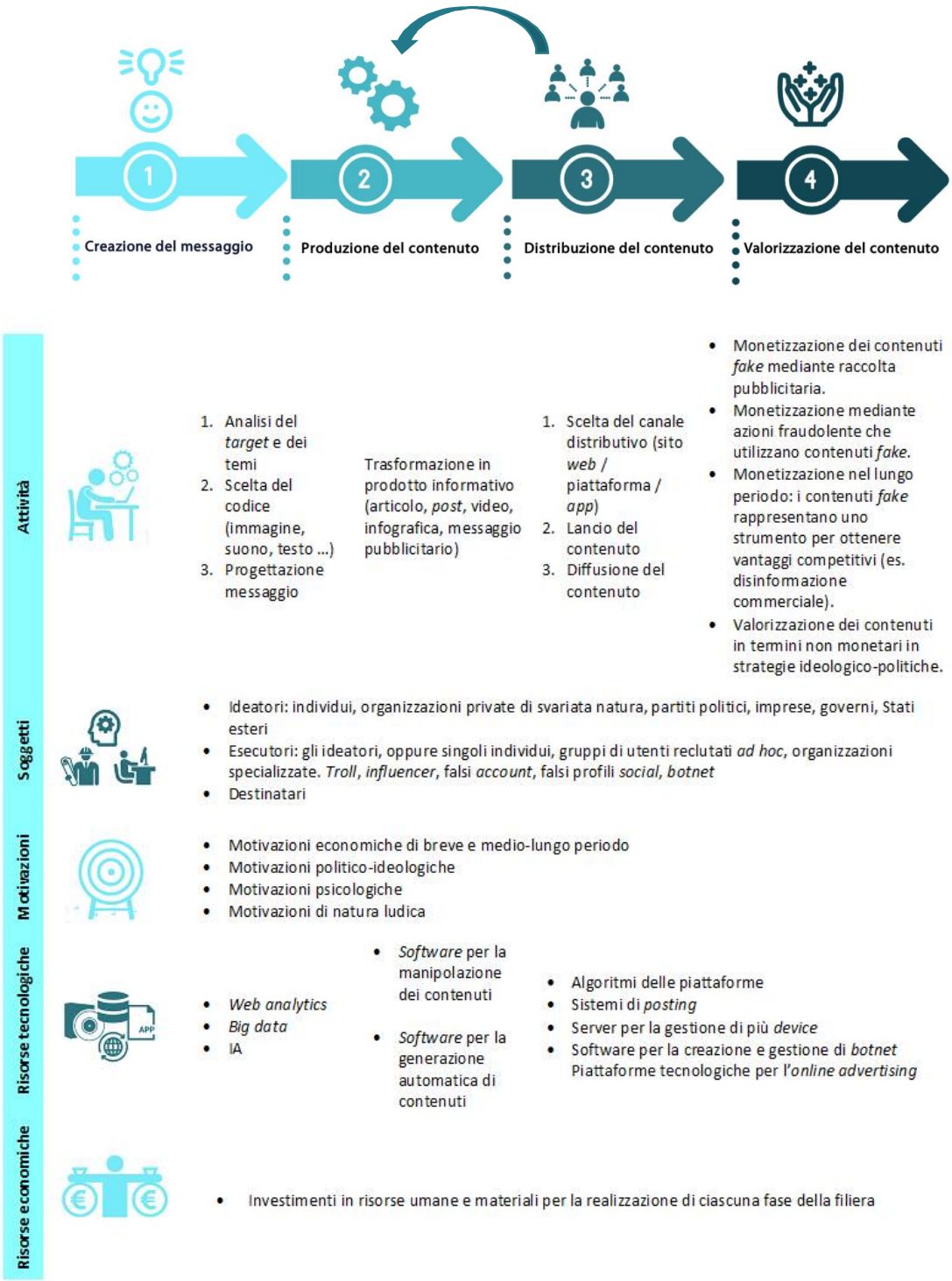
Secondo un numero crescente di studi empirici, gli **elementi di novità** investono molteplici aspetti, tra cui le tecnologie, le caratteristiche economiche della disinformazione, le dinamiche sociali. In particolare, l'evoluzione tecnologica rende più agevole la creazione, produzione e distribuzione dei contenuti *fake*, soprattutto mediante internet; i costi associati, tanto di produzione quanto di distribuzione, si sono molto ridotti così da diminuire i costi di entrata nel "mercato dei contenuti *fake*" e consentire l'ingresso di una moltitudine variegata di soggetti produttori; dal punto di vista degli utenti destinatari dei contenuti si registra una certa sfiducia nei riguardi dei mezzi di informazione tradizionali e al contempo una maggiore propensione a credere in ciò che circola sul *web*, risultato di una combinazione tra modalità di fruizione tipiche del mondo *online*, caratteristiche innate degli individui e contesto storico (cfr. *infra*).

Questi fattori incidono sugli effetti prodotti dalla disinformazione *online*, e sulle modalità con cui essa si esplica. Nello specifico, gli effetti prodotti sull'opinione pubblica, data la viralità con cui si diffondono i contenuti *fake online* e la loro capacità polarizzante, sono connotati da emozionalità, **radicamento**, **capillarità** nella **diffusione** e **persistenza** nel tempo.

La maniera con cui si manifesta la disinformazione *online*, invece, è caratterizzata dall'emergere di un preciso ***modus operandi***, che può appartenere anche al singolo individuo, più o meno consapevole, e che potenzialmente è in grado di divenire molto comune e di abbassare il livello di attenzione critica nei riguardi dei contenuti circolanti *online*, tanto da tramutarsi piuttosto in un *habitus animi.*

Tale *modus operandi* può assumere una forte strutturazione. In tal senso, i disturbi dell'informazione *online*, intesi nell'accezione ampia della relativa fenomenologia (cfr. *infra*), presuppongono tutti una sequenza ordinata di azioni che tipicamente sono effettuate per rendere disponibile *online* un contenuto *fake*. Questa successione di attività può configurarsi come una vera e propria **filiera dei contenuti *fake*** (**Figura 2**), più o meno strutturata a seconda del soggetto che promuove l'iniziativa e a seconda delle sue motivazioni. Infatti, i fenomeni critici dell'informazione *online* possono nascere e diffondersi anche in assenza di ragioni specifiche di tipo economico, ideologico o politico. In alcuni casi può trattarsi di iniziative di singoli individui che, senza essere sempre pienamente coscienti degli effetti delle proprie azioni sulla formazione dell'opinione pubblica, creano e distribuiscono *online* contenuti *fake*; in tal caso, la filiera rappresenta piuttosto un modo per schematizzare una serie di attività ordinate logicamente, prive di una precisa strategia e organizzazione a supporto.

Figura 2 – La filiera dei contenuti *fake online*



3.1 LE FASI PRINCIPALI

In maniera semplificata si possono distinguere quattro fasi principali che costituiscono la filiera dei contenuti *fake online*: una prima fase di creazione del messaggio, una seconda fase di produzione del contenuto in cui il messaggio viene incorporato ovvero trasformato in un prodotto informativo; una terza fase di distribuzione del contenuto; una quarta in cui i contenuti *fake* sono infine valorizzati, monetariamente o non monetariamente.

Nella fase I di creazione, viene elaborato il messaggio da veicolare mediante il contenuto *fake*; questo assume caratteristiche diverse in ragione dell'obiettivo degli ideatori e a seconda del *target* cui è destinata la strategia di disinformazione. In generale, per risultare efficace, il messaggio deve essere costruito in maniera tale da raggiungere l'*audience* e altresì attivarla, coinvolgendola anche nella diffusione ulteriore del contenuto. A questo fine, rilevano alcuni elementi per la preparazione degli "*asset*"[23], necessari sia alla creazione del messaggio, sia alla produzione del contenuto informativo:

- Il profilo degli utenti e il *target* di riferimento: l'accuratezza della profilazione degli utenti *online* permette di predisporre messaggi e contenuti *fake* più efficaci rispetto all'*audience target*; in fase distributiva consente una diffusione mirata dei contenuti incrementando l'efficacia della campagna di disinformazione, e, inoltre, aumenta il valore degli spazi pubblicitari da vendere agli inserzionisti pubblicitari con cui finanziare siti e piattaforme che ospitano contenuti *fake*[24].
- L'analisi del contesto e la scelta dei temi trattati nel messaggio: presuppone un'analisi dei temi "caldi" che circolano sul *web* all'interno di determinate comunità, così da individuare quelli su cui è più probabile catalizzare l'attenzione e favorire così la diffusione del contenuto *fake*. Tali *hot topic* dipendono dal contesto economico, sociale, politico, e più in generale culturale, in cui il *target* di riferimento si colloca. Inoltre, essi variano nel tempo, anche piuttosto rapidamente, per cui si rende utile un monitoraggio continuo, soprattutto se la campagna di disinformazione si pone degli obiettivi di medio-lungo termine. In particolare, risulta importante, come dimostrato da recenti ricerche scientifiche, porre l'attenzione su temi divisivi in grado di polarizzare l'utenza[25].
- Il modo in cui gli individui elaborano le informazioni: tenuto conto dell'abbondanza delle informazioni disponibili *online*, il messaggio dovrebbe sfruttare le euristiche cui si affida il sistema cognitivo umano per orientarsi nella complessità del mondo *online*, ricca di *input*, fonti di informazioni e stimoli. In particolare, il contenuto dovrebbe agire sui *bias* cognitivi degli individui e quindi soddisfare non solo i bisogni di informazione, ma soprattutto le aspettative dei destinatari in termini di corrispondenza rispetto alle proprie convinzioni, di coinvolgimento emotivo e di condivisione della visione del mondo. I contenuti dei messaggi facenti parte di campagne di disinformazione dovrebbero quindi soddisfare (e sfruttare) le tendenze degli utenti a leggere e condividere informazioni consonanti con il proprio punto di vista (*confirmation bias*) e di potenziale interesse all'interno delle *echo chambers* in cui si struttura un'opinione pubblica sempre più polarizzata. Tali messaggi dovrebbero inoltre essere "confezionati" in modalità tali da poter essere fruiti in maniera incidentale in contesti informativi (es. piattaforme *online*) caratterizzati da sovrabbondanza di contenuti provenienti da fonti disparate (amici, conoscenti, gruppi, *post* sponsorizzati, editori, ecc.)[26]

In particolare, questi aspetti (*target*, contesto e processi cognitivi) incidono sulla scelta del codice della comunicazione[27], quindi sia sul linguaggio per la composizione del messaggio (parole, immagini, suoni), sia sulla logica narrativa, ossia il modo in cui il messaggio è inquadrato.

---

[23] Sull'ecosistema delle *fake news* si veda anche l'attività di ricostruzione effettuata da PHD Italia, http://www.phdmedia.com/italy/ecosistema-fake-news/.
[24] Cfr. EDPS (2018), *Opinion on online manipulation and personal data*.
[25] Cfr. Del Vicario M., Quattrociocchi W., Scala A., Zollo F. , *Polarization and Fake News, op. cit.*
[26] Cfr. AGCOM, *Rapporto sul consumo di informazione*, op. cit..
[27] Cfr. Wardle, C. & Derakhshan, H. (2017), *Information disorder: Toward an interdisciplinary framework for research and policy making*, Council of Europe.

La considerazione congiunta dei tre elementi, dunque, porta alla definizione di messaggi con specifiche caratteristiche e una precisa struttura (cfr. *infra*, par. 5 e 6)[28].

Nella fase II di produzione del contenuto, il messaggio viene trasformato in un prodotto informativo, che può assumere la forma di un testo (ad esempio, un *post* o un articolo), di un'immagine, di un video, oppure una combinazione di questi elementi. In questa fase si affina ulteriormente il codice comunicativo, tanto che si possono rinvenire diversi "generi" di contenuti *fake* che possono essere osservati da molteplici punti di vista, come mostrano i numerosi sforzi ad oggi condotti per una loro analisi e classificazione (cfr. *supra*, par. 2). Ad esempio, in base al grado di "manipolazione" del messaggio, idealmente, si va dal contenuto completamente falso, quindi fabbricato *ex novo*, a quello basato su un'informazione originaria vera ma manipolata. La manipolazione può riguardare il messaggio contenuto (contenuti manipolati), le informazioni di contesto (false contestualizzazioni), il titolo, le immagini, o le didascalie (false connessioni), la fonte; si arriva fino ai contenuti parodistici e satirici oppure fuorvianti, in cui la manipolazione può riguardare l'inquadramento del messaggio. In particolare, una pratica comune nel mondo *online*, e che si osserva anche nell'ambito dei contenuti *fake*, è l'adozione di un codice comunicativo al confine tra comunicazione commerciale e comunicazione informativa. È il caso ad esempio dei contenuti di disinformazione sponsorizzati e dei *dark ads* che sfruttano le tecnologie che consentono una profilazione e personalizzazione dei messaggi veicolati nell'ambito del sistema di diffusione/compravendita della pubblicità *online*, e al contempo propongono agli utenti forme di comunicazione innovativa (sviluppate sempre in ambito pubblicitario come la *native advertising*), attraverso la narrazione di una "storia", maggiormente attraente e credibile rispetto alla comunicazione pubblicitaria (cfr. *infra*, par. 5)[29].

Nella fase III di distribuzione, il contenuto *fake* viene pubblicato *online* e reso quindi disponibile. In questa fase si decide il canale distributivo (o i canali distributivi) e il contesto mediatico in cui il contenuto si inserisce. Il primo può essere tipicamente un sito *web* (ad esempio il sito di un editore *online*, una piattaforma *social*, un *blog*, un forum, *etc*.) oppure un'applicazione (ad esempio di *instant messaging*). In genere, questi strumenti sono un canale preferenziale, principalmente perché rendono più facile mantenere l'anonimato, si affidano a meccanismi *peer to peer*, percepiti come più credibili dai destinatari, e costituiscono una modalità di distribuzione con costo praticamente pari a zero. Il contesto mediatico rappresenta, invece, la rete di contenuti (testi, immagini, suoni) che circola sui diversi media *online* e *offline*; in particolare, nella fase distributiva si definisce il contesto mediatico che si colloca attorno e/o insieme con i contenuti *fake* e che è importante soprattutto per conferire attendibilità al messaggio.

Nell'ultima fase, la IV, i contenuti *fake* vengono valorizzati, ossia possono produrre guadagni monetari più o meno immediati attraverso l'adozione di una serie di strategie commerciali (cfr. *infra*, par. 5), oppure possono raggiungere gli scopi desiderati senza generare necessariamente un flusso di entrate monetarie, poiché rispondono ad altre motivazioni, come si vedrà a breve e, poi, più diffusamente nella trattazione delle strategie di disinformazione a scopo ideologico-politico (cfr. *infra*, par. 6).

In particolare, per ciò che riguarda i ricavi, soprattutto in strategie commerciali di breve-medio periodo, esistono due fonti principali di remunerazione per i produttori: le risorse pubblicitarie e, in alcuni casi, il contributo diretto degli utenti ottenuto con azioni fraudolente. Nell'ambito di strategie di più lungo periodo, ritorni economici possono derivare ad esempio da campagne di disinformazione che, danneggiando l'immagine e la reputazione di un'impresa concorrente, mirano a sottrarre quote di mercato. Infine, si riscontra la presenza di strategie ibride, in cui coesistono finalità politico-ideologiche e finalità di natura economica, che possono produrre un'alterazione degli assetti di un mercato, tale da determinare un rafforzamento della

[28] In linea generale, gli studi condotti sul tema segnalano che, per catturare l'attenzione del destinatario, è importante la capacità del messaggio di indurre una reazione emotiva, la presenza di una componente visiva rilevante che permette al cervello umano di elaborare rapidamente l'informazione, una storia convincente, la ripetitività del messaggio. Per rendere il messaggio credibile, in particolare, questo deve risultare riconoscibile e familiare; deve essere ritenuto credibile da altri individui; deve essere veicolato da molteplici siti (ad esempio attraverso condivisioni, *like*, commenti, *re-tweet*); deve presentarsi nella forma e nel contesto che un individuo si aspetta; deve confermare le convinzioni possedute dal destinatario; deve avere intento persuasivo.

[29] Su questi temi cfr. anche CRS, Centro per la Riforma dello Stato, NEXA Center for Internet & Society, Fondazione P&R (2018), *Persuasori social. Trasparenza e democrazia nelle campagne elettorali digitali*.

posizione economica di alcuni soggetti d'impresa a scapito di altri, generando così dei vantaggi economici per gli ideatori.

3.2 I SOGGETTI

Una caratteristica legata alla natura decentralizzata della rete internet e, quindi, comune a tutti i disturbi dell'informazione *online*, è l'ubiquità dei soggetti che intervengono a vario titolo nelle diverse fasi della filiera e la difficoltà di poter definire la loro collocazione geografica. A questa complicazione si aggiunge la presenza di un numero elevato di soggetti, anche molto diversi tra loro, che operano nelle quattro fasi: principalmente si distinguono gli ideatori del contenuto o di un'intera campagna, gli esecutori delle diverse attività lungo la filiera, gli stessi soggetti destinatari che possono rilanciare il contenuto, oppure anche modificarlo e trasformarlo nuovamente ottenendo così un altro, più o meno diverso, contenuto *fake*.

In particolare, **gli ideatori** possono essere attivi lungo tutte le fasi: può trattarsi di singoli individui, di imprese editoriali e non, di organizzazioni con finalità svariate (culturali, ideologiche, politiche, criminali), di servizi di *intelligence*, di governi, di Stati.

Nell'ambito degli **esecutori** si collocano, invece, coloro che contribuiscono alla creazione e produzione del contenuto *fake*, talvolta coincidenti con gli stessi ideatori dell'iniziativa. Sono singoli individui, gruppi di utenti reclutati *ad hoc*, oppure anche organizzazioni vere e proprie che elaborano i messaggi e i contenuti informativi *fake*. In quest'ultimo caso alcuni studi e rapporti evidenziano l'esistenza di organizzazioni specializzate nella progettazione e implementazione di campagne di disinformazione[30].

Nella fase di distribuzione, in particolare, i soggetti che perseguono strategie di disinformazione – singoli o gruppi di individui oppure organizzazioni più o meno strutturate – possono agire con l'ausilio di meccanismi automatici come i *bot*, che consentono la pubblicazione e distribuzione dei contenuti *fake* o attraverso una molteplicità di *account* falsi o falsi profili *social*. Inoltre, a questa fase della filiera partecipano altresì gli editori e le piattaforme *online*, nonché gli stessi utenti destinatari dei contenuti che, anche inconsapevolmente, rilanciano i contenuti *fake* e ne favoriscono la diffusione.

Al riguardo, sebbene il processo di diffusione, oggetto di molti studi recenti, sia ancora da esplorare attentamente[31], tuttavia si può osservare come esistano dei soggetti chiave soprattutto per il lancio del contenuto *fake online* e per la sua promozione, in particolare *troll*, *influencer*, falsi profili *social* e falsi *account* e i cosiddetti *fake tank*[32]. In generale, i meccanismi di interazione delle piattaforme *social* consentono a qualunque individuo di farsi parte attiva nella promozione di un contenuto *fake online* trasferendo altresì stati emotivi e così contribuendo ai processi di viralizzazione.

3.3 LE MOTIVAZIONI

Se si considera il complesso delle distorsioni dell'informazione *online*, le motivazioni che spingono gli ideatori dei contenuti *fake* sono numerose, di natura svariata e spesso non univoche.

---

[30] Cfr. Wardle, C. & Derakhshan, H. , *Information dìsorder, op. cit.;* Trend Micro, (2017), *The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public.;* Lo studio di Trend Micro, che ha analizzato l'offerta di servizi di *fake news*, riporta cifre piuttosto precise per la realizzazione di una campagna di disinformazione, secondo un vero e proprio tariffario.

[31] Per una breve rassegna si veda l'*Interim report Big Data* prodotto dall'Autorità cit.; Martens B., Aguiar L., Gomez-Herrera E., Mueller-Langer F. (2018), *The digital transformation of news media and the rise of disinformation and fake news*, JCR Digital Economy Working Paper 2018-02; Zannettou S., Sirivianos M., Blacburn J., Kourtellis N. (2018), *The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and various Other Shenanigans.*

[32] Dall'inchiesta del Parlamento britannico emerge come esistano dei *think tank* completamente fittizi, nati attorno a un progetto preciso (anche di disinformazione) spesso facenti capo a finanziatori occulti, in altri casi questi sono organizzazioni più strutturate che operano su molteplici tematiche come entità indipendenti, ma che in taluni casi agiscono come *lobby* nascoste. Cfr. http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/47967.html.

Si possono distinguere essenzialmente:

- **Motivazioni economiche**, di breve-medio periodo (finalizzate al recupero di risorse economiche dalla pubblicità e/o da azioni fraudolente) e di lungo periodo (finalizzate al recupero di risorse economiche attraverso strategie varie che puntano, per esempio, al discredito di imprese concorrenti o a influenzarne il valore sui mercati finanziari).
- **Motivazioni ideologico-politiche**.
- **Motivazioni psicologiche**, legate alla volontà del singolo di affermarsi nelle comunità *online*.
- **Motivazioni ludico-satiriche**.

Si può osservare come le motivazioni di ordine economico e politico-ideologiche spesso si intrecciano e coesistono all'interno di un'unica strategia, così che non è sempre agevole distinguere gli obiettivi degli ideatori di una campagna di disinformazione.

3.4 LE RISORSE TECNOLOGICHE

La tecnologia legata al funzionamento del *web* e ai servizi che circolano su di esso rende più agevole tutte le attività che ruotano attorno ai disturbi dell'informazione *online*, ad esempio la progettazione dei messaggi, la distribuzione dei contenuti *fake*, nonché la ri-produzione effettuata dagli stessi destinatari nel momento in cui modificano e rimettono in circolazione il contenuto originariamente ricevuto.

In particolare, le **risorse tecnologiche** svolgono una funzione complessa e rilevante lungo tutta la filiera dei contenuti *fake online*; in tal senso, esse hanno un duplice ruolo nell'ambito dei disturbi dell'informazione *online* che può essere definito di tipo diretto o indiretto. Da un lato, infatti, la tecnologia rende disponibili **strumenti** che possono essere utilizzati direttamente dai soggetti della filiera che producono e distribuiscono contenuti di disinformazione; dall'altro lato, le tecnologie incidono indirettamente sul fenomeno, nella misura in cui creano un **ambiente favorevole** alla diffusione della disinformazione.

In generale, si tratta delle medesime tecnologie utilizzate per gestire contenuti *online* non *fake* e perfino delle stesse tecnologie che potrebbero coadiuvare il contrasto ai fenomeni di disinformazione *online*; da questo punto di vista, di per sé, le tecnologie si rivelano "neutrali", sebbene alcuni servizi forniti nel "mercato delle *fake news*" siano su queste basati[33].

Per ciò che riguarda gli **strumenti** tecnologici adoperati dai soggetti della filiera, all'interno di un vasto e complicato panorama, nella fase di creazione, i sistemi di *web analytics*, finalizzati al tracciamento dell'attività di navigazione degli utenti, si sono dotati nel tempo di metodi sempre più evoluti (cfr. *infra*) che consentono di rilevare le attività svolte in rete, così da ottenere o inferire informazioni circa le preferenze, i gusti, i comportamenti di acquisto (sia *online* che *offline*) e gli aspetti psicologici degli individui[34]. Ciò, unito alla disponibilità di *big data* sugli utenti e insieme con l'applicazione di sistemi di intelligenza artificiale, permette di progettare i messaggi in maniera efficace, in ragione dell'*audience target* da raggiungere e dell'analisi degli *hot topic* che circolano sul *web* nell'ambito delle comunità di utenti obiettivo della campagna[35].

Nella creazione e produzione intervengono, inoltre, *software* che permettono non solo la manipolazione agevole dei contenuti (di video, immagini, suoni e testi) ma anche la generazione automatica degli stessi.

---

[33] Cfr. Brundage M. (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*; Ghosh D., Scott B. (2018), *#Digitaldeceit – The Technologies Behind Precision Propaganda on the Internet*, Shorenstein Center on Media, Politics and Public Policy, Harvard Kennedy School.

[34] L'accesso, tramite internet, a una moltitudine di dati sugli utenti, unita al miglioramento dei modelli predittivi e ai progressi della psicometria, rende possibile elaborare anche il profilo psicologico e la personalità degli utenti. Cfr. al riguardo una delle prime applicazioni sui dati provenienti da *social media*, Kosinski M., Stillwell D., Graepel T. (2013), *Private traits and attributes are predictable from digital records of human behavior*, PNAS, 110 (15) 5802-5805.

[35] Cfr. Agcom (2018), *Big data, Interim report, op. cit.*

Nella distribuzione e diffusione dei contenuti *fake*, poi, agiscono ulteriori elementi tecnologici, tra cui i sistemi di *posting* sui *social network*, che possono essere forniti anche da soggetti che vendono *click*, *like*, *views*, o *follower*; i *software* per la creazione e gestione dei *bot*; i *server* che permettono la gestione contemporanea di una molteplicità di *device*, dunque di profili reali.

Nella fase di valorizzazione, infine, le tecnologie, gli algoritmi, i protocolli e le piattaforme per la gestione delle transazioni pubblicitarie mediante sistemi automatici (*ad server*, *demand side platform* - DSP e *sell side platform* - SSP), nonché le piattaforme per la gestione di programmi di affiliazione, oltre a quelle di *data management* (che raccolgono, ospitano e organizzano i dati sugli utenti provenienti da varie fonti anche di tipo *off line* finalizzate alla profilazione delle campagne di *digital advertising*) svolgono un ruolo importante nella monetizzazione dei contenuti *fake* mediante introiti pubblicitari (cfr. *infra*, par 5.1).

Con riferimento al **ruolo indiretto della tecnologia**, in particolare nel processo di diffusione dei contenuti di disinformazione *online*, giocano un ruolo chiave, di tipo incentivante, gli algoritmi delle piattaforme *online*, soprattutto quelli di *search* che definiscono il *ranking* dei contenuti mostrati nei risultati di ricerca[36] e quelli utilizzati dalle piattaforme *social*, che favoriscono sistemi di personalizzazione automatica dei contenuti visualizzati e permettono, inoltre, una molteplicità di azioni e reazioni da parte degli utenti[37]; questi meccanismi, da un lato, migliorano la capacità di profilare gli utenti (utile anche nella fase di creazione e produzione nonché nel monitoraggio della campagna di disinformazione) e, dall'altro lato, facilitano la diffusione dei contenuti *fake*[38].

3.5 LE RISORSE ECONOMICHE

Per ciò che riguarda le **risorse economiche da investire** per la creazione, produzione e distribuzione dei contenuti *fake online* e le prospettive di guadagno attese dalla diffusione degli stessi, per quanto non siano oggi disponibili dati quantitativi affidabili[39], si può comunque affermare che la struttura dei costi è tale per cui gli oneri di produzione sono piuttosto bassi e quelli di distribuzione tendono a zero. Questa caratteristica rappresenta un incentivo per gli ideatori di contenuti *fake* che, unita alla possibilità di generare ricavi dalla diffusione *online* di tali contenuti, rende sostenibili anche veri e propri **modelli di *business*** basati sugli stessi (cfr. *infra*, par.5).

La circostanza che il costo di produzione e distribuzione dei contenuti sia ridotto – complice non solo la struttura di internet ma anche l'evoluzione tecnologica – determina un abbassamento delle barriere all'entrata dei "mercati dei contenuti *fake*", per cui vari soggetti (individui o organizzazioni) sono incentivati a entrare nelle diverse fasi della filiera, non solo come promotori delle campagne di disinformazione, ma anche come esecutori delle svariate attività connesse alla realizzazione delle strategie di disinformazione *online*.

[36] Come verrà illustrato successivamente, l'algoritmo di *search* può essere aggirato in maniera intenzionale, in modo da dominare i risultati di ricerca per alcune ore e prima che il motore di ricerca sia in grado di correggere la distorsione (si parla di strategie di *search engine optimization* - SEO). Ad esempio i *software* SMMS (*social media management service*) per la gestione delle campagne pubblicitarie su più *social media* (frutto di una combinazione tra algoritmi di *machine learning* e tecnologie per l'*online advertising*) rappresentano degli strumenti molto potenti per coloro che perseguono strategie di disinformazione *online* indipendentemente dalla finalità sottesa alle stesse (commerciale o ideologico-politica).

[37] Cfr. AGCOM (2018), *Big data*, *cit.*

[38] Ad esempio i *software* SMMS (*social media management service*) per la gestione delle campagne pubblicitarie su più *social media* (frutto di una combinazione tra algoritmi di *machine learning* e tecnologie per l'*online advertising*) rappresentano degli strumenti potenti per coloro che perseguono strategie di disinformazione *online* indipendentemente dalla finalità sottesa alle stesse (commerciale o ideologico-politica).

[39] Alcuni numeri sono riportati da notizie di stampa nazionali e internazionali (ad esempio si stimano guadagni giornalieri che oscillano attorno ai 1.000 euro giornalieri, mensilmente si arriva a diverse decine di migliaia di euro); con riferimento agli USA in un'intervista Paul Horner, creatore di contenuti *fake*, nel 2016 ha dichiarato al Washington Post di guadagnare 10.000 dollari al mese da GoogleAdSense, con picchi di 10.000 dollari al giorno per le storie più virali. Un'altra inchiesta pubblicata da NBC riporta che l'affare dei ragazzi macedoni coinvolti nell'ultima campagna elettorale USA abbia fruttato circa 5.000 dollari al mese.

4. Una classificazione delle strategie di disinformazione *online*

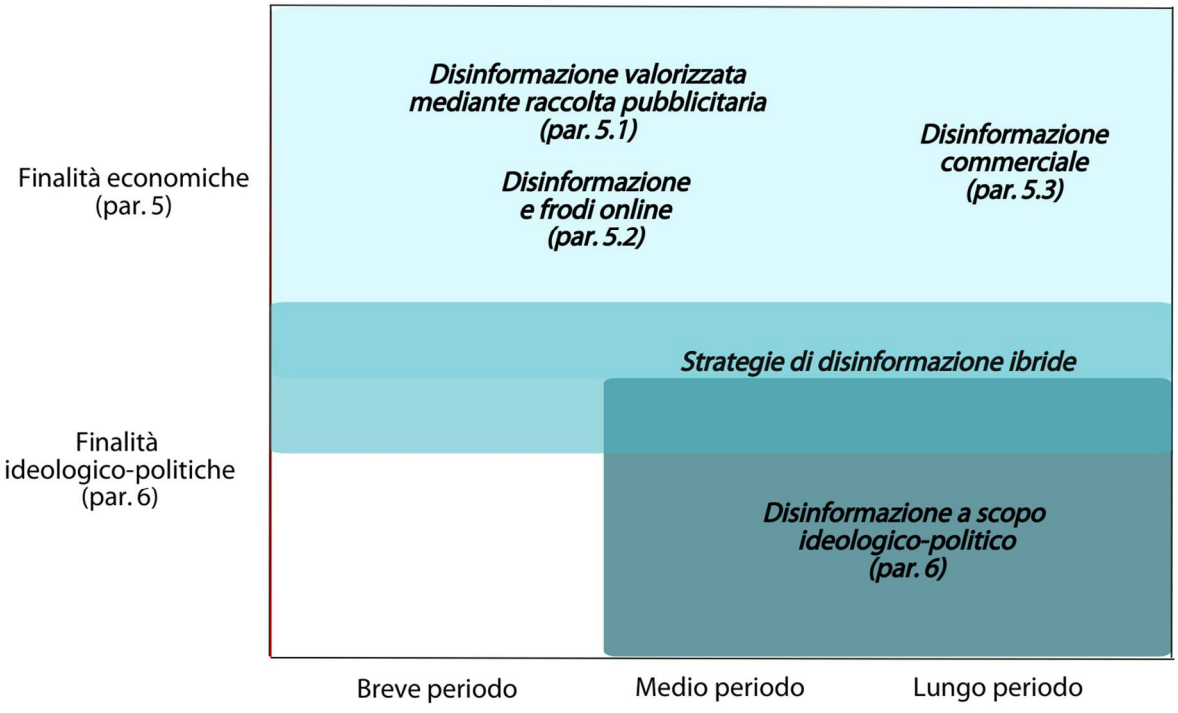
Le attività di creazione, produzione e distribuzione di contenuti *fake online* si configurano come una filiera vera e propria, in particolare quando vengono attuate nell'ambito di strategie di disinformazione.

Si può parlare di strategie di disinformazione *online* nei casi in cui gli ideatori dei contenuti non sono individui singoli ma organizzazioni stabili, o anche temporanee, accomunate da interessi specifici, mosse da precisi obiettivi di natura economica o politico-ideologica, con dotazioni finanziarie, tecnologiche e organizzative e *target* di destinatari ben individuati. Inoltre, tali strategie si manifestano in genere con una serie di azioni di disinformazione che danno luogo non a singoli sporadici episodi, ma a una serie di pubblicazioni e/o ri-pubblicazioni di contenuti *fake*; si tratta di campagne di disinformazione, dunque, che possono avere durata variabile nel tempo. In tali casi, è evidente che i contenuti *fake* diffusi rientrano essenzialmente nell'ambito dei fenomeni di disinformazione, caratterizzati, *inter alia*, proprio dalla intenzionalità e dalla diffusione massiva, oltre che dalla capacità di produrre effetti sul pluralismo informativo (cfr. *infra*, par. 1).

Per schematizzare, in base alle motivazioni degli ideatori (cfr. *supra*, par. 3.3) è possibile distinguere strategie commerciali, che perseguono finalità economiche, e strategie a sfondo ideologico-politico, sebbene, come si vedrà, un tipo di motivazione non esclude l'altro essendovi, quindi, anche strategie ibride; in base al lasso temporale in cui vengono attuate, invece, si possono osservare strategie di breve-medio periodo e strategie di lungo periodo.

Dalla combinazione di queste due dimensioni (finalità e durata) emergono tipologie distinte di strategie di disinformazione (**Figura 3**) che presentano connotati diversi quanto a soggetti ideatori, caratteristiche dei contenuti di disinformazione, flussi economici, come emerge, peraltro, anche dai casi concretamente verificatisi.

Figura 3 – Strategie di disinformazione *online*



Nel seguito tali tipi saranno analizzati separatamente; in particolare nel paragrafo 5 saranno prese in esame le strategie commerciali aventi motivazioni economiche e verranno trattate distintamente le strategie di breve-

medio periodo, sia quelle basate su risorse pubblicitarie (par. 5.1), sia quelle connesse ad azioni truffaldine a danno degli utenti (par. 5.2), e le strategie di più lungo periodo che utilizzano la disinformazione commerciale (par. 5.3). Sempre con riferimento al lungo periodo, nel paragrafo 6 saranno approfondite, infine, le strategie di disinformazione *online* con motivazioni ideologico-politiche.

5. Le strategie commerciali e i modelli di *business* della disinformazione *online*

5.1 La disinformazione *online* che si finanzia attraverso la pubblicità

In questo paragrafo sono analizzate le strategie di disinformazione *online* che hanno come obiettivo la **massimizzazione dei profitti attraverso la vendita di spazi pubblicitari**. Si tratta di distorsioni dell'informazione contraddistinte non solo dalla falsità e dalla contagiosità dei contenuti (componente oggettiva), dall'intento doloso della loro pubblicazione, dalla diffusione massiva dei contenuti e dalla capacità di produrre rilevanti effetti sul pluralismo informativo (cfr. *supra*, par. 2), ma anche da una precisa motivazione economica sottostante e consistente nell'attrarre, attraverso il *framing* e la spinta emotiva, il maggior numero di utenti (cd *click baiting*) da valorizzare attraverso la raccolta pubblicitaria (componente soggettiva).

Ciò che differenzia questo tipo di strategie è l'utilizzazione di tecnologie, protocolli, piattaforme *online*, algoritmi e meccanismi automatici che assicurano il funzionamento del sistema di compravendita della pubblicità *online* e che sono utilizzati dai soggetti che attuano questo tipo di strategia di disinformazione *online*, con l'intento di incrementare il traffico nei propri siti così da poter rivendere agli inserzionisti i contatti pubblicitari ottenuti[40].

In altri termini, i soggetti che perseguono una strategia di disinformazione *online*, allo scopo di attirare pubblico e traffico presso i propri siti da valorizzare in termini di vendita di pubblicità, si avvalgono dei medesimi strumenti tecnologici elaborati per rendere più efficace ed efficiente la campagna pubblicitaria attraverso il *web* sfruttandone le relative opacità ed aree grigie.

Tenuto conto, pertanto, che dette strategie di disinformazione si sviluppano attorno ad un modello di *business* specifico, fondato appunto sulla pubblicità *online*, è utile prima di tutto richiamare brevemente il funzionamento del sistema di compravendita degli spazi pubblicitari *online* e, a seguire, ripercorrere i principali cambiamenti dello stesso sistema che spiegano come sia possibile rendere profittevole la diffusione *online* di contenuti di disinformazione, tanto da fare emergere strategie di disinformazione basate su tale modello di *business*.

La compravendita di spazi pubblicitari *online* si basa su meccanismi diversi in base alla tipologia di pubblicità (*search*, *social*, *display*, *email*), al canale di vendita (*diretto* mediante le proprie forze di vendita e *indiretto* ossia mediante intermediari di pubblicità: centri media, *ad network*) e alla modalità di contrattazione (*tradizionale*, che presuppone la negoziazione diretta fra le parti, oppure attraverso *modelli automatici* di compravendita). In particolare, la modalità di compravendita della pubblicità *online* che si serve di modelli automatici presuppone l'impiego di piattaforme tecnologiche automatizzate che mettono in contatto gli acquirenti (inserzionisti/centri media mediante la *demand side platform* - DSP) con i venditori (concessionarie/editori mediante la *sell side platform* - SSP), permettendo alle loro inserzioni di raggiungere gli utenti profilati che, sulla base delle informazioni analizzate, saranno verosimilmente in grado di attirare l'attenzione da parte proprio del *target* desiderato.

Come illustrato nella **Figura 4** (che riporta in modo semplificato il processo di compravendita di pubblicità *online* secondo il *programmatic*), dal lato della domanda si collocano gli inserzionisti di pubblicità *online* che

[40] Ghosh D.., Scott B., *#Digitaldeceit, op. cit.*

sono interessati al raggiungimento di un determinato *target* di utenza. Tali operatori, avvalendosi del canale diretto o di quello indiretto, procedono all'acquisto di spazi pubblicitari all'interno dei siti internet che in base ai dati raccolti consentono loro di veicolare il messaggio esattamente al profilo socio-demografico desiderato, corrispondendo un prezzo in funzione delle visualizzazioni, ovvero del numero di azioni svolte (*click*, compilazione di moduli, acquisto) o ancora del tempo speso nella navigazione. Dal lato dell'offerta, si riscontra la presenza di editori/*publisher* ossia di fornitori di contenuti e servizi *web* orizzontali (motori di ricerca, portali, *social network*, aggregatori di contenuti) e verticali (siti di informazione, di intrattenimento) che hanno delineato il proprio sito con l'intento di creare appositi spazi per accogliere uno o una combinazione di formati pubblicitari cui sono associati dei codici (*ad tag*) che consentono l'inserimento della pubblicità da parte dell'*ad server* (ossia del *software* per la gestione e distribuzione delle campagne pubblicitarie attraverso le pagine *web* e i relativi contenuti).

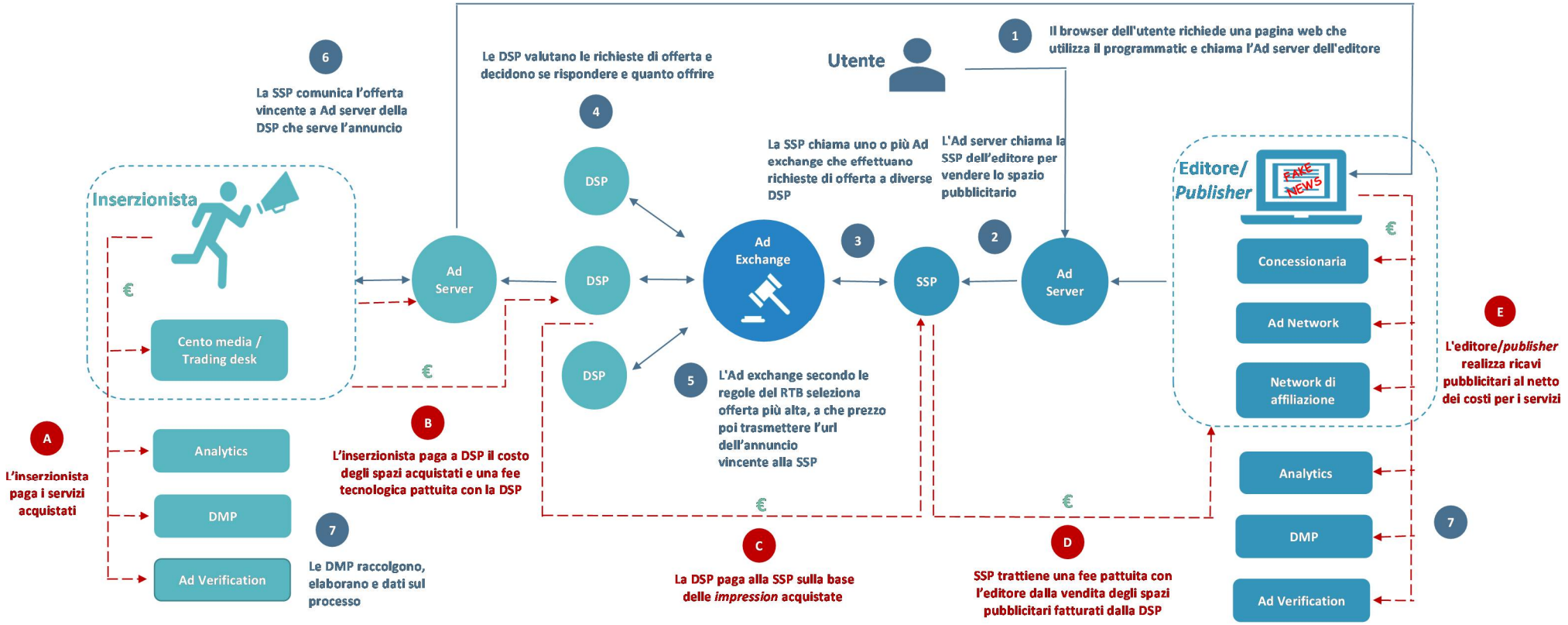
Gli editori/*publisher* (eventualmente attraverso le proprie concessionarie di pubblicità) mettono a disposizione il proprio inventario attraverso gli intermediari (*ad network* o altri intermediari come motori di ricerca, *social network*) ricevendo un corrispettivo dalle diverse piattaforme di intermediazione in funzione del numero delle *impression* distribuite o delle azioni svolte dall'utente (*click*, condivisione, acquisto, tempo speso nella visualizzazione).

La possibilità offerta dal mezzo di avvalersi di avanzati e innovativi strumenti tecnologici, che riducono l'incidenza dei costi di transazione mettendo in contatto un gran numero di inserzionisti con altrettanto numerosi fornitori di contenuti *web*, ha favorito la diffusione di numerosi attori attivi nell'intermediazione di pubblicità *online*. Gli intermediari hanno il compito di acquistare e gestire gli spazi pubblicitari dei proprietari dei siti *web* per poi rivenderli agli inserzionisti. Accanto alle concessionarie di pubblicità *online* (proprietarie o terze) operanti anche su altri mezzi, si riscontra la presenza di piattaforme tecnologiche quali gli *ad network*, i *network* di affiliazione[41] e gli *ad exchange*[42].

[41] Le *Ad network* o *affiliation network* sono intermediari che aggregano o acquisiscono spazi pubblicitari da un insieme di siti *web* e vendono tale inventario agli inserzionisti. La funzione della piattaforma è quella di aggregatore degli spazi pubblicitari che aderiscono al *network*, di promozione degli stessi presso gli inserzionisti e di distribuzione delle inserzioni tra i siti appartenenti al *network*. Cfr. Assocom, FCP-Assointernet, Fedoweb, Fieg, Iab Italia, Netcomm, Unicom, UPA, *Libro Bianco sulla comunicazione digitale*, 2017.

[42] L'*Ad exchanges* rappresentano delle piattaforme tecnologiche basate sul protocollo *real time bidding* (RTB), ossia sulle aste *online*, che consente ad inserzionisti ed editori/*publisher* (fornitori di contenuti e servizi *web*) di acquistare e vendere dinamicamente spazi pubblicitari. Cfr. *Libro Bianco sulla comunicazione digitale*, *op. cit.*

Figura 4 - *Programmatic advertising*



Il processo di compravendita di spazi pubblicitari in *programmatic advertising*

1. Il processo inizia nel momento i cui il browser dell'utente richiede una pagina web di un sito che utilizza il programmatic e chiama l'ad server
2. L'ad server chiama la *supply side platform* (SSP) utilizzata dall'editore
3. La SSP effettua una chiamata ad uno o più *ad exchanges* che a sua volta effettua diverse chiamate o richieste di offerta a diverse *demand side platform* (DSP) che cercano di ottenere il miglior prezzo
4. Le DSP valutano le richieste di offerta (*bid request*) e decidono in base ai parametri impostati come discriminanti (prezzo/targetting/localizzazione...) se rispondere all'*ad exchange* e quanto offrire
5. L'*ad exchange* con le regole del RTB e le impostazioni dell'editore sulla SSP seleziona l'offerta più alta e a che prezzo per poi trasmettere l'url dell'annuncio vincente alla SSP
6. La SSP comunica l'offerta vincente all'*ad server* che risponde con l'annuncio selezionato al browser dell'utente e procede a passare la creatività sul sito del publisher
7. Le *data management platform* (DMP) raccolgono, aggregano ed elaborano i dati relativi a tutto il processo di distribuzione della pubblicità

Flussi finanziari della compravendita di spazi pubblicitari in *programmatic advertising*

A. L'inserzionista paga per i servizi resi da: centro media/*trading desk* indipendente, *ad server*, fornitore di servizi di *web analytics*, affitto o costruzione DMP, *ad server* o società terze per *ad verification*
B. L'inserzionista (o centro media/*trading desk* indipendente) paga alla DSP il costo degli spazi acquistati comprensivo della fee pattuita con la DSP
C. La DSP paga la SSP sulla base all'*impression* acquistate (SSP può applicare o meno una fee alla DSP)
D. La SSP trattiene una *fee* tecnologica pattuita e retrocede all'editore/concessionaria il valore ottenuto dalla vendita degli spazi pubblicitari fatturati alla DSP
E. L'editore/*Publisher* fattura alla concessionaria che trattiene una parte e retrocede allo stesso. L'editore paga i servizi ottenuti (*ad server*, *web analytics*, affitto/costruzione DMP, *ad verification*)

Il meccanismo appena descritto è il frutto di un **processo di trasformazione**, favorito dallo sviluppo tecnologico, che ha investito il settore della raccolta di pubblicità *online* negli ultimi 10 anni e ha riguardato diversi aspetti: i) i formati pubblicitari; ii) le tecniche di profilazione e targetizzazione degli utenti; iii) il crescente impiego di tecnologie (in particolare dell'intelligenza artificiale di tipo *low*); iv) le modalità di interazione fra domanda e offerta di pubblicità, sempre più attraverso modelli automatici di compravendita piuttosto che attraverso la contrattazione diretta; v) il processo di formazione dei prezzi[43].

Tale evoluzione del sistema pubblicitario ha consentito, innanzitutto, margini di guadagno per tutti gli operatori coinvolti nella filiera della pubblicità *online* (editori, piattaforme, inserzionisti, intermediari); d'altra parte, essa ha comportato la creazione di un **contesto ideale** nel quale strategie di disinformazione con finalità commerciali di breve-medio periodo (valorizzazione del traffico attraverso la vendita di spazi pubblicitari *online* agli inserzionisti) possono dispiegarsi.

In particolare, fra gli **elementi di mercato che hanno favorito la diffusione di fenomeni di disinformazione *online*** e strategie di valorizzazione attraverso la pubblicità è possibile evidenziare:

A. la crescita della distanza fra inserzionista ed editore/*publisher* e la diminuzione del livello di trasparenza circa i flussi di distribuzione delle inserzioni pubblicitarie;

B. l'incremento della complessità del sistema pubblicitario e del rischio di esposizione ad azioni di fraudolente (*Ad Fraud*);

C. il potenziamento della capacità di profilazione dell'utenza e di personalizzazione del messaggio, compresi i contenuti di disinformazione a *target* specifici.

Nel seguito questi tre aspetti saranno ripresi ed esaminati distintamente, evidenziando le implicazioni che da essi scaturiscono per le strategie di disinformazione *online*.

A. Crescita della distanza fra inserzionisti ed editori/*publisher* e diminuzione della trasparenza dei flussi di distribuzione delle inserzioni pubblicitarie

L'intero sistema pubblicitario *online* e il guadagno ottenuto dai diversi protagonisti sono direttamente correlati alla capacità di individuare il *target* desiderato, al numero delle visite ottenute e alle *impression* servite che aumentano la probabilità di un'azione da parte dell'utente (come ad esempio i *click* sugli annunci pubblicitari) a fronte della visualizzazione del messaggio pubblicitario. La capacità di un sito internet di **generare *click*** su annunci pubblicitari in virtù del contenuto offerto (compresi pertanto i contenuti *fake*) rappresenta, in effetti, l'elemento principale che consente di attrarre investimenti pubblicitari.

La circostanza che il sistema di compravendita di pubblicità *online* sia basato sulla ricerca dei *click* da parte di utenti attentamente profilati, ha fatto sì, inoltre, che l'attenzione degli inserzionisti fosse sempre più incentrata sul *target* piuttosto che sul contenuto/sito *web*/piattaforma *online* che ospita la pubblicità e sulla sua effettiva capacità di attrarre *audience*. A ciò si è unito sia la possibilità dell'automatismo dei meccanismi di negoziazione degli spazi pubblicitari *online* in *programmatic*, che determina una distanza tra la domanda di spazi pubblicitari e l'offerta, sia il conseguente proliferare lungo la filiera di un numero elevato di operatori, piattaforme tecnologiche e intermediari anch'essi frapposti tra inserzionisti e editori/*publisher*. Inoltre, l'elevata complessità dei sistemi di negoziazione della pubblicità *online* spesso induce gli inserzionisti, soprattutto quelli con *budget* più limitati, a servirsi di intermediari esternalizzando tutta o molta parte delle

[43] La pubblicità *online* può essere venduta a CPM (cost per mille *impression*) modello utilizzato prevalentemente nella display e video advertising; CPC (*cost per click*) tipicamente impiegata per la pubblicità di tipo *search e classified/directory*; a *performance* ovvero in base al raggiungimento di un risultato (fra cui, CPA – *cost per action*; CPL *cost per lead*, CPS – *costo per sale*, CPO - *cost per order*)) ovvero in base ad altri parametri come il tempo o in numero di visitatori (CPV - *cost per visit*). La pubblicità all'interno dei social network utilizza anche modelli specifici in funzione delle azioni di comunicazione social (ad. es. CPL - *cost per like*).

attività di gestione delle proprie campagne pubblicitarie (dalla pianificazione, alla compravendita, fino alla verifica *ex post*).

Tali aspetti hanno un impatto diretto sulla proliferazione della disinformazione *online*, poiché favoriscono un ambiente in cui vi è una **scarsa capacità di controllo** sull'intero processo di negoziazione da parte degli investitori pubblicitari che non sono in grado di conoscere con esattezza – e in anticipo – in quali siti verranno visualizzati i messaggi pubblicitari.

Ciò accade nonostante la possibilità di adottare **comportamenti orientati alla trasparenza** da parte di tutti gli operatori coinvolti. È possibile, infatti, per l'inserzionista impostare attraverso le DSPs i parametri di acquisto in modo che le proprie *impression* non siano servite su siti illegali (fra cui quelli che propongono disinformazione) o siti con contenuti ritenuti contrari alla *brand policy* definita dall'azienda (introducendo *blacklist/whitelist*; concordando preventivamente i parametri semantici e le *keyword* in modo da escludere contenuti indesiderati). Inoltre, è possibile utilizzare *partner* commerciali che operino in modalità *trasparent*[44] ovvero adottare degli strumenti di *ad verification* per il rispetto della brand *policy* dei singoli contenuti, nonché politiche di intervento e azioni volte alla rimozione dell'annuncio visualizzato e/o di blocco dell'erogazione dell'annuncio.

In generale, queste misure, improntate alla trasparenza da parte dei diversi attori economici, appaiono non efficaci nel contrastare l'uso del sistema della pubblicità *online* ai fini del *business* dei contenuti di disinformazione. Restano, infatti, **spazi di inserimento** per chi sia in grado di aggirare i meccanismi di controllo dei sistemi di compravendita della pubblicità *online*; in tal senso, persino singoli utenti o gruppi di utenti sono capaci di sfruttare da un lato le tecnologie, i protocolli e le piattaforme automatiche che consentono il funzionamento dello stesso sistema di compravendita della pubblicità *online*, dall'altro lato i meccanismi e le logiche sottese alla distribuzione e propagazione delle notizie attraverso i *social network*: questo sembra indicare il caso dei ragazzi e degli studenti residenti a Veles (vedi **Caso 1**) scoperto attraverso un'inchiesta condotta dalla CNN[45]. Nel dettaglio si osserva come l'elevato grado di **polarizzazione** dei soggetti che si informano attraverso i *social network*, insieme ai processi per i quali le notizie più radicalizzate diventano virali attraverso tali piattaforme, favorendo la disseminazione della disinformazione *online*, abbiano svolto un ruolo fondamentale per il successo della strategia. In altri termini, la tendenza alla polarizzazione degli utenti attorno a tematiche dibattute è stata sfruttata dai ragazzi residenti a Veles per creare notizie false che grazie alle azioni di condivisione delle stesse all'interno del *social network* ne hanno consentito la diffusione virale allo scopo di attirare traffico nei propri siti e generare introiti pubblicitari in un orizzonte di breve-medio periodo

Tale strategia si è, pertanto, fondata su un fenomeno del ***click baiting*** (letteralmente "esca da click") che si riferisce ad un contenuto diffuso *online* la cui funzione principale è quella di attirare il maggior numero di *click* e, pertanto, generare maggiori ricavi da pubblicità *online*. Generalmente il *click bait* rappresenta un contenuto contraddistinto da titoli sensazionalistici ed accattivanti, suscettibili di fare leva sulla sensibilità e le emozioni di chi accede, con l'intento di incoraggiarne la condivisione attraverso i *social network*, incrementando, pertanto, gli introiti derivanti dalla vendita di pubblicità *online*. Sussiste un legame tra polarizzazione ideologica degli utenti attorno ad una determinata tematica e maggiore impegno (*engagement*) degli stessi nei confronti delle notizie divulgate sui *social network* che trattano tale argomento. Tale legame, unito all'operare degli algoritmi di presentazione automatica dei contenuti, si riflette sul concretizzarsi di fenomeni di viralità connessi alla diffusione di posizioni radicalizzate e disinformazione. Pertanto, maggiore è il livello di polarizzazione ideologica degli individui, maggiore sarà la probabilità che attorno agli stessi si diffondano contenuti di disinformazione allo scopo di generare *click* che si traducano in maggiori ricavi pubblicitari. In tale contesto di riferimento, complice del successo di tale strategia di disinformazione, è altresì **l'automatismo dei sistemi di compravendita** di pubblicità che, in assenza di adeguate misure di trasparenza, consente di

[44] Si tratta di una modalità di azione delle piattaforme e degli intermediari utilizzati che comporta una dichiarazione preventiva alla campagna dell'elenco dei bacini dei siti su cui l'annuncio di un *brand* potrebbe essere servito, con evidenza dei siti aggregati da *property* terze (*source traffic*); consentendo in fase di *post* valutazione il controllo censuario a posteriori di tutti i siti su cui è stata servita la campagna pubblicitaria. Cfr. *Libro Bianco sulla comunicazione digitale, op. cit.*.

[45] CNN, *The Fake news machine: inside a town gearing up for 2020.*

vendere *target* allo scopo di massimizzare le *impression* servite o i *click* sui *banner* pubblicitari e, quindi, i guadagni di tutti gli attori economici coinvolti nel processo, veicolando pubblicità sui siti *web* indipendentemente dalla falsità o meno dei contenuti presenti.

È possibile, tuttavia, distinguere una prima fase nella quale tali strategie di disinformazione *online*, fondate sul *click baiting*, hanno avuto un discreto successo generando guadagni pubblicitari rilevanti per gli ideatori ed esecutori della stessa.

Successivamente, in seguito alle modifiche degli algoritmi di selezione automatica dei contenuti realizzate dalle principali piattaforme orizzontali – Google e Facebook – per contrastare e contenere la diffusione del fenomeno della disinformazione *online*, la sostenibilità economica di tali strategie si è decisamente ridotta[46]. In particolare, Google ha recentemente modificato la *policy* di Google Ads esplicitando le categorie di contenuti vietati[47] e di attività vietate, a da novembre 2016, ha rafforzato le misure di trasparenza riferibili alla propria *policy* sui "Contenuti ingannevoli", esplicitando in particolare che i *"Google Ads non possono essere inseriti in pagine che nascondono informazioni o che forniscono informazioni ingannevoli"* e fornendo, inoltre, esempi concreti di contenuti non accettabili[48]. Inoltre, sono state introdotte delle misure di trasparenza a protezione degli inserzionisti finalizzate ad evitare che le inserzioni compaiano in siti con contenuti di dubbia provenienza o contrari ai propri valori[49], ovvero ad accrescere la consapevolezza o il controllo dei siti presso i quali saranno visualizzate le inserzioni pubblicitarie[50]. Con riferimento a Facebook, la strategia volta a contrastare il fenomeno della diffusione di contenuti *fake* si basa sui seguenti elementi: (i) rimozione di "*fake account*" contrari agli standard della comunità e alla politica dell'azienda; (ii) promozione di strumenti di *fact-checking*; (iii) riduzione della convenienza economica (vendita di pubblicità, *Instant Articles*) da parte degli *account* e/o delle pagine che in modo sistematico diffondono e/o condividono contenuti *fake*; (iv) contenimento della diffusione di contenuti *fake* attraverso il *ranking* nell'ambito della News Feed ovvero sviluppo di strumenti di informazione sempre nella News Feed (editore, condivisioni precedenti) volti ad accrescere la consapevolezza dell'utenti; vi) promozione di programmi, azioni di educazione e alfabetizzazione dell'utenza finalizzati a riconoscere eventuali contenuti di disinformazione[51].

---

[46] Cfr. i comunicati stampa di Facebook, *Increasing Our Efforts to Fight False News*, By Tessa Lyons, Product Manager, del 21 giugno 2018; *Hard Questions: What's Facebook's Strategy for Stopping False News?* del 23 maggio 2018; *Update on Our Advertising Transparency and Authenticity Efforts*, del 27 ottobre 2017. Sul punto si veda anche il verbale di Facebook del 9 maggio 2018 nell'ambito del *Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali.*

[47] Si vedano le norme di Google ads che riportano i contenuti vietati (articoli contraffatti, prodotti o servizi pericolosi, comportamenti disonesti, contenuti inappropriati) e le attività vietate (abuso della rete pubblicitaria, raccolta e utilizzo di dati, false dichiarazioni) consultabile al seguente link https://support.google.com/adspolicy/answer/6008942?hl=it

[48] Nell'ambito della *policy* sui contenuti ingannevoli sono stati individuati degli esempi di comportamenti che possono trarre in inganno l'utente fra cui: Attirare gli utenti a interagire con contenuti sotto falsi o poco chiari pretesti; "Phishing" per le informazioni degli utenti; Promozione di contenuti, prodotti o servizi che utilizzano affermazioni false, disoneste o ingannevoli (ad esempio schemi "Get Rich Quick"); Impersonare i prodotti Google; Affermare falsamente di avere un'affiliazione o un'approvazione da parte di un altro individuo, organizzazione, prodotto o servizio; Dirigere i contenuti su questioni politiche, sociali o di interesse pubblico per gli utenti in un Paese diverso dal proprio, se si travisano o nascondono il proprio paese di origine o altri dettagli materiali su di sé. La policy di Google è consultabile al seguente link: https://support.google.com/adsense/answer/1348688?hl=it#Misrepresentative_content

[49] In particolare, con riferimento alla piattaforma You Tube sono state individuate categorie precise di video non adatti alla pubblicità che si aggiunge al meccanismo degli avvertimenti relativi alle Norme della community che possono portare alla restrizione di alcune opzioni o addirittura all'esclusione dallo YouTube Partner Program o, nei casi più gravi, all'esclusione completa da YouTube. Cfr. https://support.google.com/youtube/answer/6162278?hl=it e https://support.google.com/youtube/answer/2802032?hl=it

[50] Sempre per la pubblicità di tipo video Google mette a disposizione degli inserzionisti, diversi strumenti per il targeting, che consentono di controllare dove vengono visualizzati gli annunci su YouTube, tra cui strumenti sviluppati da terze parti, come la tecnologia di verifica di Integral Ad Science che consente il monitoraggio del livello di rischio della sicurezza del *brand*, evitando, pertanto, che il proprio *brand* compaia su siti a tema a rischio o semplicemente non coerente con i desideri dell'inserzionista stesso. Cfr. https://support.google.com/youtube/answer/2454017 e https://support.google.com/displayvideo/answer/3297897?hl=it

[51] Cfr. Si vedano i comunicati stampa di Facebook, *Increasing Our Efforts to Fight False News*, By Tessa Lyons, Product Manager, del 21 giugno 2018; *Hard Questions: What's Facebook's Strategy for Stopping False News?* del 23 maggio 2018; *Update on Our Advertising Transparency and Authenticity Efforts*, del 27 ottobre 2017. Sul punto anche il verbale di Facebook del 9 maggio 2018 nell'ambito del *Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali.*

In generale, appare chiaro come l'evoluzione tecnologica consenta ai soggetti che intendono perseguire tali strategie di disinformazione di pianificarne ogni singolo dettaglio della stessa, adattandosi ai nuovi algoritmi di presentazione automatica dei contenuti delle piattaforme, nonché sfruttando le difficoltà di controllo da parte degli inserzionisti dei flussi di distribuzione delle inserzioni pubblicitarie *online* attraverso i modelli compravendita automatica, richiedendo pertanto ulteriori misure volte ad ampliare il livello di trasparenza del sistema pubblicitario.

CASO 1 - Veles (Repubblica di Macedonia), la capitale della disinformazione *online*: polarizzazione, *clickbaiting* e pubblicità *online*

Il caso presentato in questo approfondimento illustra come singoli utenti o gruppi di utenti siano in grado, a fronte di investimenti tecnologici piuttosto limitati, di adeguate conoscenze dei meccanismi di funzionamento delle piattaforme di *digital advertising*, nonché di conoscenze dei processi di propagazione delle notizie attraverso i *social network*, di creare, produrre, distribuire *online* contenuti di disinformazione, da monetizzare, poi, attraverso la vendita di pubblicità *online*.

L'esempio sotto riportato risulta particolarmente calzante, tenuto conto del coinvolgimento nella strategia di disinformazione sia di una delle principali piattaforme di intermediazione pubblicitaria *online* (Ad Sense) – riconducibile al primo operatore per ricavi da pubblicità *online* a livello mondiale (Google) –, sia del principale *social network* (Facebook), secondo operatore in termini di risorse pubblicitarie complessive.

Nei primi mesi del 2017 la CNN ha pubblicato gli esiti di un'inchiesta che ha indagato sulla presenza di almeno 100 siti dedicati all'informazione politica nei quali venivano (e in alcuni siti sono tuttora) pubblicati contenuti *fake* a favore di Donald Trump, siti tutti gestiti e di proprietà di utenti ubicati nella città di Veles, comune della Repubblica di Macedonia (poco più di 55.000 abitanti)

L'inchiesta ha messo in luce l'esistenza di una strategia, perseguita da singoli individui o gruppi, finalizzata alla realizzazione di introiti attraverso la vendita di pubblicità *online* diffusa su siti di disinformazione e fondata sui seguenti elementi:

- *I soggetti*. Gli ideatori ed esecutori della strategia di disinformazione scoperta dalla CNN sono studenti, spesso minorenni o giovani privi di prospettive occupazionali in considerazione delle scarse opportunità offerte dal territorio. Essi hanno creato un proprio sito internet, nel quale, durante la campagna elettorale statunitense, sono apparse notizie *fake*. Al riguardo, l'inchiesta ha evidenziato come alcuni dei siti esaminati presentassero esclusivamente contenuti *fake*, mentre in altri si potessero trovare anche notizie vere, o parzialmente vere e rielaborate a partire dai contenuti apparsi in altri siti di informazione. Alcuni dei ragazzi protagonisti della vicenda si avvalevano, inoltre, della collaborazione nella produzione dei contenuti di disinformazione di altri utenti, ubicati negli Stati Uniti, così da rendere la notizia *fake* più aderente alle aspettative del *target* dei destinatori. Determinante nella fase di distribuzione è stata la creazione di falsi *account* Facebook – spesso acquistati da minori a fronte del pagamento di cifre piuttosto irrisorie – attraverso i quali le notizie di disinformazione venivano postate, commentate e rilanciate all'interno del *social network*, sfruttando i processi di propagazione delle notizie attraverso tale piattaforma. Più il contenuto *fake* veniva letto, cliccato, condiviso, postato o commentato all'interno della *social network*, riconducendo, pertanto, l'utente a consultare il sito di disinformazione *online*, maggiori erano le opportunità di guadagno per il proprietario dello stesso, collegate alle *impression* ovvero ai *click* sui *banner* pubblicitari serviti dalla piattaforma di intermediazione pubblicitaria scelta. Quanto ai destinatari della strategia, si osserva che le notizie *fake* si rivolgevano ad uno specifico *target* di statunitensi ossia ai conservatori sostenitori di Donald Trump.
- *Le motivazioni*. La strategia ha lo scopo di attirare traffico presso i propri siti sui quali venivano servite le inserzioni pubblicitarie. L'intervista condotta dalla giornalista della CNN ad uno dei principali protagonisti della vicenda, ha evidenziato come nel corso delle ultime elezioni presidenziali statunitensi il guadagno giornaliero fosse della misura di 2.000-2.500 euro al giorno, a fronte delle *impression* servite o dei *click* ricevuti, superando di gran lunga il reddito corrispondente all'intera vita lavorativa di un lavoratore medio macedone, il cui reddito mensile non supera 400 euro.
- *Le risorse tecnologiche*. I protagonisti hanno utilizzato dei nomi che richiamano quelli di testate di informazione *online* (es. USADailyPolitics.com, WordlPoliticus.com) con una grafica non particolarmente raffinata ma sufficientemente credibile. Secondo le dichiarazioni di uno dei principali protagonisti della vicenda alla giornalista della CNN, la piattaforma di intermediazione pubblicitaria *online* utilizzata è Ad Sense di Google che ha provveduto a somministrare la pubblicità sui siti di disinformazione remunerando i proprietari degli stessi in funzione delle *impression* servite o dei *click* effettuati sugli annunci. La strategia prevedeva, come detto, la creazione attraverso il

social network Facebook di *account* falsi che venivano utilizzati per postare, condividere e commentare le notizie *fake* apparse nei siti di disinformazione allo scopo di aumentarne le visite.

- *Gli investimenti*. Tale strategia commerciale ha richiesto limitati investimenti tecnologici (è sufficiente un pc) ma adeguate conoscenze sul funzionamento dei meccanismi automatici di compravendita di pubblicità *online* e delle modalità di propagazione delle notizie attraverso i *social network*. L'adesione alla piattaforma di intermediazione pubblicitaria Ad Sense è gratuita in quanto è sufficiente un *account* Google, così come per la creazione di un proprio sito internet ci si può avvalere di servizi completamente gratuiti. L'indagine della CNN ha, inoltre, evidenziato come uno dei primi ideatori di tale strategia abbia ampliato il proprio *business* proponendo delle attività di formazione rivolta agli studenti che, come lui, sono interessati a perseguire una strategia commerciale fondata su contenuti *fake* nella prospettiva di prepararsi per le prossime elezioni presidenziali del 2020 negli Stati Uniti.

Il caso appena illustrato ha dimostrato che esiste un modello di *business* sostenibile dei contenuti di disinformazione *online* e basato sulla pubblicità *online* In particolare, i protagonisti di tale strategia sostenendo costi estremamente contenuti sono stati in grado di sfruttare a proprio vantaggio alcuni elementi del sistema di compravendita di pubblicità *online* che ostacolano la capacità di controllo da parte degli inserzionisti, nonché le peculiarità dei meccanismi di diffusione delle notizie attraverso i *social network*. Emerge, infatti, come l'elevato livello di polarizzazione dei soggetti più attivi *online*, unito all'esistenza di algoritmi che operano una personalizzazione automatica dei contenuti di informazione, creino un contesto che offre maggiori opportunità di *business* a chi produce contenuti *fake* che rispondono a posizioni politiche piuttosto radicalizzate affinché si diffondano e diventino virali.

B. Incremento della complessità del sistema pubblicitario ed esposizione ad azioni fraudolente

Le evoluzioni che hanno caratterizzato il funzionamento del sistema di compravendita di pubblicità *online* negli ultimi anni hanno sicuramente accresciuto il **grado di complessità** dello stesso con riferimento ai seguenti aspetti:

- **numerosità dei soggetti** che intervengono nei processi automatizzati di compravendita ed erogazione della campagna pubblicitaria (e in particolare se gestita secondo modello automatici, come nel caso del *programmatic*) e conseguente difficoltà di comprensione dei ruoli, delle funzioni e dei legami societari (partecipazioni e/o società controllanti/controllate) fra gli stessi;
- **crescente affidamento a sistemi automatizzati** basati sui dati nel processo di compravendita che sebbene consenta agli inserzionisti di intercettare *target* degli utenti perfettamente profilati non offre sempre adeguate garanzie circa i siti presso i quali verranno effettivamente distribuite le inserzioni pubblicitarie;
- **varietà delle transazioni pubblicitarie** che risultano molto più complicate rispetto allo schema rappresentato in **Figura 4** (schematizzato attraverso i punti 1-7), prevedendo numerosi processi di re-intermediazione (fra *ad network/affiliation network*) nonché la partecipazione di numerose *ad exchanges* nelle singole transazioni. Questo conduce ad una complicazione dei flussi finanziari (il cui schema è molto più articolato rispetto ai passaggi da A a E illustrati in **Figura 4**) comportando anche un aumento dei costi di intermediazione pubblicitaria e una riduzione degli introiti realizzati dagli editori/*publisher* rispetto all'investimento sostenuto dagli inserzionisti[52].

Tali elementi contribuiscono a rendere il sistema di negoziazione della pubblicità *online* più **vulnerabile**, creando spazi di azione per coloro che intendono perseguire strategie di disinformazione con finalità commerciali.

L'interesse di coloro che perseguono tali strategie di disinformazione – ottenere traffico e vendere *inventory* in modo da intercettare quanto più volume di investimento pubblicitario possibile – trova perfetto allineamento con l'interesse delle piattaforme automatiche dal lato della domanda (DSP, *trading desk*, *ad server*) che, salvo

[52] Secondo le stime di Warc il 60% dell'investimento pubblicitario complessivo sarebbe assorbito dagli intermediari, mentre solo il restante 40% rappresenta quanto rigirato ai *publisher*. Cfr. Emarketer, *Why tech firm ontain most of the money in Programmatic Ad Buy*s, 16 aprile 2018.

diverse impostazioni, attraverso i modelli di compravendita in *programmatic* distribuiscono *impression* in funzione dei volumi di traffico e del *target* desiderato. Il sistema così configurato, caratterizzato da sistemi automatici e dall'assenza di un contatto diretto fra inserzionista e *publisher*, è in effetti **intermediato e re-intermediato** da diversi soggetti e articolato secondo schemi di compravendita pubblicitaria complessi e in continua evoluzione. Il rischio in cui si incorre è di favorire il finanziamento di siti che producono informazione di scarsa qualità, compresi quelli di disinformazione, creando spazi di azione per forme di **distrazione dell'investimento pubblicitario**.

I protagonisti di strategie di disinformazione, infatti, possono avvalersi di tecnologie idonee a **manipolare il traffico** proveniente dai propri siti in modo da associare ad essi una *viewability* più alta, secondo i parametri delle *ad exchanges*, e, inoltre, possono **modificare ad *hoc* il proprio inventario** per renderlo più desiderabile ed appetibile secondo le impostazioni degli algoritmi di distribuzione automatica di pubblicità utilizzati dalle piattaforme dal lato della domanda (*Ad server*, *DSP*, *trading desk*, cfr. **Figura 4**).

Come illustrato dal rapporto della World Federation of Advertiser, analizzando i primi 5000 siti ritenuti più pregiati in termini di *inventory* disponili attraverso le ad *exchanges*, si osserva come la maggioranza degli stessi siano riconducibili a "*viral-news-site*", ossia a siti di informazione basati su contenuti accattivanti, di moda, clamorosi, in grado di suscitare interesse e stimolare emozioni negli utenti allo scopo di generare traffico, innescando fenomeni di viralizzazione[53]. Tali siti sembrano attualmente assorbire una quota significativa delle transazioni, e dei corrispondenti investimenti pubblicitari che avvengono in *programmatic*, mentre le analisi circa la qualità del traffico generato dagli stessi mostrano una ridotta efficacia dell'investimento pubblicitario e un elevato rischio che allo stesso siano associati fenomeni di frode pubblicitaria (cfr. *infra*).

Inoltre, gli editori che producono informazione primaria e di qualità competono con tale categoria di siti nel versante degli utenti per ottenere livelli di *audience* da valorizzare in termini di raccolta pubblicitaria. La pressione competitiva esercitata dai siti di informazione più popolari indurrebbe gli editori *online* ad acquistare *sourced traffic*, derivante da *property* terze, maggiormente soggetto a rischio di *ad fraud.*

Infatti, come illustrato nel caso *Methbot* presentato nel *box* di approfondimento (**Caso 2**), le complessità del funzionamento del sistema pubblicitario *online* sopra richiamate e l'accresciuto livello di anonimato dei diversi attori coinvolti possono essere sfruttate da operatori, gruppi di utenti, organizzazioni più o meno strutturate, che, attraverso opportuni investimenti tecnologici e impostando adeguatamente le piattaforme di *digital advertising* (*ad server*, SSP), sono in grado di **aggirare i meccanismi di controllo e verifica** improntati alla trasparenza, oltre ai sistemi deputati alla verifica del traffico non-umano, per adottare azioni fraudolente con l'intento di realizzare ricavi pubblicitari.

Si parla di ***ad fraud*** per identificare la creazione illegittima di traffico – composto sia da traffico non umano sia da pratiche illegittime poste in essere da esseri umani – al fine di cercare deliberatamente di distrarre parte degli investimenti pubblicitari. Esistono diverse tipologie di frodi, per le cui definizioni si può fare riferimento alla tassonomia TAG (Trustworthy Accountability Group), fra cui sinteticamente si ricordano:

- *impression fraud*,
- *click fraud* – come nel caso del traffico generato dai *bot* utilizzati da Methbot - e *click baiting*,
- *conversion fraud*,
- *data fraud*.

In tutti questi casi, è possibile adottare degli accorgimenti da parte degli inserzionisti volti a ridurre la possibilità di vedersi attribuiti i costi delle campagne pubblicitarie collegate a pratiche fraudolente[54]. Tuttavia, il caso presentato nel *box* di approfondimento dimostra come gli ideatori di strategie commerciali a danno del

[53] WFA, *Compendium of ad fraud knowledge for media investor*, 2016
[54] Cfr. *Libro Bianco sulla comunicazione digitale*, 2017, cit. p. 39 e ss.

sistema pubblicitario siano in grado di sviluppare sistemi sempre più articolati che sfruttano tecnologie in grado di evolversi nel tempo allo scopo di aggirare i meccanismi di controllo e di *ad verification*.

CASO 2 - AD Fraud Comanda /AFK$_{13}$: disinformazione, frodi e raccolta pubblicitaria

Il caso presentato in questo approfondimento illustra come un'organizzazione con finalità criminali sia stata in grado, attraverso un'accurata pianificazione di una strategia di medio/lungo periodo basata sulla combinazione di rilevanti investimenti tecnologici (sia *hardware*, *software*) e di adeguate competenze informatiche, ingegneristiche e di *digital marketing*, di attuare una pratica di *ad fraud* innovativa e senza precedenti.

Sfruttando i limiti e le complessità caratterizzanti il sistema pubblicitario *online*, la pratica è stata fondata sui seguenti elementi che ne hanno determinato il successo:

- *I soggetti*. In base alle verifiche effettuate da White Ops[55], società specializzata in servizi di monitoraggio di pratiche fraudolente nel settore della pubblicità *online*, la strategia è stata progettata e gestita da una vera e propria organizzazione criminale ubicata in Russia. White Ops già dal settembre 2015 aveva iniziato a monitorare il traffico non umano (generato da alcuni *robot* chiamati "C3") a danno del sistema pubblicitario che, tuttavia, ancora operava su scala piuttosto ridotta. A partire da ottobre 2016, il sistema fraudolento si è evoluto rapidamente trasformandosi in "Methbot", così denominato in ragione della stringa "Meth" presente nel codice dei *bot*, e ha iniziato ad operare su una scala molto più ampia (137 milioni di *impression* sono state richieste dagli indirizzi IP attribuibili alla rete in esame). I destinatari dell'azione fraudolenta sono stati (e potrebbero essere tuttora) gli inserzionisti pubblicitari interessati a raggiungere specifici *target* di utenti, le cui caratteristiche ed azioni (*click*) sull'inserzione pubblicitaria sono state artificialmente generate dal sistema per finalità commerciali.
- *Le motivazioni.* L'organizzazione ha predisposto il sistema Methbot con l'intento di realizzare ricavi dalla vendita agli inserzionisti pubblicitari dei contatti derivanti dalle visite e dalle visualizzazioni prodotte artificialmente dai *bot* che in apparenza sembravano effettuate su domini riconducibili ad importanti editori *online*. In base alle stime di White Ops, tenuto conto di un CPM medio di 13 dollari, la strategia avrebbe garantito all'organizzazione un guadagno giornaliero (per almeno 2 mesi) compreso fra 2,6 e 5,6 milioni di dollari (tenuto conto di una stima delle *impression* medie giornaliere generate artificialmente da Methbot all'interno di un *range* di 200-300 milioni)
- *Le risorse tecnologiche utilizzate.* L'organizzazione criminale ha creato una vera e propria rete caratterizzata sia da elementi *hardware*, sia da *software* programmati in modo da raggiungere lo scopo commerciale, fondata sulla:

  (i) registrazione di almeno 852.922 indirizzi numerici (IP) falsificandone la relativa documentazione in modo da simulare che gli stessi fossero riconducibili ai principali fornitori di servizi di connessioni *online* statunitensi come AT&T, Comcast, Verizon, ecc. In alcuni casi, sono stati utilizzati anche nomi di falsi *provider* che richiamano quelli di aziende esistenti (come AmOL al posto di AOL) così da fingere la provenienza del traffico dagli Stati Uniti;

  (ii) falsificazione di oltre 6.111 domini o *url* in modo da sembrare che gli stessi fossero associati ad editori *online* particolarmente conosciuti;

  (iii) predisposizione di una rete di *server* (800 – 1200) dedicati, collocati in *data center* ubicati in Stati e continenti differenti (Stati Uniti e Europa) per evitare la provenienza del traffico da una unica fonte;

  (iv) associazione di *robot* o *bot* agli indirizzi IP in modo da simulare, attraverso alcuni accorgimenti tecnici, l'attività di navigazione da parte un utente reale (manipolazione della localizzazione geografica diversa fra i vari *bot*, simulazione di una cronologia di siti già visitati e predisposizione di credenziali fasulle legate ai principali *social network* come Facebook).

  Nel dettaglio, i *bot* erano programmati in modo da simulare una normale sessione di navigazione *desktop* da parte di un utente attraverso uno dei principali *browser* (Crome, Firefox, Internet Explorer, Safari). L'organizzazione ha adottato una serie di azioni volte ad ingannare i sistemi automatici di compravendita della pubblicità *online* e quelle deputati al controllo delle *ad fraud* e del traffico non umano. In particolare, attraverso la creazione di *url* e domini falsi, Methbot ha simulato la provenienza delle visite artificiali da siti di editori *online* particolarmente conosciuti, fra cui importanti siti di informazione come *Wall Street Journal*, *Fox News*, *Vogue*, ESPN. In realtà, il traffico generato dai *bot* corrispondeva a siti misconosciuti creati *ad hoc* dall'organizzazione criminale per ospitare quale unico contenuto la pubblicità di tipo video. Inoltre, utilizzando il protocollo standard VAST, i *bot* hanno inoltrato quotidianamente delle chiamate di richiesta di *impression* (pubblicità di tipo video) ai vari *ad network*. Infine, una

[55] White Ops, *The Methbot Operation*, 20 dicembre 2016.

volta servita la pubblicità attraverso i sistemi automatici di *adverstising*, i *bot* procedevano alla produzione di false visualizzazioni del video pubblicitario e di *clicks* (in media si stima che i *bot* abbiamo prodotto da 200 a 300 *video advertising impressions* al giorno), simulandone l'azione da parte di un *browser*.

- *Gli investimenti economici*. L'organizzazione criminale ha effettuato rilevanti investimenti in tempo, attività di ricerca e sviluppo, risorse tecnologiche, infrastrutture *hardware* e in *software* (acquisto di *server* e di spazio per gli stessi nei *data center*, programmazione dei *bot*) per la predisposizione della propria strategia che si è evoluta ed adattata nel tempo ai sistemi di *ad fraud* e di verifica del traffico non umano per crescere rapidamente su scala internazionale.

In definitiva, l'attuale complessità, interconnessione e il conseguente anonimato dell'ecosistema della pubblicità *online* è stato sfruttato da Methbot per perseguire la propria strategia commerciale producendo danni ingenti, sia dal lato della domanda di pubblicità, sia da quello della offerta, anche in considerazione della scala globale dell'azione fraudolenta. Dal lato della domanda, infatti, accanto ai costi in termini di CMP pagato dagli inserzionisti a fronte dei video serviti e visualizzati in modo fraudolento, il danno subito a livello di sistema economico risulta essere più ampio (considerando sia l'incremento dei costi a carico degli investitori del sistema pubblicitario collegati al monitoraggio delle campagne pubblicitarie – servizi di *viewabilty*, *brand safety* e *brand security* – sia i costi all'economia del paese, come ad esempio quelli riconducibili ai mancati introiti fiscali). Dal lato della offerta, tenuto conto che il sistema simulava traffico proveniente da siti dei maggiori editori statunitensi *online*, l'operazione Methbot ha prodotto un altrettanto ingente costo consistente nel mancato guadagno e finanziamento dell'attività di produzione primaria di informazione.

C. Ampliamento delle capacità di profilazione dell'utenza e personalizzazione dei contenuti di disinformazione a *target* specifici

Negli anni più recenti abbiamo assistito, da un lato, ad una crescente possibilità di avvalersi di tecnologie che consentono la trasformazione in dati digitali di qualunque elemento della vita economica, sociale e privata di una persona sia *online* che *offline* – cd processo di **datizzazione** – dall'altro lato, allo sviluppo vertiginoso della strumentazione tecnologica per la raccolta, l'elaborazione, la classificazione e il processamento dei *big data* sugli utenti, che, accedendo al *web* soprattutto attraverso *device* mobili, rilasciano in ogni momento una quantità inesauribile di dati e informazioni.

A questo si aggiunge il crescente impiego dell'intelligenza artificiale, e in particolare delle tecniche di *machine learning* e computazionali fondate sull'analisi psicometrica dei *big data* originati dagli utenti, che ha consentito di affinare le tecniche di profilazione, aumentando la possibilità di **personalizzazione dei contenuti**, compresi i messaggi pubblicitari, cui sono esposti gli utenti (cfr. *supra*, par. 3).

Inoltre, la tecnologia propria di internet consente non solo un monitoraggio del comportamento dell'utente nel momento in cui è esposto al messaggio pubblicitario (misurazione del contatto effettivo dell'utente con la pubblicità, cd. *impression*), ma anche il tracciamento delle relative azioni successive all'esposizione alla pubblicità, di interazione con il messaggio pubblicitario (es. numero di *click*, condivisioni) e fino all'atto di acquisto *online* o a un'altra azione ritenuta importante per l'inserzionista (*conversion*).

Tale accresciuta capacità di profilazione dell'utenza e personalizzazione dei messaggi pubblicitari, che risultano bene allineati alle preferenze dei destinatari, produce immediati vantaggi economici per tutti gli attori coinvolti nel sistema pubblicitario: per gli inserzionisti, interessati a raggiungere uno specifico *target* di persone, e, attraverso le tecnologie disponibili (ad *technology platform*, cfr. **Figura 4**), a somministrare la propria pubblicità mirata e personalizzata, anche in tempo reale; per gli utenti, che si presume, in base ai dati raccolti, abbiano il maggiore interesse ad acquistare il prodotto/servizio; per i *publisher*, che sono in grado di valorizzare il proprio inventario ottenendo un corrispettivo commisurato, a seconda delle diverse modalità di vendita, al raggiungimento dei diversi obiettivi (visualizzazioni, azioni - fra cui i *click* -, tempo speso, *etc*.); infine, per le piattaforme, coinvolte nel sistema di compravendita di pubblicità e in grado, così, di offrire servizi e informazioni aggiuntive, sia agli inserzionisti, sia ai *publisher* per il raggiungimento dei reciproci obiettivi, ricevendo un corrispettivo in funzione delle *impression* servite e/o delle azioni effettuate dagli utenti.

Con la diffusione dei *social network*, i servizi di personalizzazione delle campagne pubblicitarie si sono ulteriormente evoluti fino a giungere ad una forma di comunicazione pubblicitaria innovativa, per quanto controversa, detta ***dark advertising***.

Le campagne pubblicitarie diffuse sui *social network* che utilizzano *dark ads* si caratterizzano, infatti, per la somministrazione di messaggi pubblicitari estremamente mirati che sono **visibili solo alla categoria di utenti profilata** in base a specifici criteri (cd. pubblico bersaglio) restando, pertanto, del tutto invisibili agli altri utenti.

In altri termini, la disponibilità dei *big data* sugli utenti, unita a sofisticate tecniche psicometriche di profilazione dell'utenza, sopra descritte, ha permesso di sviluppare uno strumento di *marketing* all'interno dei *social media* molto potente, che è in grado non solo di personalizzare il messaggio pubblicitario, differenziandone il contenuto e adeguandolo alle specifiche caratteristiche del *target* selezionato, ma anche di renderlo visibile solo nelle *newsfeed* dello specifico gruppo di utenti, mentre altri gruppi non possono avervi accesso, ignorandone, addirittura, l'esistenza.

Una delle applicazioni principali del *dark advertising* è avvenuta nell'ambito delle campagne pubblicitarie elettorali diffuse attraverso i *social media* che saranno oggetto di approfondimento specifico nei paragrafi successivi (cfr. *infra*, par. 6). Vale qui ricordare, che, a partire dal caso dell'interferenza russa nelle elezioni statunitensi, l'attenzione della ricerca accademica, nonché delle inchieste giornalistiche si è concentrata sulle tecniche volte ad influenzare l'opinione pubblica durante le campagne elettorali e connaturate ai meccanismi di personalizzazione dei messaggi elettorali a pagamento diffusi attraverso le piattaforme (*social network*, motori di ricerca). In tale specifico contesto, la personalizzazione dei messaggi elettorali e le tecniche di *microtargeting* assumono particolare rilievo, ma il fenomeno dei *dark ads* assume, tuttavia, una portata più generale. Come evidenziato dalla letteratura scientifica sull'argomento, i meccanismi di **personalizzazione automatica** dei contenuti proposti agli utenti operata dai *social network*, che sono gli stessi utilizzati nella somministrazione dei messaggi pubblicitari, e le **azioni di condivisione** degli stessi compiute dagli utenti – insieme – favoriscono la creazione e diffusione di notizie false e la propagazione virale dei contenuti (cfr. *infra*, par. 6.2).

Con riferimento alle *dark ads*, più i messaggi pubblicitari diventano effimeri – la durata di una campagna pubblicitaria o di un *refresh* – e visibili solo in maniera altamente selettiva – da parte di un *target* bersaglio e non da altri – maggiore è il rischio che al loro interno vengano veicolati contenuti di disinformazione e che gli stessi passino inosservati, tenuto conto dell'assenza del controllo intersoggettivo altrimenti presente all'interno del *social network*[56]. Si osserva, infatti, come il prezzo dell'inserzione pubblicitaria, e quindi il guadagno della piattaforma di *social network* e degli altri intermediari coinvolti, vari in funzione della capacità del messaggio di influenzare il comportamento del destinatario, generando un *click* o favorendo un'azione (condivisione, acquisto, *etc.*). Sussiste, pertanto, un evidente incentivo economico, sia per la piattaforma di condivisione sociale, sia per gli inserzionisti, nel "confezionare" messaggi pubblicitari in modo da favorire il livello di *engagement* degli utenti ai quali sono destinati, ricercando argomentazioni – anche se ideologicamente ed emotivamente estreme – che possano suscitare la sensibilità o l'interesse degli stessi, influenzandone il comportamento. In altri termini, in linea con la letteratura di settore, il meccanismo intrinseco di funzionamento della pubblicità diffusa attraverso i *social network*, che si basa sulla presentazione dei contenuti pubblicitari in modo selettivo (personalizzazione) con l'intento di ottenere maggiore *engagement* e, pertanto, maggiori ritorni economici, insieme alle azioni di condivisione compiute dagli utenti più polarizzati, favorisce il formarsi di bolle ideologiche e fenomeni di disinformazione.

Pertanto, se a tale modello pubblicitario, che si nutre di personalizzazione e di azioni da parte degli utenti che favoriscono la propagazione virale dei contenuti, si uniscono contenuti visibili esclusivamente da gruppi di utenti selezionati (*dark ads*), il rischio che possano manifestarsi disturbi dell'informazione in assenza di adeguati livelli di trasparenza – circa gli inserzionisti, la tipologia di messaggi pubblicitari veicolati ai diversi *target* bersaglio – risulta ancora più elevato.

[56] Cfr. Nexa Center, *Persuasori Social, Trasparenza e democrazia nelle campagne elettorali digitali*, 30 maggio 2018.

In definitiva, anche con riferimento ai meccanismi di diffusione della pubblicità personalizzata sui *social network*, l'analisi condotta evidenzia l'assenza di adeguati livelli di trasparenza e l'esigenza di adottare delle misure concrete affinché l'utente sia messo nelle condizioni di comprendere tutti gli elementi che compongono la specifica campagna pubblicitaria cui è stato esposto.

L'analisi condotta nei precedenti paragrafi (punti A. B. e C.) ha mostrato come le trasformazioni intervenute nel settore pubblicitario, favorite dall'evoluzione tecnologica, abbiano contribuito alla creazione di un contesto ideale nel quale possono essere realizzate strategie di disinformazione con finalità commerciali, fondate sulla valorizzazione del traffico attraverso la vendita di spazi pubblicitari *online*.

In particolare, da un lato, il fatto che il sistema di compravendita di pubblicità *online* sia basato sulla ricerca dei *click* e sul *target* di utenti che verosimilmente effettuerà un'azione (*convertion*), dall'altro, l'automatismo dei meccanismi di negoziazione degli spazi pubblicitari ha favorito la proliferazione di operatori, piattaforme tecnologiche e intermediari, accrescendo la distanza fra domanda (inserzionisti) ed offerta (editori/*publisher*) di pubblicità.

Tale elemento ha un impatto diretto sulla proliferazione della disinformazione *online*, poiché favorisce un ambiente in cui vi è una scarsa capacità di controllo sull'intero processo di negoziazione da parte degli investitori pubblicitari che non sono in grado di conoscere sempre con esattezza – e in anticipo – in quali siti verranno visualizzati i messaggi pubblicitari.

Inoltre, il sistema di negoziazione di spazi pubblicitari *online* è diventato gradualmente più complesso e caratterizzato da sistemi automatici che, salvo diverse impostazioni, non presuppongono un contatto diretto fra inserzionista e *publisher*, bensì intermediato e re-intermediato da numerosi soggetti e articolato secondo schemi di compravendita pubblicitaria complessi e in continua evoluzione.

In questo contesto, l'interesse dei siti di disinformazione – ottenere traffico e vendere *inventory* in modo da intercettare quanto più volume di investimento pubblicitario possibile – trova perfetto allineamento con l'interesse delle piattaforme automatiche dal lato della domanda (DSP, *trading desk*, *ad server*) che – salvo diverse impostazioni – distribuiscono *impression* in funzione dei volumi di traffico e del *target* desiderato. Il rischio del sistema così configurato è favorire il finanziamento di siti che producono informazione di scarsa qualità, compresi quelli di disinformazione, distraendo le risorse pubblicitarie attraverso l'adozione di pratiche fraudolente.

Infine, la personalizzazione dei messaggi pubblicitari all'interno dei *social network* resa possibile dall'evoluzione tecnologica, dall'impiego dei *big data* sugli utenti e dall'utilizzo di tecniche di profilazione sempre più sofisticate, unite alle azioni degli utenti più polarizzati, favorirebbe il formarsi di bolle ideologiche e la creazione e proliferazione di fenomeni di disinformazione. Nel dettaglio, le campagne pubblicitarie sui *social* che si basano sui *dark ads*, caratterizzate da messaggi visibili solo in maniera altamente selettiva da parte di un *target* bersaglio e non da altri, risultano essere quelle maggiormente idonee a veicolare contenuti di disinformazione anche in considerazione del fatto che eventuali messaggi falsi non possono essere visualizzati dagli altri utenti, rischiando, pertanto, di passare inosservati.

Dal quadro appena descritto emergono **alcuni problemi di trasparenza** del sistema pubblicitario *online* in particolare in alcune fasi della filiera; opacità che viene sfruttata da organizzazioni, singoli utenti o gruppi, algoritmi automatici di selezione e personalizzazione dei contenuti con l'intento di incrementare traffico, *engagement* e realizzare guadagni pubblicitari derivanti dalla valorizzazione dei contatti e/o delle azioni ottenute.

Tale situazione richiede l'adozione di un **approccio multidisciplinare** che coinvolga istituzioni, attori del sistema e centri di ricerca sia per la comprensione di fenomeni complessi e in continua evoluzione che caratterizzano il sistema pubblicitario, sia per incoraggiare forme di autoregolamentazione degli attori finalizzate all'adozione di misure concrete a salvaguardia dei cittadini che rischiano, altrimenti, di essere esposti a contenuti di disinformazione per finalità commerciali.

5.2 Disinformazione e truffe *online*

All'interno delle strategie commerciali di disinformazione *online*, oltre a quelle basate sulla pubblicità online, esiste un ulteriore meccanismo di sfruttamento economico dei contenuti di disinformazione, basato sul **contributo diretto** degli utenti. Si tratta di campagne che sconfinano in condotte che possono essere rilevanti, oltre che sotto il profilo del pluralismo e della correttezza dell'informazione, anche sotto il profilo del diritto civile (ad esempio, possono sfociare in pratiche commerciali scorrette) e penale (è il caso delle truffe, dei casi in cui le false informazioni configurano ipotesi di distorsione del mercato e naturalmente dei casi più generali di diffamazione, procurato allarme e abuso della credulità popolare).

In generale, nel mondo *online*, specialmente nei mercati dei servizi di *e-commerce*, sono diffuse le condotte fraudolente a danno dei consumatori; l'ambiente di internet è, infatti, un ecosistema particolarmente adatto a questo genere di comportamenti criminosi, poiché garantisce l'accesso agevole a una serie di informazioni, un certo livello di anonimato e una capacità di azione in tempo reale.

In questo quadro, la diffusione di contenuti di disinformazione *online* può diventare uno strumento molto efficace per attirare i consumatori e risorse economiche, attraverso **raggiri** di vario tipo, che possono arrivare a configurare vere e proprie **truffe**. Gli elementi centrali che garantiscono efficacia a queste campagne sono la contagiosità dei contenuti proposti e la presenza di elementi, all'interno del contenuto di disinformazione, che sfruttano i *bias* cognitivi degli individui, in particolare proponendo il raggiro in un contesto che risulta familiare (il marchio contraffatto di una testata o la riproduzione della pagina *web* di una testata)[57].

Data la natura truffaldina, si tratta di campagne che si sviluppano nel breve periodo e orientate a ottenere un guadagno immediato; da questo punto di vista si può osservare che in tali strategie commerciali la sola motivazione sottostante sembra essere il reperimento di risorse economiche nel più breve tempo possibile. Cionondimeno, al di là dell'obiettivo contingente degli ideatori, queste condotte determinano un **impatto sul pluralismo e sulla formazione dell'opinione pubblica** soprattutto nel più lungo periodo, nella misura in cui, oltre a compromettere la credibilità e l'autorevolezza degli operatori dell'informazione, minano la capacità dell'utente di discernere le fonti di informazione autentiche.

In queste strategie, il contenuto di disinformazione costituisce un elemento di contorno necessario a richiamare l'attenzione dell'utente, a conferire credibilità all'iniziativa promossa e, quindi, ad attuare la condotta fraudolenta.

In particolare, l'ideatore della campagna utilizza contenuti *fake* per riprodurre artatamente un **contesto informativo noto** al consumatore e da questo ritenuto affidabile e, sfruttando tale credenza, all'interno del contesto falsato propone l'acquisto di un prodotto, ingannando il consumatore circa l'importo e/o l'autenticità della transazione stessa (**Figura 5**).

[57] Akerlof, G. A., and Shiller, R.J. (2015). *Phishing for phools: The economics of manipulation and deception.* Princeton, NJ: Princeton University Press.

Figura 5 - La disinformazione nel contesto delle truffe *online*

Dalla casistica osservabile ad oggi emerge come il prodotto di cui si promuove l'acquisto in genere appartiene a categorie come elettrodomestici, elettronica, informatica, telefonia, abbigliamento, e di solito il consumatore può ottenerlo mediante la partecipazione a finti concorsi o dietro rilascio di una serie di dati, compresi quelli della carta di credito.

Ciò che ingenera confusione nel consumatore e diventa determinante per la riuscita della campagna è proprio il contesto in cui è collocato l'oggetto dell'azione truffaldina, il quale è caratterizzato dall'inserimento di contenuti di disinformazione. A questo riguardo, i casi verificatisi evidenziano una varietà di modi di presentazione dei contenuti di disinformazione. Ad esempio, come illustrato più in dettaglio nel **Caso 3**, questi possono presentarsi come finti articoli di note testate giornalistiche di cui è replicato il *layout* e talvolta anche la firma di giornalisti conosciuti; attorno alla veste grafica, poi, si compone un testo breve con notizie roboanti, che può, in taluni casi, essere tradotto più o meno palesemente da un'altra lingua. Un altro modo di generare il contenuto di disinformazione è riprodurre una finta anteprima di una pagina *web* di una testata giornalistica oppure, attraverso un falso profilo *social*, diffondere dei *post* sponsorizzati che invitano all'acquisto di un prodotto a prezzi irrisori, sfruttando essenzialmente il nome della testata e il marchio. Da un *link* posizionato all'interno del finto articolo o che compare sulla finta pagina *web* o nel *post* il consumatore approda infine alla pagina malevola dove si finalizza la truffa.

Gli ideatori di questo tipo di strategie tendono a restare anonimi e utilizzano domini registrati adoperando identità false o rubate e molteplici falsi profili *social* le cui pagine, anche quando rimosse dalla piattaforma, vengono nuovamente ricreate continuando a sfruttare la fidelizzazione degli utenti che seguono la testata.

Uno dei punti di forza di queste strategie è la relativa semplicità ed economicità. Per ciò che riguarda le risorse adoperate, nella fase di creazione e produzione del contenuto è sufficiente un qualunque *software* che permetta di manipolare un marchio o di imitare il *layout* di una pagina di una testata. Per la distribuzione e la diffusione, poiché queste campagne si basano fondamentalmente sull'affidamento degli utenti nei riguardi di una testata e sulle possibilità di circolazione massiva e rapida dei contenuti attraverso le piattaforme *online*, è necessaria l'iscrizione a una piattaforma e la registrazione di un dominio, il che può avvenire utilizzando identità false o rubate; inoltre, i promotori di queste strategie possono servirsi di *post* sponsorizzati e stanziare un certo *budget* per promuovere il *post* che ospita il contenuto di disinformazione e consentirgli una circolazione mirata e più efficace. Gli investimenti necessari per attuare queste strategie restano, in ogni modo, piuttosto ridotti rendendole nel complesso **accessibili e remunerative** per i promotori.

CASO 3 – Truffe *online* che sfruttano contenuti di disinformazione

Alcune testate giornalistiche italiane sono state negli ultimi tempi le protagoniste di strategie ingannevoli finalizzate alla truffa che hanno fatto non solo uso improprio del nome e del marchio delle testate stesse, ma hanno altresì creato dei contenuti di disinformazione articolati, che si sono rivelati cruciali per catturare l'attenzione e la fiducia dell'utente e portare a buon fine la truffa.

Le campagne che saranno esaminate a seguire hanno interessato nello specifico la testata *online* Fanpage.it e la testata Repubblica.it.

Nel primo caso, il contenuto di disinformazione è stato costruito attorno al nome (lievemente modificato) e al marchio contraffatto di Fanpage.it ed è stato essenzialmente utilizzato un *post* sponsorizzato su Facebook per promuovere la diffusione del contenuto.
Questi, in breve, i fatti.
Nel febbraio 2018, la società Ciaopeople, editrice del giornale *online* Fanpage.it, tramite una segnalazione pervenuta da un proprio dipendente, viene a conoscenza dell'esistenza di un *post* sponsorizzato dalla testata, in cui è pubblicizzata la vendita di uno *smartphone* al prezzo irrisorio di 1 euro.
Il *post* sponsorizzato utilizza impropriamente il nome della testata e il marchio e risulta pubblicato da pagine Facebook create da ignoti che riportano nomi simili a Fanpage.

Il *link* posto all'interno del *post* rimanda a indirizzi *web* con nomi che possono parzialmente richiamare quello della testata (www. fanpagepromo. com) e all'interno delle pagine *web* che si aprono viene richiesto all'utente di inserire i dati della propria carta di credito per poter acquistare il prodotto pubblicizzato.
A marzo 2018, un *follower* della testata lamenta presso Ciaopeole di aver effettuato l'acquisto del prodotto – che risultava essere stato messo in palio da una azienda francese – e di aver visto prelevare dalla propria carta di credito una cifra, pari a 49,90 euro, superiore a quella pubblicizzata.
A seguito della segnalazione effettuata da Ciaopeople a Facebook, la piattaforma ha rimosso le pagine false, ma nonostante ciò nuove pagine venivano continuamente ricreate.
Questa pratica truffaldina è continuata per un periodo con modalità lievemente differenziate: nel *post* veniva utilizzato un nome simile a quello della testata (Fanpage.lt) e sono stati richiesti agli utenti diversi tipi di informazioni, dal nome all'*e-mail*, fino all'indirizzo e il codice postale.

Per ciò che riguarda il caso che ha interessato Repubblica.it, la strategia di disinformazione adottata ha puntato in particolare sulla creazione di articoli falsi, che riproducevano il *layout* autentico, diffusi mediante false pagine Facebook della testata con nomi simili a quello ufficiale ("LaReppublica.it"), oppure attraverso false pagine *web* della testata che sfruttano le finte anteprime della pagina della testata utilizzate dai social (*snippet*).

Negli articoli veniva annunciata la vendita da parte della Apple in Italia degli iPhone a 1 euro; dal *link* posto all'interno, poi, gli utenti venivano rimandati al sito truffa (www. coopaff. com).

la Repubblica.it

Ecco come gli Italiani avranno il nuovo iPhone X della Apple a solo 1€

14

Se vivete in Italia e volete il nuovissimo iPhone X, allora questo potrebbe essere l'articolo più entusiasmante che abbiate mai letto.
Ecco come: la Apple Inc. (NASDAQ: AAPL) offre il suo nuovissimo telefono iPhone X agli Italiani a solo 1€. Si, avete letto bene: 1€

26 Aprile 2018

1' di lettura

Tutto ciò fa parte di una strategia di marketing, la Apple sta lavorando con RUniverse , suo partner affidabile, per offrirvi i cellulari iPhone X più convenienti del 99% rispetto al prezzo di vendita al pubblico.

Infine, come a Ciaopeople anche a Repubblica è accaduto che circolassero *post* sponsorizzati facenti capo alla pagina Facebook di "LaReppublica" che pubblicizzavano l'acquisto dello *smartphone* a 1 euro e reindirizzavano a un sito malevolo.

I casi presentati, al di là del profilo di interesse legale, mostrano come i contenuti di disinformazione *online* siano uno strumento estremamente efficace per indurre in errore il consumatore, riducendo il suo livello di attenzione critica. Inoltre, si conferma la persistenza del fenomeno, poiché, infatti, queste campagne (a basso costo) si sono dimostrate difficili da contrastare nonostante l'impegno da parte delle stesse piattaforme *online* coinvolte. Infine, l'associazione da parte del consumatore tra l'offerta commerciale truffaldina pubblicizzata e la reputazione della testata, nel più lungo periodo, comporta che tali strategie rischiano di minare la credibilità e l'autorevolezza degli operatori dell'informazione e più in generale di compromettere la capacità dell'utente di riconoscere le fonti di informazione autentiche.

Sul piano delle **possibili soluzioni**, l'analisi della disinformazione *online*, utilizzata nell'ambito di azioni fraudolente ai danni dei consumatori, suggerisce come questo tipo di strategie richiedano un intervento specifico, oltre gli strumenti di tutela già disponibili.

L'azione di contrasto, esercitata nei singoli casi mediante l'applicazione del Codice del consumo nonché delle altre norme civili e penali, infatti, è uno strumento necessario ma presenta delle limitazioni nel contesto della disinformazione *online*. Da un lato i contenuti di disinformazione, attorno ai quali viene elaborato il raggiro, si diffondono molto rapidamente tra i consumatori rendendo queste strategie difficili da perseguire, molto efficaci e di rapida esecuzione; dall'altro lato, oltre al **danno economico** ai consumatori e **all'immagine** delle testate associate ai contenuti di disinformazione, creano anche un effetto ulteriore e più ambiguo, poiché nel lungo periodo contribuiscono a creare sul *web* un ambiente in cui l'utente ha **difficoltà a riconoscere le fonti di informazione autentiche**, con effetti, dunque, sulla correttezza dell'informazione e sul pluralismo.

Di conseguenza, accanto ai meccanismi di tutela già esistenti, sarebbe opportuno predisporre un altro tipo di intervento, che abbia l'obiettivo specifico di colpire la disinformazione, fondato sia su un'azione preventiva che innalzi il livello di attenzione critica degli utenti in rete, sia su strumenti che aiutino i consumatori nel riconoscimento delle fonti di informazione autentiche.

5.3 Disinformazione commerciale

Un'altra modalità, molto diffusa sul *web*, di sfruttare la disinformazione per scopi economici è rappresentata da campagne di disinformazione commerciale che diffondono *online* **informazioni false su prodotti e aziende**. I rischi che da queste possono derivare sono molteplici e investono non solo le imprese, ma anche i consumatori e possono arrivare a condizionare le scelte di interi gruppi, con effetti su tutta la collettività.

L'alterazione del *set* informativo disponibile agli individui, infatti, rappresenta un problema dal punto di vista del **benessere sociale**. L'informazione costituisce una risorsa fondamentale per tutti gli agenti economici che operano sui mercati, nella misura in cui contribuisce alla formulazione delle scelte di consumo e di quelle di produzione. La sua incompletezza e la sua distribuzione asimmetrica tra i diversi attori nei mercati, infatti, producono delle inefficienze sia nella formazione dei prezzi, nelle quantità e nella qualità dei beni e servizi scambiati in equilibrio sul mercato, sia nella distribuzione del benessere tra le imprese e i consumatori. A questi effetti generali, si aggiungono, inoltre, le conseguenze negative non strettamente economiche generate dalle imperfezioni informative che investono altri aspetti della vita degli individui (la salute, la sicurezza, l'ambiente, *etc.*).

In questa prospettiva, internet rappresenta un ambiente potenzialmente ricco di informazioni che può contribuire notevolmente al miglioramento dei processi decisionali di tutti gli operatori economici e correggere sensibilmente le imperfezioni di natura informativa. D'altra parte, il *web*, proprio per la sua struttura fortemente decentralizzata e globale, e in virtù della presenza di una molteplicità di contenuti, costituisce un ambito in cui è – quantomeno – più **complessa** e più **costosa** sia la ricerca di informazioni da parte degli utenti, sia la scrematura di quelle rilevanti e attendibili. Ciò implica che sui mercati *online* possono permanere, e perfino complicarsi, le imperfezioni dell'informazione, che finiscono con l'assumere in rete maggior rilievo per la loro capacità di diffondersi su scala globale, di **resistere nel tempo** e di coinvolgere una **pluralità di mercati**.

Da questo punto vista, le strategie che diffondono contenuti di disinformazione *online* relativamente a specifici prodotti non solo colpiscono la **reputazione** delle aziende e lo **scenario competitivo**, fino a configurare atti di concorrenza sleale veri e propri, ma soprattutto inducono il **consumatore in confusione**, impedendo in tal modo una corretta informazione e la corretta formazione di un'opinione sui prodotti e sulle loro caratteristiche, determinando così una diminuzione del benessere. Esse, inoltre, sono in grado di produrre effetti su platee di consumatori molto vaste e di interessare, almeno potenzialmente, tutti i mercati presenti su internet e quindi il sistema economico nel suo complesso; infine possono alterare le conoscenze e i comportamenti degli individui su temi rilevanti socialmente, quali la sicurezza dei prodotti, la salute, l'ambiente.

Con riguardo alle caratteristiche della disinformazione commerciale *online*, questa appare un fenomeno poliedrico per quanto riguarda i soggetti ideatori, le loro motivazioni e le modalità con cui esso si manifesta.

In generale, si osserva che il contenuto messo in circolazione sul *web* viene progettato come una combinazione tra una comunicazione commerciale e un contenuto di natura informativa sulla base di elementi veritieri, oppure può essere un contenuto manipolato completamente (vedi **Caso 5**). In ogni modo la componente visiva del messaggio è preponderante, per consentire al destinatario di associare immediatamente il messaggio al prodotto e al marchio oggetto della campagna. Attorno al prodotto e/o al marchio viene elaborato il contenuto di disinformazione, che in genere consiste nel proporre un breve testo in cui viene riportata una notizia *shock* sulle proprietà o sulle caratteristiche del prodotto e sulle sue conseguenze per il consumatore, oppure viene creata una falsa notizia sul prodotto – più articolata – in cui è citata una fonte riconoscibile come autorevole (ad esempio, un sedicente studio scientifico o un'autorità pubblica) cui si attribuisce la notizia stessa.

Ad essere colpito è spesso il marchio, o uno specifico prodotto finale o anche una materia prima. Secondo uno studio effettuato da Centromarca su un campione di 46 imprese industriali del settore *grocery* (35 alimentari e 11 non alimentari), da settembre 2016 a marzo 2018, 22 aziende sono state protagoniste di "crisi" da contenuti *fake*. I risultati hanno evidenziato, in particolare, che il problema ha interessato in misura superiore le aziende del settore alimentare (16 su 22) e che i contenuti *fake* sono stati veicolati in special modo attraverso le piattaforme *online*[58].

In altri casi, la campagna non prende di mira un prodotto o un marchio preciso, ma piuttosto, sulla base di informazioni false, parziali o fuorvianti, indirizza l'opinione pubblica verso l'adozione di condotte (per esempio alimentari) che premiano il consumo di alcuni prodotti a scapito di altri.

Gli ideatori di queste strategie possono essere imprese che perseguono obiettivi di natura economica, poiché mirano a gettare discredito sui concorrenti, con effetti immediati sulla reputazione che nel più lungo periodo possono portare anche ad una diminuzione delle vendite, fino a produrre effetti negativi sulla stabilità economico-finanziaria dell'impresa e di conseguenza sullo scenario competitivo. D'altra parte, talune campagne possono non avere natura strettamente economica ma piuttosto possono rappresentare strategie ibride in cui coesistono anche delle motivazioni ideologico-politiche; in questo caso, i promotori della campagna possono essere soggetti diversi, organizzazioni a sfondo ideologico oppure anche stati esteri che utilizzano la disinformazione come strumento nell'ambito di vere e proprie guerre commerciali (cfr. *infra*, par. 6).

Per ciò che riguarda gli strumenti tecnologici, in particolare nella fase di distribuzione, le strategie di disinformazione commerciale utilizzano essenzialmente tecnologie e relativi servizi delle piattaforme di *social networking* e altresì delle applicazioni di *instant messaging*, veicoli privilegiati per ottenere una diffusione rapida e massiva. Nel processo di diffusione, infatti, appare determinante il meccanismo del **"passaparola"**, che rinforza progressivamente la credibilità del messaggio; ciò che, infatti, risulta decisivo è la **partecipazione attiva** dello stesso consumatore al processo di diffusione.

È utile notare, infine, che in queste strategie la fase di valorizzazione dei contenuti è caratterizzata da un meccanismo di recupero nel lungo periodo; i ritorni per gli ideatori, infatti, sono rilanciati nel tempo e dipendono essenzialmente da quanta eco la campagna ha ottenuto tra il pubblico, da quanto in profondità ha intaccato il rapporto di fiducia tra azienda e consumatore e da quanto ha inciso, infine, sulle scelte effettive di consumo degli utenti.

[58] Cfr. lo studio di Centromarca su *fake news* e industria di marca presentato al seminario AGCOM dell'11 giugno 2018 "Gli effetti della disinformazione commerciale sulle scelte dei consumatori".

CASO 5 – Disinformazione commerciale: il caso del settore alimentare

Le strategie di disinformazione commerciale fondano il loro successo sull'individuazione di temi sensibili per i consumatori e sulle modalità di presentazione del messaggio mediante i contenuti *fake* di cui si servono. L'elemento centrale è suscitare l'emotività e le preoccupazioni più profonde dei consumatori per attrarne l'attenzione e promuovere, in tal modo, la propagazione dei contenuti di disinformazione. Questa logica, peraltro comune anche agli altri fenomeni di disinformazione, e soprattutto quelli con motivazioni politico-ideologiche (cfr. *infra*, par. 6), è stata applicata a diversi prodotti, mescolando, talvolta, questioni e motivazioni di natura economica con quelle ideologiche (ad esempio nelle campagne contro l'olio di palma e il latte vaccino).

In diversi casi le strategie di disinformazione commerciale hanno fatto leva sul tema - molto sentito - della salute, prendendo di mira in maniera più o meno diretta specifici marchi e/o aziende del settore alimentare (ad esempio la campagna di disinformazione sulla ripastorizzazione del latte scaduto). Il caso esaminato brevemente nel seguito può essere considerato un tipico esempio di questo genere di disinformazione e riguarda le vicende che hanno interessato l'azienda Mutti.

A novembre 2017 si diffonde tra gli utenti, prima tramite Whatsapp poi tramite *post* su Facebook, la notizia che il Ministero della Salute avrebbe emanato un atto di richiamo prodotti nei riguardi l'azienda Mutti per aver trovato tracce di arsenico in un lotto di produzione della passata di pomodoro dell'impresa. Il contenuto *fake* diffuso si presenta come un falso documento che riproduce un modulo di richiamo prodotti del Ministero, in cui compaiono una serie di dettagli sul prodotto, il motivo del richiamo, le avvertenze, le foto del prodotto.

Nei giorni immediatamente successivi alla circolazione del contenuto, dopo la denuncia alla Polizia postale, l'impresa e lo stesso Ministero della Salute smentiscono la notizia attraverso i rispettivi canali ufficiali. L'azione dell'azienda per far rientrare la "crisi" si prolunga per più di un mese con comunicati stampa sui siti *web*, smentite sulle piattaforme Facebook, Whatsapp e Twitter, comunicati sui principali quotidiani nazionali e all'interno di programmi televisivi, *etc*.

La campagna di disinformazione che ha investito la Mutti non sembra aver avuto conseguenze sui risultati economici di *business* dell'impresa, sebbene abbia comportato investimenti da parte dell'azienda per gestirne le conseguenze; tuttavia l'aspetto rilevante che rende questo caso emblematico è che, benché la notizia fosse stata prontamente smentita, anche dalle autorità, la combinazione tra la logica di funzionamento dei *social network* e una limitata capacità di attenzione ne ha comunque determinato la diffusione virale.

Tale efficacia della disinformazione commerciale può diventare molto problematica quando, oltre a investire la reputazione di un'azienda, provoca ricadute sugli utenti e sulla collettività in generale; a questo proposito uno dei temi spesso oggetto di campagne di disinformazione è proprio quello della salute; è il caso, ad esempio della disinformazione in campo sanitario, quando ad essere oggetto di strategie di disinformazione sono specifici farmaci o protocolli di cura.

L'analisi della disinformazione commerciale *online* evidenzia un **fenomeno complesso**, che in alcuni casi può configurarsi anche come concorrenza sleale o come forma di pubblicità ingannevole; più spesso, al di fuori di tali situazioni nettamente distinguibili, la disinformazione commerciale risulta piuttosto difficile da inquadrare (ad esempio può essere espressione di una guerra commerciale, di una strategia a sfondo politico-ideologico, *etc.*) e l'individuazione da parte degli utenti appare problematica. Il contenuto di disinformazione, infatti, è progettato per colpire l'emotività degli utenti, la quale assume un peso determinante anche rispetto ai dati fattuali e il suo effetto si rafforza man mano che la falsa notizia si diffonde. In questa ottica, gli interventi *ex post* di ripristino della correttezza delle informazioni – quando possibili – hanno scarso, se non nullo, effetto sui consumatori. Inoltre, in tutti i casi in cui le strategie di disinformazione si collocano in una zona grigia tra strategie commerciali e strategie politico-ideologiche (cfr. *infra*, par. 6), le conseguenze prodotte sulle credenze e sui comportamenti degli utenti investono temi di rilevante interesse pubblico come la salute e l'ambiente, con **risvolti non solo di carattere economico ma anche sociale**.

Il quadro che emerge, dunque, suggerisce l'opportunità di adottare specifiche **soluzioni per il contrasto** ai fenomeni di disinformazione commerciale, che potrebbero coinvolgere sia direttamente le imprese che offrono i propri beni e servizi *online*, e quindi investono in comunicazione *online*, sia gli stessi consumatori, con l'obiettivo duplice: promuovere forme di comunicazione commerciale *online* veritiere e corrette, favorendo

l'isolamento di tutti i contenuti di disinformazione commerciale così da evitare che si finanziano attraverso la pubblicità, nonché migliorando il riconoscimento degli stessi da parte degli utenti anche attraverso il potenziamento delle iniziative di autoregolamentazione a tutela dei consumatori che offrano strumenti concreti per la riconoscibilità della comunicazione commerciale digitale[59]. A tal proposito un ruolo importante lo possono svolgere delle forme di autoregolamentazione efficaci nel garantire un elevato livello di tutela dei consumatori[60].

6. Le strategie di disinformazione *online* di natura politico-ideologica

Le strategie di disinformazione *online* di natura politico-ideologica sono caratterizzate, così come quelle viste in precedenza, dall'analisi del *target*, dall'individuazione di specifici temi ad alto contenuto emotivo, e dalla scelta di determinati codici comunicativi e *framing* nella progettazione del messaggio (fase I di creazione); dalla trasposizione in un prodotto informativo, dal messaggio all'immagine (fase II della produzione), dalla scelta del canale distributivo e dal lancio del contenuto (fase III di distribuzione). Si differenziano, invece, dalle precedenti nella **fase IV della valorizzazione**, che in questo caso non è infatti di natura economica: tali iniziative di comunicazione, condotte da organizzazioni di varia natura, (partiti politici, Stati e governi), sono piuttosto guidate da **motivazioni di natura ideologica o politico-elettorale**, e hanno un forte impatto, per tematiche, tempistiche e numeri, sul pluralismo e sulla correttezza dell'informazione suscettibile di produrre effetti concreti sulle scelte degli utenti-cittadini sia sotto il profilo della formazione delle preferenze sia per quanto riguarda il rafforzamento di preferenze polarizzate pre-esistenti.

Sebbene anche in ambito politico-ideologico sia possibile far riferimento a *fake news* in relazione ad esempio a quelle notizie "*provenienti da fonti che forniscono, in maniera tendenziosa, informazioni a supporto di specifici punti di vista e orientamenti politici*[61]", alle "*falsità coscientemente distribuite per minare un candidato o orientare una competizione elettorale*", alle "*notizie che si oppongono da un punto di vista ideologico*", alle "*notizie che sfidano l'autorità costituita*[62]" , ovvero alle false contestualizzazioni, ai contenuti manipolati o fabbricati ad arte o ingannevoli per motivi di influenza politica[63], questo ampio ventaglio di casistiche non può certamente rientrare nell'oggetto di questo studio, visto che si tratta di azioni che non sempre appaiono strutturate come condotte strategiche di lungo periodo e cioè sistematiche, pervasive, ripetute.

In particolare, queste azioni sfruttano quegli stessi meccanismi di funzionamento delle piattaforme *online* (algoritmi di ricerca o *news feed*), che favoriscono la nascita di ***echo chambers***, caratterizzate da individui che discutono solo all'interno di una cerchia di persone ideologicamente vicine[64], ricalcando e acuendo i problemi di esposizione selettiva e *confirmation bias*[65] e, soprattutto, la **polarizzazione ideologica** dei cittadini, con le relative conseguenze sulla formazione dell'opinione pubblica[66].

Come rilevato dal recente Rapporto dell'Autorità sul consumo di informazione, l'analisi della relazione sussistente tra la polarizzazione ideologica degli utenti dei *social network* e le loro attività in rete ha mostrato

[59] Fra le iniziative volte ad assicurare l'onestà, la correttezza e la veridicità della comunicazione commerciale digitale, si ricorda la Digital Chart adottata dall'Istituto dell'Autodisciplina pubblicitaria (IAP) che si prefigge una ricognizione delle forme di comunicazione commerciale digitale e la definizione dei criteri di riconoscibilità delle stesse nel rispetto dell'art. 7 del Codice di autodisciplina della Comunicazione Commerciale (C.A.). Cfr. IAP, Digital Chart.

[60] In questa direzione va l'iniziativa intrapresa dall'Istituto dell'Autodisciplina Pubblicitaria (IAP) con la "Digital Chart" che offre strumenti concreti in relazione al tema della riconoscibilità della comunicazione commerciale digitale.

[61] Zimdars M., *False, op. cit.*

[62] Tambini, D. (2017), *How advertising fuels fake news*, London: Media Policy Project Blog, London School of Economics and Political Science.

[63] Wardle C., *Fake news, op. cit.*. Per una problematizzazione delle diverse sfere semantiche legate all'espressione "fake news" si veda Caplan R., Hanson L., Donovan J. (2018), Dead Reckoning. Navigating Content After "Fake News", Data&Society.

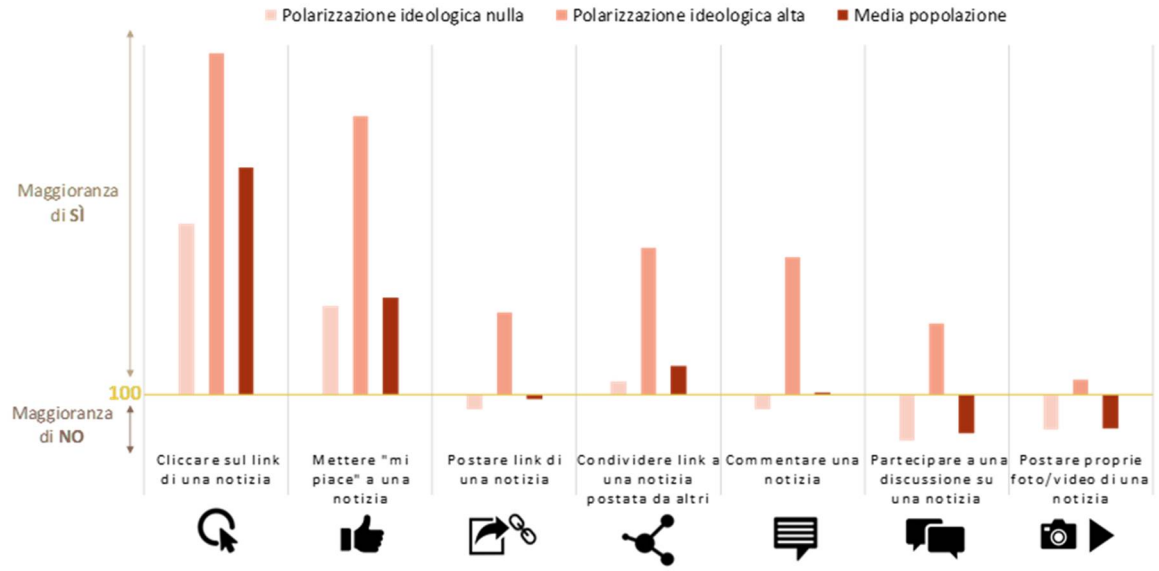
[64] Cfr. Quattrociocchi, W., Scala, A. & Sunstein, C., *Echo Chambers on Facebook, op. cit.*

[65] Cfr. Flaxman S., Goel S., Rao J.M. (2016). "Filter Bubbles, echo Chambers, and Online News Consumption". *Public Opinion Quarterly*, 80(1), pp. 298–320.

[66] Come rilevato anche da Nicita A., *È possibile il libero scambio nel mercato della verità?*, Il Foglio, 31 gennaio 2017.

come la polarizzazione possa avere un effetto significativo sul maggior impegno (*engagement*) nei confronti delle notizie divulgate dai *social network* (**Figura 6**).

Figura 6 - Rapporto tra azioni informative svolte sui social network e polarizzazione ideologica (2017)



Fonte: Elaborazioni Autorità su dati GFK Italia

Il legame tra lo svolgimento di tutte le **azioni informative sui *social network***, incluse quelle azioni a più alto tasso di coinvolgimento dell'utente (ad es. postare proprie foto e video di una notizia e partecipare alla discussione su una notizia) e la **polarizzazione** ha evidenti riflessi sul concretizzarsi di fenomeni di diffusione di posizioni radicalizzate e creazione di bolle ideologiche, che a loro volta possono produrre casi di *hate speech*[67], con una doppia causazione circolare cumulativa tra notizie false e discorsi d'odio, nella quale la polarizzazione si trasforma in **estremizzazione**[68].

Più gli utenti dei *social network* diffondono contenuti orientati ideologicamente, più le persone coinvolte nella stessa *echo chamber* confermano i propri pregiudizi (*confirmation bias)* reagendo con contenuti offensivi o comunque lesivi della dignità di un gruppo di persone, producendo campagne di odio contro individui specifici e il gruppo (etnico, religioso, di genere, ecc.) da esso/essa rappresentato – caratteristica tipica, come vedremo, delle strategie di disinformazione *online* polarizzanti di natura ideologica.

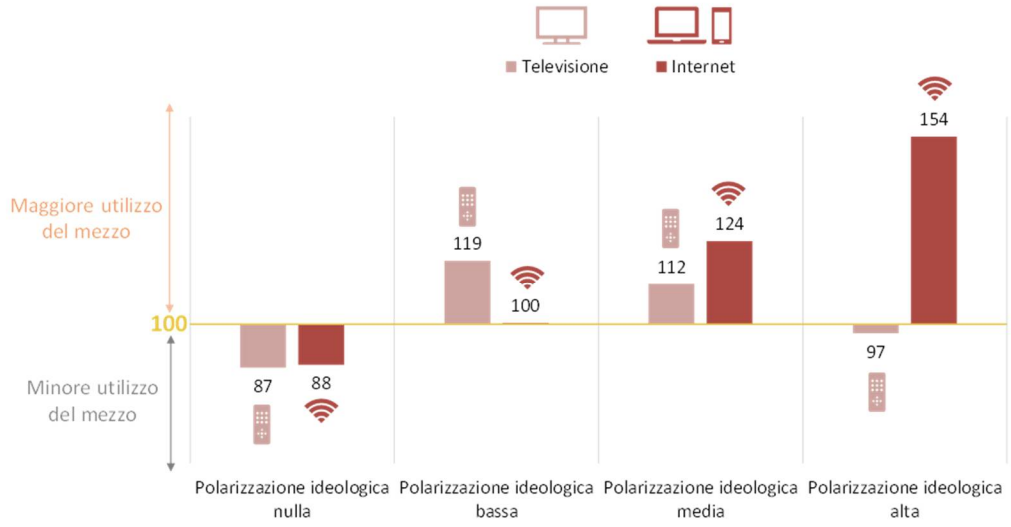
Chiaramente, queste dinamiche hanno un importante **riflesso anche sulle strategie di disinformazione politica basate sulla manipolazione a fini elettorali**. Esaminando il legame tra i mezzi informativi utilizzati

[67] L'*hate speech* – espressione spesso confusa con "l'incitamento all'odio" – è una categoria ben precisa, anche sotto il profilo giuridico, elaborata negli anni dalla giurisprudenza americana, in confronto con il *freespeech*, per indicare un genere di parole e discorsi che non hanno altra funzione a parte quella di esprimere e sollecitare odio e intolleranza verso una persona o un gruppo, e che rischiano di provocare reazioni violente (*hate harm*) contro quel gruppo o da parte di quel gruppo. Nel linguaggio ordinario indica più ampiamente un genere di offesa fondata su una qualsiasi discriminazione (etnica, religiosa, di genere o di orientamento sessuale) ai danni di un gruppo. Secondo la Commissione europea contro il razzismo e l'intolleranza del Consiglio d'Europa (ECRI), l'incitamento all'odio può essere definito come "promozione o incitamento, in qualsiasi forma, di denigrazione, odio o diffamazione di una persona o di un gruppo di persone, nonché qualsiasi molestia, insulto, stereotipizzazione negativa, stigmatizzazione o minaccia nei confronti di tale persona o gruppo di persone e la giustificazione di tutti i precedenti tipi di espressione, sulla base di razza, colore, provenienza, nazionalità, etnia, età, disabilità, lingua, religione o convinzioni personali, sesso, genere, identità di genere, orientamento sessuale e altre caratteristiche o status personali"(cfr. ECRI General Policy Recommendation No. 15 on Combating Hate Speech adopted on 8 December 2015, trad. nostra). L'Autorità per le garanzie nelle comunicazioni ha avviato, con la Delibera n. 403/18/CONS, un procedimento finalizzato all'adozione di un regolamento in materia di rispetto della dignità umana e del principio di non discriminazione e di contrasto all'*hate speech* e all'istigazione all'odio, nelle radio e nelle tv.

[68] Sunstein, C.R. (2009). *Going to extremes: How like minds unite and divide.* Oxford: Oxford University Press.

per formare le proprie scelte politico-elettorali e il livello di polarizzazione ideologica, al fine di verificare l'esistenza di eventuali *trend* di consumo informativo a fini politico-elettorali da parte dei cittadini più o meno ideologicamente schierati, è stata svolta, nell'ambito del Rapporto sul consumo dell'informazione, un'analisi circa la correlazione tra livello di polarizzazione ideologica dei cittadini (intesa nel senso qui specificato) e l'accesso medio ai due mezzi per informarsi. Nello specifico, dalla **Figura 7** si evince come le differenze tra l'utilizzo delle due principali fonti informative per livello di polarizzazione ideologica emergano in particolare per internet, non per la televisione: gli individui politicamente più attivi, e quindi più schierati anche dal punto di vista ideologico, ricorrono in maniera piuttosto ampia ad internet come mezzo di comunicazione per informarsi sulle scelte politico-elettorali.

Figura 7 - Indice di differenziazione tra l'uso dei mezzi di comunicazione per le scelte politico-elettorali, per livello di polarizzazione ideologica dei cittadini (2017; popolazione 18 anni e più)



Fonte: Elaborazioni Autorità su dati GFK Italia

Ne risulta un quadro in cui la **polarizzazione opera già a livello di selezione del mezzo**, per poi "viralizzarsi" a seguito delle **azioni compiute sui *social network*** dagli utenti più attivi e del concomitante operare di **algoritmi personalizzati**. Questi ultimi, in particolare, appaiono favorire l'emergere di bolle ideologiche, con le relative conseguenze sulla formazione di un dibattito democratico e informato sui *social network*, e, di rimando, in tutte le sfere in cui si articola l'opinione pubblica, sulla formazione della decisione di voto e, in generale, sulla dimensione sociale e politica della vita di ogni individuo.

Nonostante i molti tratti comuni, relativi anche alle dinamiche di diffusione dei contenuti su cui si basano le relative condotte strategiche, può risultare opportuno distinguere – per tipo di contenuti, attori coinvolti, finalità, natura e modalità di distribuzione dei messaggi, ovvero risorse tecnologiche utilizzate – tra disinformazione a carattere ideologico e disinformazione a carattere politico.

6.1. Le strategie di disinformazione *online* di natura ideologica

Le strategie e le campagne di disinformazione *online* **di tipo ideologico** sono caratterizzate, per quanto riguarda la creazione e la produzione dei contenuti, da messaggi che provocano un'immediata reazione emotiva, anche inconsapevole e istintiva[69] con un'importante componente visiva una forte componente

[69] Kahneman, D. (2012). *Pensieri lenti e veloci*. Milano: Mondadori.

narrativa (*framing*) e; tali messaggi seguono lo *storytelling* del gruppo/organizzazione, sono chiaramente ripetuti in un arco temporale piuttosto ampio e tendono, infine, sistematicamente e deliberatamente a sottolineare e ad alimentare le differenze e le divisioni, con riferimento ad esempio a temi quali la difesa della nazionalità, dell'identità etnica, o della religione di cui la rispettiva ideologia si fa portatrice[70] rispetto a 'minacce' concrete provenienti da gruppi di individui o soggetti politico-istituzionali o religiosi individuati come target e rispetto alla cui azione viene sistematicamente alimentato un sentimento di paura, disgusto, pericolo.

Tra i **soggetti promotori ed ideatori** di queste strategie di disinformazione, si possono certamente riconoscere gruppi di odio e ideologici (ad es. suprematisti della razza o dell'uomo bianco, nazionalisti, ecc.) e relativi personaggi-chiave o *leader* carismatici[71], teorici della cospirazione (ad es. antisemiti revisionisti storici), gruppi di cittadini impegnati su determinante tematiche (ad es. gruppi informali, gruppi *single-issue*[72] o associazioni locali accomunate da un atteggiamento di tipo NIMBY[73] nei confronti di opere pubbliche). In tutti i casi, si tratta pertanto di gruppi o organizzazioni con forte motivazione ideologica, che si contraddistinguono dunque per porre in essere campagne non di breve periodo.

Dal punto di vista della **distribuzione del contenuto**, queste campagne comunicative si caratterizzano per: volume e multi-canalità dell'azione informativa; rapidità, continuità e ripetitività nella diffusione dei messaggi (da parte dei cd. *troll*); presenza di gruppi *networked*[74] e dalla struttura agile.

Gli strumenti che le organizzazioni protagoniste di campagne di disinformazione ideologica – veri e propri *network* caratterizzati dalla condivisione di una medesima subcultura – utilizzano sono ambienti digitali di cultura partecipativa, comunità caratterizzate da barriere all'ingresso relativamente basse per la libera espressione e per forme di impegno civico da parte di qualunque utente, una forte supporto per la creazione e condivisione di contenuti generati dagli utenti, e alcune forme di tutoraggio virtuale da parte di utenti con maggiore esperienza[75]. Tra le **tecniche espressive** maggiormente utilizzate da questi *network* spiccano certamente quelle di amplificazione strategica e di *framing* – mutuate dall'universo delle pubbliche relazioni, che coinvolgono fonti di informazione locale o specializzate e *influencer* anche con lo scopo di raggiungere i media tradizionali –, ovvero i cd. *meme*, messaggi (spesso con preponderante componente visiva) caratterizzati da stile e linguaggio riconducibili alla dimensione satirico/parodistica o comunque scherzosa, ma anche da una forte capacità di influenza attraverso la trasmissione *online* che li rende oggetti di comunicazione socialmente condivisa[76].

---

[70] Cfr. Wardle C., Derakhshan H. (2017), *Information disorder: Toward an interdisciplinary framework for research and policy making*, *op. cit.*, pp. 29-38.

[71] Il ruolo dei leader carismatici è stato messo in evidenza anche da alcuni studi sulle strategie di disinformazione scientifica. Ad esempio, Sara e Jack Gorman analizzano nello specifico il ruolo di alcuni leader carismatici del negazionismo scientifico – Andrew Wakefield, che per primo ha messo in connessione autismo e vaccini, il negazionista del virus HIV Peter Duesberg, la paladina del movimento anti-vaccini Jenny McCarthy, la leader del movimento anti-OMG Gilles-Eric Sérafini (cfr. Gorman S. E., Gorman J.M. (2017). *Denying to the Grave.* Oxford: Oxford University Press).

[72] Secondo un linguaggio mutuato dalla scienza politica (Sartori G. (1976), *Parties and Party Systems. A framework for analysis*, Cambridge: Cambridge University Press), possiamo intendere come *single-issue* quei gruppi, spesso a carattere (iper)locale, caratterizzati dalla proposta di e dal tentativo di far entrare nell'agenda del pubblico una specifica questione o un singolo tema di interesse del gruppo o della comunità locale.

[73] Sull'uso dell'acronimo NIMBY (Not In My Back Yard) nelle scienze sociali, cfr. Borell K., Westermark A. (2018), "Siting of human services facilities and the not in my back yard phenomenon: a critical research review", in *Community Development Journal*, 53, 2 (1), pp. 246-262.

[74] Nel senso attribuito al termine da Rainie L., Wellman B. (2012), *Networked. The New Social Operating System*, Boston: MIT Press, trad. it. *Networked. Il nuovo sistema operativo sociale*, Roma: Guerini, 2012.

[75] Cfr. Jenkins H. (2006), *Convergence Culture*, New York: New York University Press, trad. it. *Cultura Convergente*, Milano: Apogeo, 2007.

[76] Cfr. Marwick A., Lewis R. (2017), *Media Manipulation and Disinformation Online*, Data & Society pp. 34-39. Anche la tendenza ai toni sarcastici e cinici e il gusto tipicamente post-moderno per l'ironia rientrano tra le cifre stilistiche delle strategie comunicative dei *troll* che animano i *network* di disinformazione ideologica (Kakutani M. (2018), *The Death of Truth. Notes on Falsehood in the Age of Trump*, New York: Tim Duggans Books, trad.it. *La morte della verità. La menzogna nell'era di Trump*, Milano: Solferino, 2018).

I *meme*, così come altre forme di comunicazione utilizzate da gruppi e organizzazioni ideologicamente o politicamente orientate, devono chiaramente essere contraddistinte da una forte diffondibilità, ovvero capacità del messaggio di propagarsi attraverso le reti di comunicazioni sociali[77], che rende possibile una continua riproduzione del messaggio in prodotti mediali, anche generati dagli utenti.

Seguendo questo **schema circolare di diffusione**, tali messaggi riescono a ricevere anche una forte copertura da parte dei media tradizionali, influenzando quindi l'agenda del pubblico, secondo il noto principio dell'*agenda-setting*, per il quale l'aumento della copertura mediatica su specifiche tematiche influenza la presunta importanza delle stesse tra il pubblico[78].

Per diventare diffondibili e **influenzare l'agenda dei media**, questi messaggi devono in particolare rispettare una serie di euristiche, o scorciatoie mentali, che gli utenti dei *social media* sono soliti utilizzare quando valutano la credibilità di una fonte o di un messaggio, quali la reputazione, la credibilità, la consistenza, la violazione dell'aspettativa, la conferma delle proprie credenze, l'intento persuasivo[79]. Inoltre, nel caso specifico di messaggi facenti parte di campagne di disinformazione ideologica, conta particolarmente il riferimento a tematiche in grado di polarizzare il pubblico[80].

Dal punto di vista, infine, della **valorizzazione**, l'intento di queste campagne di disinformazione si esplica, nella diffusione di un'ideologia, nella radicalizzazione di chi già condivide determinate idee e convinzioni, negli attacchi personali nei confronti di personaggi e *opinion leader* che portano avanti e condividono pubblicamente posizioni opposte o ritenute discordanti.

CASO 6 - Disinformazione ideologica: GamerGater, Far Right e scie chimiche

Tra le principali campagne che si fondano su strategie di disinformazione ideologica aventi tutte le caratteristiche finora evidenziate, possiamo citare, per quanto riguarda l'esperienza statunitense, i casi delle subculture Gamergater[81] e Far-Right[82], e, in Italia, i teorici delle scie chimiche.

Nel primo caso, una specifica comunità di *gamer*, ideologizzata e politicizzata sulla base dell'assunto che l'identità *geek* abbia per anni sofferto di una reputazione molto bassa a causa di numerose forme di oppressione sociale (a partire dalla semplice identificazione dei *gamer* con epiteti offensivi), sfrutta le piattaforme *web* (in particolare 4Chan) per organizzare attacchi ripetuti (che includono forme di dossieraggio, *revenge porn*, intimidazione e *social shaming*) contro femministe e altri nemici della loro ideologia. L'utilizzo di "brigate" organizzate e di *network* agili per intercettare nuovi adepti e diffondere i propri messaggi sono caratteristiche strutturali che sono state poi fatte proprie dal movimento *far-right*, che condivide con i *gamergater* numerose caratteristiche ideologiche e soprattutto alcuni esponenti di spicco o *influencer*.

Il movimento *far-right* (caratterizzato da un'ideologia estremista, con tratti di suprematismo razziale, misoginia ma anche anti-capitalismo) infatti utilizza le diverse piattaforme per convertire e reclutare nuovi attivisti, spingendoli esplicitamente verso forme di partecipazione *online*, attraverso la diffusione di messaggi e *meme*, e azioni da *troll*, aventi l'obiettivo di raggiungere anche i media *mainstream*, ponendo determinate tematiche nell'agenda mediale e

[77] Cfr. Jenkins H., Ford S., Green J. (2013), *Spreadable media. Creating Value and Meaning in a Networked Culture*, New York: New York University Press, trad. it. *Spreadable media. I media tra condivisione, circolazione, partecipazione*, Milano: Apogeo, 2013

[78] McCombs M.E., Shaw D. (1972), "The Agenda Setting Function of Mass Media", in *The Public Opinion Quarterly*, 36 (2), pp. 176-187. In particolare, al fine di verificare l'ipotesi di agenda-setting, le strategie e i disegni delle ricerche che rientrano in questo filone di studi prevedono tre fasi: a) l'analisi dell'agenda dei media (tramite analisi del contenuto), b) l'analisi dell'agenda del pubblico(ovvero l'insieme degli argomenti e dei temi che l'opinione pubblica ritiene più importanti in un determinate momento), attraverso *survey* rivolte ai cittadini; c) il confronto, in termini di tematiche di maggiore interesse, tra l'agenda dei media e l'agenda del pubblico..

[79] Metzger M. J., Flanagin, A. J. (2013), "Credibility and trust of information in online environments: The use of cognitive heuristics", in *Journal of Pragmatics*, 59, pp. 210-220.

[80] Cfr. Del Vicario M., Quattrociocchi W., Scala A., Zollo F. (2018), *Polarization and Fake News, op. cit.* Un recente studio ha inoltre dimostrato che la preferenza (di stampo ideologico) per determinate politiche pubbliche influenza *a priori* persino la credibilità scientifica dei fatti dai quali derivano quelle risposte politiche (Nisbet E.C., Cooper K.E., Garrett R.K. (2015). "The Partisan Brain. How Dissonant Science Messages Lead Conservatives and Liberals to (Dis)Trust Science". *Annals of the American Academy of Political and Social Science*, 658 (1), pp. 36-66).

[81] Cfr. Marwick A., Lewis R. (2017), *Media Manipulation and Disinformation Online, op. cit.*, pp. 7-9.

[82] Cfr. Lewis R., Marwick A., *Taking the Red Pill, op. cit.*

quindi del pubblico. In particolare, per ottenere questo tipo di risultati, i loro messaggi, che richiamano il suprematismo dell'uomo bianco o il cospirazionismo, sono spesso caratterizzati da un forte tensione emotiva basata sull'ostilità di "altri" (ebrei, immigrati musulmani, *élite* globali, femministe, ecc.), la cui presenza giustifica anche l'utilizzo di toni aggressivi e di vere e proprie azioni di *online harassment*. In particolare, questa subcultura si caratterizza per un forte discredito attribuito alle fonti di informazione tradizionali e per il presunto tentativo di disvelare la mutevole natura della verità ai suoi seguaci, secondo un processo chiamato *redpilling*, inteso come rivelazione e risveglio verso le realtà del mondo circostante[83].

L'utilizzo di tecniche e strategie coordinate di aggressione *online* contro esponenti e *opinion leader* avversi alle proprie posizioni è diventato di pubblico dominio in Italia, a seguito del processo che ha visto il *leader* di un gruppo cospirazionista delle scie chimiche condannato a otto mesi di carcere per diffamazione a mezzo *web* nei confronti della giornalista Silvia Bencivelli, medico e giornalista scientifico. Quest'ultima, vittima per cinque anni di messaggi di minacce e insulti quasi tutti a sfondo sessuale, ovvero di violente denigrazioni e di video a carattere minatorio, era rea, secondo l'imputato condannato e il suo gruppo, di aver smentito con un dettagliato articolo sul quotidiano nazionale La Stampa[84] la teoria cospirazionista delle scie chimiche, secondo la quale le scie bianche di condensa che a volte si formano nel cielo al passaggio degli aerei sarebbero diffuse da qualcuno (il Nuovo ordine mondiale, la Kasta, o comunque persone o organizzazioni dei "poteri forti") per avvelenare l'umanità. Anche questo gruppo, che porta avanti strategie di disinformazione su base ideologica (avversione ai "poteri forti", anti-capitalismo, complottismo, ecc.), si basa su campagne integrate di comunicazione che mirano a "svelare" e ribaltare acquisizioni e verità scientifiche con toni emotivi e rivelatori, come dimostrano i numerosi contenuti testuali e video ancora circolanti in Rete in cui vengono smentiti persino i contenuti della sentenza di condanna di uno dei loro *leader*.

I casi qui mostrati dimostrano come, nonostante le differenze tra i diversi gruppi e le relative ideologie sottese, le strategie di disinformazione ideologica evidenziano alcuni tratti comuni, con riferimento ad esempio ai toni emotivi delle campagne comunicative, che sfociano spesso in attacchi personali ed episodi di molestie *online*.

CASO 7 - Replica, rettifica e altri rimedi tradizionali alla prova della disinformazione online: il caso dei "Birthers"

Nel febbraio del 2007, Barack Obama annunciò la sua candidatura alle elezioni presidenziali del 2008, iniziando la corsa per le primarie contro una favorita Hillary Clinton. Già nel dicembre del 2007, un editoriale apparso sulla rubrica 'Smears 2.0' del Los Angeles Time evidenziava come il web 'fosse benzina sul fuoco per le odiose insinuazioni sulla presunta fede musulmana di Obama, ravvivando un odio antico" nei confronti di un possibile "Manchurian candidate".

In quell'editoriale, si evidenziava come i rumors contro Obama fossero iniziati nel 2006 con una campagna avviata da una virale e-mail nella quale Obama veniva definito 'The Enemy Within', il nemico all'interno, titolo di un film del 1994 che narrava di un piano militare per rimuovere il Presidente degli Stati Uniti. Nel gennaio di quell'anno, i responsabili della campagna di Obama furono costretti a denunciare Fox News per la ripetuta falsa notizia che Obama avrebbe speso quattro anni della sua vita a frequentare una scuola islamica indonesiana. Sebbene la CNN avesse dimostrato che la scuola frequentata da Obama non avesse nulla a che fare con gli incubatori pakistani per jihadisti, nel 2007 il Washington Post pubblicò una storia su Obama dal titolo 'Muslim ties'.

Il 9 giugno del 2008, Jim Geraghty sulla National Review Online, testata online di orientamento politico dichiaratamente conservatore, avanzò la richiesta ad Obama di pubblicare il suo certificato di nascita per smentire i rumors che lo indicavano come kenyota. Tre giorni dopo, i sostenitori della campagna di Obama pubblicarono il certificato di nascita sul sito web 'Fight The Smears', combatti quelli che infangano.

A fronte dell'applicazione di questi tradizionali metodi di contrasto alla diffusione di notizie false, l'effetto delle repliche o rettifiche fu di amplificare la polarizzazione del dibattito intorno a teorie cospirazioniste alimentate da un movimento di opinione che negli USA ha preso il nome di "Birthers". La pubblicazione del certificato, infatti, non solo non sopì le voci sul 'vero' luogo di nascita di Obama, ma ne attivò altre circa la veridicità del documento. In rete si diffusero, intorno alla notizia, narrazioni improntate a una tipica linea discorsiva del cospirazionismo: "se pubblicano un documento è per nascondere la scomoda verità, quindi, il documento deve essere falso. In fondo, ci vuole ben poco a contraffare o a sostituire un documento, no?".

[83] Lewis R., Marwick A., *Taking the Red Pill, op. cit.* Fra l'altro nel loro studio sul caso Far-Right le due ricercatrici tendono a sottolineare la differenza tra strategie di disinformazione mirate che hanno come obiettivo la diffusione dell'ideologia, l'attacco a determinati gruppi e il tentativo di raggiungere anche i media tradizionali e casi di *misinformation*, basati sulla circolazione di prodotti mediali contenenti ricostruzioni storiche inaccurate.

[84] Bencivelli S., *"Le scie chimiche" la leggenda di una bufala*, La Stampa, 16 settembre 2013.

Un anno dopo, nel 2009, con Obama già insediato alla Casa Bianca, un sondaggio commissionato dal blog The Daily Kos mostrava come il 28% degli elettori repubblicani riteneva che Barack Obama non fosse nato negli Stati Uniti. Ma ce ne era anche per l'ex Presidente George Bush Jr. Infatti, in quello stesso sondaggio, il 35% degli elettori democratici riteneva che George W. Bush fosse stato informato in anticipo dell'attacco alle torri gemelle: "both parties have their fanatics" sentenziò David Paul Kuhn, sul sito Real Clear Politics. I fanatici ci son sempre stati e sono dappertutto. Il che sembra suggerire che non dovremmo preoccuparcene più di tanto, perché, per fortuna, le elezioni sono decise da una maggioranza più saggia, meno tifosa e meno credulona.

Da allora, la falsa notizia su cui si fondano le accuse dei birthers riemerge ad ogni campagna elettorale statunitense, alimentando la polarizzazione delle opinioni espresse dall'elettorato su questioni sempre meno rilevanti rispetto alle proposte politiche dei soggetti che competono nelle singole consultazioni elettorali.

La vicenda, pertanto, dimostra che la propagazione di notizie false online produce effetti reputazionali duraturi difficilmente arginabili con i tradizionali metodi di contrasto, poiché le stesse smentite si prestano ad essere strumentalizzate nell'ambito di strategie mirate di disinformazione.

6.2. Le strategie di disinformazione *online* di natura politica

Le strategie di disinformazione a **carattere più strettamente politico** tendono invece a collocarsi a metà strada tra ***disinformation*, propaganda e *information operations***[85].

Tali strategie di disinformazione sono caratterizzate, per quanto riguarda la **creazione e la produzione** del messaggio, da contenuti che fanno esplicito riferimento ad opposte visioni e fazioni politiche, sia mettendo in luce temi e personaggi della propria parte, sia soprattutto denigrando personaggi e affrontando in maniera negativa temi portati avanti dalla fazione opposta.

Le relative **campagne** di comunicazione sono portate avanti da stati e singoli partiti politici di varia matrice (destra, populisti, *single-issue parties*, ecc.)[86], ma anche da fonti di informazione partigiana[87] e possono essere dirette ai propri concittadini e, in taluni casi, anche a cittadini esteri, attraverso le piattaforme *online*. Da un'analisi comparata tra le diverse campagne curate da stati e partiti o movimenti politici, tutte di lungo periodo e ben strutturate, emerge una comune **forma organizzativa definita** *cyber troop*, espressione ampia che fa riferimento all'esistenza di *team* governativi, militari[88], politici o partitici che hanno l'obiettivo di manipolare l'opinione pubblica sui *social media*, anche attraverso l'esplicito uso di "narrazioni dinamiche" per combattere opposte operazioni di propaganda[89].

Per quanto riguarda la **distribuzione dei messaggi**, gli attori che conducono campagne di disinformazione politica attraverso le piattaforme *online* utilizzano un insieme ampio e composito di prodotti e servizi capaci di diffondere messaggi e soprattutto determinati sentimenti e stati emotivi attraverso internet.

In particolare, tali organizzazioni utilizzano una varietà di **strumenti e tecniche** per la manipolazione dell'opinione pubblica attraverso i *social media*. La loro strategia comunicativa globale comprende: la creazione di siti e piattaforme governative o comunque ufficiali per la pubblicazione di contenuti; l'utilizzo di *account*, sia reali, sia *fake*, sia automatizzati, per interagire con gli utenti dei *social media*; la creazione di contenuti validi quali immagini, video, *post*, articoli, storie, *meme* da diffondere sulle piattaforme *online*. Similmente a quanto accade per i gruppi che diffondono campagne di disinformazione ideologica, queste organizzazioni – vere e proprie *fake tank* – sono molto attive nel commentare e interagire sui *social media*, anche attraverso contenuti a forte carica emotiva, perseguono strategie di *invidual targeting*, non solo mediante

[85] Nel senso attribuito alle due espressioni da Jack C., *Lexicon of Lies, op. cit.*,. Cfr. anche *infra*, par. 1.
[86] Cfr. Marwick A., Lewis R. (2017), *Media Manipulation and Disinformation Online, op. cit.*, pp. 4-22
[87] Silverman C. et al., *Hyperpartisan Facebook Pages, op. cit.*
[88] Sui i *team* militari si veda il caso della 77ma brigata dell'Esercito inglese creata dal Ministero della Difesa nel 2015 (cfr. Cholewa E. (2018), "Trovare l'arma giusta per un nuovo modello di conflitto: la 77a brigata dell'esercito inglese", in Melchior C., Romoli A. (a cura di), *La strategia della persuasione: Comunicazione e media nell'era della post-verità*, Milano: Franco Angeli).
[89] Howard P., Bradshaw S. (2017), *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*, Oxford: Computational Propaganda Project. Peraltro, gli autori del Rapporto distinguono in maniera netta le *cybertroops* da *hackers* ed esperti di *cybersecurity* eventualmente al servizio dei Governi.

l'uso di messaggi pubblicitari targetizzati (che sono solo una piccola parte della loro strategia), ma anche con campagne di *online harassment* verso utenti-*leader* portatori di una visione opposta[90], condotte magari attraverso *fake account*[91]o *bot*.

Se dal punto di vista della loro composizione, questi *team* possono coinvolgere tanto attori governativi e soggetti politici, quanto cittadini, volontari e organizzazioni o aziende esterne, anche dal punto di vista organizzativo le *cyber troop* si differenziano per struttura organizzativa, *budget* e ricorso ad attività di formazione, di ricerca e in generale di *capacity building*[92]. In ogni caso, si tratta di gruppi formati da un insieme di attori che anche attraverso la diffusione di *hate speech*, l'uso di *troll* o campagne automatizzate di comunicazione politica[93], hanno l'obiettivo di influenzare l'agenda mediale, politica e istituzionale[94].

In particolare, dal punto di vista delle **risorse tecnologiche**, la disinformazione politica segue la logica strutturale, beneficia dei prodotti e perfeziona le strategie del più ampio settore della pubblicità *online*[95], in particolare con riferimento a:

- tecniche di raccolta dei dati comportamentali degli utenti nel *web* (*web tracking*, *location tracking*, *cross-device tracking*) per poter indirizzare specifici contenuti (organici o a pagamento, testuali o multimediali), attraverso le diverse piattaforme;
- diversi tipi e formati di pubblicità *online*, anche automatizzate, che si rivelano profondamente utili per individuare specifici *target* (ad es. durante una campagna elettorale), poiché sfruttano il ciclo virtuoso per il quale più spazi pubblicitari vengono acquistati, più efficienti saranno le successive compravendite effettuate attraverso le *ad platform*, che avranno avuto modo di "imparare" e progressivamente affinare i *target* su cui indirizzare gli investimenti;
- attività di ottimizzazione della propria presenza tra i risultati delle *query* operate dagli utenti sui motori di ricerca (*search engine optimization* - SEO), che, nel caso specifico dei contenuti politici, può consentire agli agenti delle strategie di disinformazione di indirizzare il pubblico non solo e non tanto su *fake news* ma anche su notizie e informazioni tendenziose e maliziose che costituiscono la zona grigia della *malinformation*, o su notizie e informazioni provenienti dai propri canali;
- *software* di *social media management*, che non solo creano maggiori efficienze nella condivisione o nella promozione dei contenuti verso pubblici targetizzati, ma si caratterizzano anche per l'integrazione delle proprie piattaforme con algoritmi di *machine learning* in grado di produrre raccomandazioni sull'*audience*, sul contenuto e sulle tempistiche della campagna pubblicitaria, indirizzando in maniera automatica, nel caso della comunicazione politica, messaggi diversi a utenti diversi e aiutando così l'inserzionista a individuare maggiori sinergie tra contenuti organici e a pagamento; a rispondere in tempi giusti rispetto all'evoluzione di fatti di cronaca oggetto di attenzione politica, anche attraverso l'ascolto del *sentiment* verso queste vicende espresso dagli utenti in internet; infine, a proporre campagne negative e di *harassment* verso determinati utenti;

[90] In ogni caso, spesso questi team governativi possiedono sistemi altamente organizzati per identificare e targetizzare su base individuale il proprio pubblico di riferimento.
[91] *Idem*, pp. 8-13.
[92] *Idem*, pp. 14-21. Ad esempio, è stata dimostrata la forte capacità organizzativa della *troll factory* coordinata dall'Internet Research Agency russa nel triennio 2015-2017 (cfr. *infra*, caso 7).
[93] Come hanno anche sottolineato in interventi pubblici alcuni manager della Società, piattaforme *online* come Facebook sono diventate uno strumento per manipolare il discorso e l'impegno su temi civici e ingannare il pubblico (Weedon J., Nuland W., Stamos A., *Information Operations on Facebook*, *op. cit.*), ma anche, in maniera più ampia, un'ambiente che condiziona il corretto svolgimento del dibattito pubblico in maniera democratica, a causa dell'acuirsi di problematiche relative alla diffusione di *false news*, alla formazione di *echo chambers*, alla presenza di interferenze straniere su temi locali, e infine al manifestarsi di forme di *harassment* a sfondo politico verso singoli utenti (Chakrabarti S. (2018), *Hard Questions: What Effect Does Social Media Have on Democracy?*).
[94] Secondo un processo di *agenda building* così come definito da Cobb R. W., Elder C. D. (1971), The politics of agenda-building: An alternative perspective for modern democratic theory, in *Journal of Politics*, 33, 892-915; sul ruolo dei social media nell'influenzare l'agenda politica cfr. Parmelee J. H. (2014), The agenda-building function of political tweets, in *New Media & Society*, 16, 434-450.
[95] Ghosh D., Scott B., #DigitalDeceit, op. cit. .

– tutte le innovazioni in materia di intelligenza artificiale utili a prevedere comportamenti degli utenti nel tempo, a migliorare l'applicazione di tutti gli strumenti tecnologici della pubblicità *online*, e più in generale a creare capacità di avanzamento in ogni fase del processo[96].

Dal punto di vista, infine, della **valorizzazione** delle campagne di disinformazione politica, l'intento doloso si esplica nel tentativo di fare propaganda a favore di una determinata forza o visione politica ovvero di influenzare temi ed esiti di campagne elettorali di diverso tipo (politiche, locali, referendarie, *etc.*).

CASO 8 - Disinformazione politica in UK e USA: il caso delle interferenze russe durante le elezioni presidenziali USA 2016

Per comprendere al meglio strategie e campagne di disinformazione di natura più espressamente politico-elettorale, possiamo citare i casi relativi alle campagne di comunicazione portate avanti da partiti inglesi nel corso della campagna elettorale per le elezioni politiche 2017, e il caso delle interferenze russe nella campagna elettorale presidenziale USA 2016.

Nel caso britannico, alcune analisi esplorative hanno evidenziato che, tra i messaggi pubblicitari diffusi attraverso le piattaforme *online* (in particolare Facebook) da tutti i maggiori partiti in campo durante le recenti elezioni politiche, un'ampia percentuale era focalizzata su fatti in apparenza falsi o chiaramente manipolati per confondere i lettori (ad es. la mancata condanna dell'Irish Republican Army da parte del candidato laburista), e in generale su contenuti di tono emotivo atti a discreditare apertamente l'avversario. Inoltre, molti di questi contenuti erano non solo declinati su temi specifici legati ai programmi dei singoli partiti (ad es. la Brexit per i liberal-democratici), ma avevano anche un inquadramento ben preciso della narrativa intorno al tema, che attaccava apertamente le posizioni degli avversari politici, e si basavano su una targetizzazione ben precisa degli utenti (nel caso dei Lib-Dem coloro che avevano votato Remain al referendum 2016); infatti, le scelte dei contenuti da promuovere attraverso *post* sponsorizzati sono state effettuate sulla base di attente analisi del comportamento degli utenti *online*, che spesso ha portato alla scelta di concentrare i propri sforzi pubblicitari su collegi storicamente appannaggio di altri partiti (in particolare nel caso dei Laburisti), sempre attraverso *post* e contenuti che rimandavano ad attacchi negativi nei confronti dei principali avversari[97].

Se nei casi appena descritti non tutti gli strumenti e le tecnologie proprie della pubblicità digitale venivano sfruttate congiuntamente, diverso è stato il caso relativo alla strategia di disinformazione russa nella campagna elettorale presidenziale USA del novembre 2016: si tratta infatti di un esempio perfetto di integrazione di tattiche e strumenti diversi aventi lo scopo comune di diffondere idee politiche e notizie false al fine di condizionare la formazione dell'opinione pubblica su temi di campagna e di conseguenza l'esito del voto[98]. In particolare, a seguito delle prime ricostruzioni di organizzazioni governative che si occupano di sicurezza interna negli Stati Uniti, già nei primi mesi del 2017 sono emerse informazioni relative alla produzione e diffusione organizzata di contenuti di disinformazione su temi elettorali coordinate a livello governativo dalla Russia[99]. Inoltre, nello stesso periodo, alcuni studi hanno evidenziato le numerose notizie false su questi temi circolate grazie ai *social media*[100], provenienti da organizzazioni controllate dal Governo russo e da *account* ad esse collegate[101].

Nel settembre 2017, Facebook ha quindi ammesso che centinaia di *account fake* finanziati dall'Internet Research Agency (IRA), una società legata al Governo russo, avevano acquistato spazi pubblicitari indirizzati a specifiche categorie di

[96] Ghosh D., Scott B. (2018), *#DigitalDeceit, op. cit.*

[97] Cfr. Tambini D., Anstead N., Magalhaes J.C. (2017), *How the Liberal Democrats are using Facebook ads to court 'remainers'*; Idem (2017), *Labour's advertising campaign on Facebook (or "Don't Mention the War")*; Idem (2017), *Is the Conservative Party deliberately distributing fake news in attack ads on Facebook?*.

[98] Alcuni esperti hanno evidenziato come l'attività di propaganda russa attraverso siti di informazione specializzata sia iniziatata già nel 2014: in particolare secondo Clint Watts, esperto di sicurezza udito dalla Commissione speciale sull'intelligence del Senato USA, sin dal 2014 Internet e i social media hanno offerto alla Russia un accesso economico, efficiente ed efficace ai pubblici stranieri con una negazione plausibile della propria influenza diretta (Watts C., *Statement Prepared for the U.S. Senate Select Committee on Intelligence hearing: "Disinformation: A Primer In Russian Active Measures And Influence Campaigns"*, March 30, 2017). In un altro contributo, l'autore ricostruisce quattro diverse fasi dell'attività di propaganda russa tra il 2014 e il 2016: sviluppo delle capacità, espansione dell'audience, influenza dei cittadini, utilizzo di *leak* e narrative divisive (cfr. Watts C., *So What Did We Learn? Looking Back on Four Years of Russia's Cyber-Enabled "Active Measures"*, January 18, 2018). Sull'utilizzo del termine propaganda in questo contesto si veda *infra*, par. 1.

[99] USA Senate – Committee on Armed Services, *Hearing to receive testimony on Foreign Cyber Threats to the United States*, January 5, 2017; Office of the Director of National Intelligence, *Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution*, January 6, 2017.

[100] Howard P.N., Gorwa R., *Facebook could tell us how Russia interfered in our elections. Why won't it?*, May 20, 2017

[101] Weisburd A., Watts C., Berger J.M., *Trolling for Trump – How Russia is trying to destroy our democracy*, November 6, 2016

utenti in vista delle elezioni presidenziali[102]. Successivamente a questa rivelazione l'attenzione di politici e delle stesse piattaforme si è indirizzata sul tema della trasparenza delle pubblicità elettorali, come dimostrato da una proposta di legge del Senato USA[103] e dalle nuove politiche in materia di Facebook[104] e Twitter[105].

A seguito di queste numerose evidenze, il Ministero della Giustizia USA ha avviato una commissione speciale di inchiesta guidata a partire dal maggio 2017 dall'ex direttore dell'FBI Robert Mueller[106], che ha portato alla messa in accusa per cospirazione di società (ad es. IRA e Defendant) e di cittadini russi, con particolare riferimento all'utilizzo dei *social media* a fini elettorali (anche attraverso *troll* e profili falsi) e alla diffusione di campagne pubblicitarie targetizzate e *dark ads*[107], e successivamente ad una seconda messa in accusa nei confronti di dodici componenti dell'agenzia di intelligenza militare GRU, per presunte attività di *hacking* nei confronti di account di posta elettronica personali (anche della candidata democratica Hillary Clinton) e istituzionali[108]. Parallelamente all'azione della Commissione Mueller, l'Amministrazione Trump ha imposto una serie di sanzioni economiche verso oligarchi, pubblici funzionari, banche e aziende russe[109].

Come segnalato in un recente studio (che prende come riferimento l'ambiente Twitter nel triennio 2015-2017), i diversi *account* facenti capo all'IRA utilizzavano linguaggi, temi e schemi di comportamento diversi, tanto da poter essere distinti in almeno cinque gruppi principali (*right troll*, *left troll*, *news feed*, *hashtag gamer* e *fearmonger*), che agivano secondo modalità e tempistiche differenti, nel tentativo di incidere significativamente sull'agenda mediatica e politica (secondo un processo di *agenda building*), a dimostrazione quindi di una forte e stabile organizzazione alle spalle delle cosiddette "interferenze russe"[110]. Altri studi sul tema, alla base anche di alcune indagini e di numerosi documenti istituzionali, hanno invece testimoniato come nel tentativo di influenzare la campagna elettorale statunitense si incrocino molte delle tecniche e degli strumenti prima illustrati: pubblicità elettorali e notizie false indirizzate a pubblici precisi su base geografica (i cosiddetti *swing states*, tra cui il Michigan, dove nella settimana precedente le elezioni quasi la metà delle notizie su temi elettorali circolate su Twitter erano false o faziose[111]),"misure attive" di disinformazione basate su messaggi politici che attaccano *leader* democratici e minano le istituzioni, riguardanti temi socialmente divisivi, ovvero basati su sentimenti di paura[112]. In particolare, tali "misure attive" sono state perpetrate su tutte le diverse piattaforme di condivisione e scambio di informazioni, perseguendo, grazie alla caratteristiche specifiche di tutti gli strumenti utilizzati (si veda la tabella successiva[113]) cinque obiettivi complementari: (i) minare la fiducia dei cittadini nella forma di governo democratica; (ii) fomentare ed esacerbare fratture politiche divisive; (iii) erodere la fiducia tra cittadini e i politici eletti

[102] In particolare Facebook ha dichiarato che nel periodo intercorrente fra giugno 2015 e maggio 2017 un totale di circa 3000 inserzioni pubblicitarie, per una spesa complessiva superiore ai 100.000 dollari, è stato acquistato da circa 470 *fake accounts*. Circa la metà della spesa è stata realizzata da *account* statunitensi ma di lingua russa, o comunque legati a società, organizzazioni e istituzioni russe (Cfr. Stamos A. (2018), *An Update On Information Operations On Facebook*). Anche Twitter ha segnalato la presenza di inserzioni pagate da *account* russi collegati all'IRA (cfr. Testimony of Jack Dorsey, United States Senate Select Committee on Intelligence, September 5, 2018).

[103] S. 1989, *Honest Ads Act*, October 19, 2017.

[104] Facebook, *Making Ads and Pages more transparent*, April 6, 2018 .

[105] Twitter, *Increasing Transparency for Political Campaigning Ads on Twitter*, May 24, 2018.

[106] Negli ultimi mesi l'attività della Commissione speciale è stata alla ribalta sulla stampa internazionale soprattutto per le varie ripercussioni che ha avuto sulla politica interna, con particolare riferimento alla posizione del Presidente in carica Donald Trump e di parte del suo staff.

[107] Indictment, *United States v. Internet Research Agency LLC et al.*, docket entry 1, Feb. 16, 2018, case no. 18-cr-00032-DLF, U.S. District Court for the District of Columbia.

[108] Indictment, *United States v. Viktor Borisovich Netykshoo et al.*, Jul 13., 2018, Case 1:18-cr-00215-ABJ, U.S. District Court for the District of Columbia. A seguito di questo nuovo capo di accusa, sono stati segnalati e bloccati siti *fake* di numerosi candidati, esponenti politici e *think thanks* legati al Partito Repubblicano (Cfr. Smith B. (2018), *We are taking new steps against broadening threats to democracy*).

[109] Cfr. US Department of Treasury, *Treasury Designates Russian Oligarchs, Officials, and Entities in Response to Worldwide Malign Activity*, April 6, 2018.

[110] Linvill D.L., Warren P.L. (2018), *Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building*, working paper. Il dataset alla base dello studio è stato integralmente pubblicata sul sito di informazione statistica FiveThirtyEight ed è divenuto la base di partenza di numerosi reportage giornalistici che hanno raggiunto anche le prime pagine di alcuni quotidiani italiani il 3 agosto 2018 (nel dataset una categoria di account è costituita dai *non-english*, tra cui spiccano alcuni profili/*troll* di lingua italiana).

[111] Howard P. N., Bolsover G., Kollanyi B., Bradshaw S., Neudert L. M. (2017), *Junk news and bots during the US election: What were Michigan voters sharing over Twitter*, Oxford: Computational Propaganda Research Project.

[112] Tali "misure attive" di disinformazione sono gestite attraverso organizzazioni e fonti informative (RT, Sputnik News), siti cospirazionisti (InfoWars, ZeroHedge), aggregatori di incerta attribuzione, siti specializzati in trasferimento e condivisioni di dati e informazioni riservati (WikiLeaks), bot e un esercito di *social media operatives*, "un mix di account controllati dai russi, utili idioti e spettatori innocenti" (cfr- Weisburd A., Watts C., Berger J.M., *Trolling for Trump*, *op. cit.*)

[113] Tabella adattata e tradotta da Watts C., *Extremist Content and Russian Disinformation Online: Working with Tech to Find Solutions*, Testimony before the Senate Judiciary Committee, Subcommittee on Crime and Terrorism, October 31, 2017.

e tra i cittadini e le istituzioni democratiche; (iv) rendere popolari le *policies* tipiche dell'agenda russa nella popolazione straniera; (v) creare sfiducia o confusione sulle fonti informative rendendo sfocate le differenze tra fatti e finzione[114].

Interferenze russe e "misure attive"		
Obiettivo	**Piattaforme**	**Scopi e Benefici**
Posizionamento	**Primarie: 4chan, Reddit**	* inserzioni false nelle discussioni sui socialmedia
	Secondarie: 8chan, Youtube, Facebook	* insinuare cospirazioni tra il target di riferimento
		* diffondere *kompromat* (materiali compromettenti) su avversari mirati/selezionati, informazioni sia false che vere
		* nascondere le attribuzioni del Cremlino, fornendo una negabilità plausibile
Propagazione/diffusione	**Twitter**	* diffondere narrazioni/racconti attraverso gli accounts aperti del Cremlino e troll sotto copertura
		* amplificare la selezione del target delle storie e la preferenza per le narrazioniche supportano gli scopi del Cremlino (la propaganda computazionale fa apparire le falsità molto più credibili attraverso la ripetizione e il volume)
		* diffondere le narrazioni nei media principali in tutto il mondo
		* attaccare gli oppositori politici, esperti di politica estera e le personalità avversarie dei media
Saturazione	**Primarie/di base: Facebook**	* amplificare le divisioni sociali e politiche, erodere la fede nella democrazia attraverso le discussioni e lla pubblicità
	Secondarie: Google, LinkedIn, Instagram, Pinterest	* tirar fuori argomenti da altri programmi nelle discussioni tra amici e in famiglia
		* reclutare il target per la creazione e distribuzione di una propaganda organica o provocazioni/stimoli fisiche (proteste, manifestazioni o perfino violenza)
Hosting	**Youtube**	* post di aperta propaganda che offuscano la mano del Cremlino (RT)
		* condivisione di contenuti video al pubblico di riferimento tramite produttori e reporter piuttosto che attraverso i canali televisivi

In particolare, fonti informative quali RT e Sputnik hanno concentrato i loro sforzi sui *social media*, utilizzando tutte le piattaforme di *social networking* e condivisione di contenuti audiovisivi più note, monitorando l'opinione pubblica attraverso i commenti ai propri contenuti, e cercando, con diversi *account* ufficiali e non, di raggiungere sempre maggiore utenza[115]. Nonostante le numerose evidenze a supporto della tesi delle interferenze russe nella campagna elettorale presidenziale del 2016, la propaganda russa sta continuando ad agire anche con riferimento alle cd. elezioni di *mid-term* del 6 novembre 2018, come rilevato anche dagli esperti di sicurezza delle principali piattaforme *online*: in particolare alcuni account e pagine su Facebook, seguite da molti utenti statunitensi e autori di numerose inserzioni e post sponsorizzati sulla piattaforma, hanno anche creato eventi sulla piattaforma, che sono stati rimossi prima di dar vita a manifestazioni e incontri dal vivo[116].

L'insieme di tecniche di raccolta dei dati comportamentali degli utenti in Rete, pubblicità *online* targettizzate, creazione di siti e piattaforme comunque ufficiali per la pubblicazione di contenuti, utilizzo di *account*, sia reali, sia *fake*, sia automatizzati, per interagire con gli utenti dei *social media*, contenuti a forte carica emotiva, campagne di odio verso i *leader* avversi, rende il caso delle interferenze russe durante la campagna elettorale presidenziale USA un efficace esempio di strategia di disinformazione politica, replicabile da altri Stati. Nell'agosto 2018, infatti, Facebook ha rimosso numerose pagine, gruppi, account, tutti coinvolti in campagne coordinate di "azioni non autentiche", provenienti non solo dalla Russia, ma anche dall'Iran, precisando peraltro che si tratta di due campagne distinte, accomunate però dall'utilizzo di tattiche simili nella creazione di network di account che ingannavano gli utenti (soprattutto del Medio

[114] Watts C., *Statement Prepared for the U.S. Senate Select Committee on Intelligence hearing: "Disinformation: A Primer In Russian Active Measures And Influence Campaigns", op. cit.*

[115] Come evidenziato da Office of the Director of National Intelligence, *Background to "Assessing Russian Activities and Intentions in Recent US Elections", op. cit.*

[116] Facebook, *Removing Bad Actors on Facebook*, July 31, 2018.

Oriente, Americana Latina, Stati Uniti e Regno Unito) circa la loro provenienza e i loro scopi[117]. Nei giorni immediatamente successivi, anche Twitter ha rimosso numerosi account[118], mentre Google ha segnalato la rimozione di 39 canali su YouTube riferibili al governo iraniano, la cui presenza online tramite campagne e azioni coordinate di propaganda è – secondo esperti di sicurezza – molto più focalizzata sulla diffusione dei suoi interessi di politica estera[119], rispetto a quella russa, che mira più direttamente a interferire nel dibattito pubblico di altri Stati.
Attualmente, il tema delle interferenze straniere in campagna elettorale è oggetto di numerosi studi e ambito di discussione per l'opinione pubblica anche nel contesto europeo, con riferimento a molte delle elezioni nazionali o referendum occorsi negli ultimi due anni, ed è costantemente monitorato dalle istituzioni europee in vista delle elezioni dei membri del Parlamento Europeo previste nel giugno 2019[120].

I casi di disinformazione ideologica e politica, nel loro insieme di azioni sinergiche e coordinate, dimostrano l'importanza di una **costante azione di monitoraggio** di tutte le attività che possono minare il pluralismo dell'informazione e, di conseguenza, la formazione dell'opinione pubblica nel nuovo ecosistema digitale. In particolare, il forte legame consequenziale tra campagne di disinformazione, fenomeni di *hate speech*, radicalizzazione degli utenti, frammentazione sociale e rischi di deviazione del dibattito pubblico dal binario del corretto esercizio dei diritti e doveri democratici, richiede una decisa attenzione delle istituzioni, e in particolare delle autorità nazionali di regolazione, sia sul rispetto dell'esercizio della libertà di informazione sulle piattaforme digitali, sia sull'osservanza dei relativi limiti, ovvero su casistiche peculiari, che sfruttando i meccanismi di funzionamento delle piattaforme, possono influenzare il diritto all'informazione del cittadino.

Per quanto riguarda più nel dettaglio la lotta alla disinformazione a scopo politico-ideologico, **soluzioni condivise** tra operatori presenti sul mercato dell'informazione e piattaforme *online*, coordinate e promosse dal regolatore, relativi alla **rilevazione e monitoraggio** dei fenomeni di disinformazione *online*, a meccanismi di **trasparenza** e **valorizzazione** delle organizzazioni editoriali, a **tecniche e strumenti di *fact-checking***, anche tecnologicamente avanzati, **all'educazione dei consumatori/utenti**, possono certamente rappresentare una strada più utile rispetto a singole azioni legislative prive degli strumenti conoscitivi adeguati.

7. Osservazioni conclusive

L'analisi svolta in questo rapporto rileva che le distorsioni **dell'informazione *online*** si presentano come un vasto **fenomeno multiforme**, per caratteristiche, soggetti coinvolti, motivazioni sottostanti, tecniche di comunicazione impiegate per comporre i contenuti *fake*, strumenti e tecnologie utilizzate, risorse investite.

Nell'ambito di tale complessità, a partire dai risultati della letteratura e della ricerca scientifica sull'argomento e dalla casistica concreta, sono stati individuati alcuni **caratteri distintivi**, utili a chiarire meglio i contorni del fenomeno, per evidenziarne le criticità sotto il profilo del pluralismo e della correttezza dell'informazione e, di conseguenza, gli strumenti più adatti a gestirle.

Un primo elemento qualificante riguarda la caratterizzazione delle distorsioni dell'informazione *online* che agiscono tanto dal lato dell'offerta quanto dal lato della domanda. Al riguardo, la disinformazione *online* si colloca all'interno di un più vasto ambito di fenomeni, classificabili nelle macro-categorie della mala-informazione, mis-informazione e disinformazione. In particolare, la **disinformazione *online*** costituisce un

[117] Facebook, *Taking Down More Coordinated Inauthentic Behavior*, August 21, 2018. In particolare, Facebook ha comunicato la rimozione di 652 pagine, gruppi e account (su Facebook e Instagram) di origine iraniana (con contenuti in lingua inglese e araba).
[118] Cfr. Testimony of Jack Dorsey, United States Committee on Energy and Commerce, September 5, 2018.
[119] Come segnalato da Volz D., McMillan R., *Tech Giants Target Iran*, The Wall Street Journal, August 24, 2018.
[120] Si veda in particolare il progetto di monitoraggio, ricerca, raccolta di casi e *fact-checking* Disinformation Review coordinato dalla task force East StratCom, organizzazione facente parte dell'amministrazione dell'Unione europea, creata nel 2015 e incentrata sulla comunicazione strategica delle politiche e delle attività dell'UE nei Paesi del vicinato orientale (Armenia, Azerbaijan, Bielorussia, Georgia, Moldavia e Ucraina) e nella stessa Russia.

tipo specifico di distorsione dell'informazione, che presenta un livello di problematicità più elevato nella prospettiva di tutela del pluralismo informativo (**paragrafo 2**).

Un secondo elemento distintivo emerge allorché si analizzi il processo che presiede alla produzione, distribuzione e diffusione dei contenuti *fake*. A questo proposito, si osserva che i disturbi dell'informazione *online* possono dar luogo a una vera e propria **filiera dei contenuti *fake***, più o meno strutturata a seconda del soggetto che promuove l'iniziativa e delle sue motivazioni, nonché dei *target* verso i quali è diretta la strategia informativa. Dalla ricostruzione della filiera emerge un preciso *modus operandi*, articolato in fasi distinte, di creazione, produzione, distribuzione e valorizzazione dei contenuti *fake* (**paragrafo 3**).

Un terzo elemento, che scaturisce dall'analisi della disinformazione *online*, svolta in particolare sulla filiera e sulla casistica, riguarda la possibilità di individuare delle **strategie di disinformazione *online***, in particolare nei casi in cui gli ideatori dei contenuti *fake* non sono individui singoli ma organizzazioni stabili, o anche temporanee accomunate da interessi specifici, mosse da precisi obiettivi di natura economica e/o politico-ideologica, con dotazioni finanziarie, tecnologiche e organizzative e *target* di destinatari ben individuati. Inoltre, tali strategie si manifestano in genere con una serie di azioni di disinformazione che danno luogo non a singoli sporadici episodi, ma a una serie di pubblicazioni e/o ri-pubblicazioni di contenuti *fake*; si tratta di campagne di disinformazione, dunque, che possono avere durata variabile nel tempo (**paragrafo 4**).

Un quarto elemento, messo in luce dall'esame delle strategie di disinformazione *online*, è la **varietà di modalità con cui esse si manifestano**; in maniera semplificata, infatti, si possono distinguere strategie di disinformazione di breve periodo e strategie di più lungo respiro, strategie con finalità economiche e strategie con finalità politico-ideologiche, nonché azioni con finalità ibride. Dall'incrocio di tali dimensioni, emerge una possibile classificazione, utile sia per l'analisi del fenomeno, sia per la predisposizione di specifiche soluzioni in via regolamentare (**paragrafo 4**).

Tale classificazione distingue, in particolare, le strategie commerciali da un lato, e quelle di natura politico-ideologica dall'altro; l'analisi di ciascuna di esse, inoltre, mette in risalto una **serie di criticità per il pluralismo e la correttezza dell'informazione** prodotte dai fenomeni di disinformazione *online* inclusi quelli particolarmente polarizzanti ed estremizzanti che precipitano in *hate speech* (**paragrafi 5 e 6**).

Considerata la varietà di profili problematici emersa dall'analisi effettuata nel documento, le possibili **misure di contrasto** alla disinformazione *online*, di tipo autoregolamentare o, in prospettiva, regolamentare, devono essere anch'esse **molteplici e complementari** tra loro. Inoltre, per essere efficaci ed equilibrate, esse potrebbero essere **concordate e coordinate** tra i diversi attori dell'informazione *online*, richiedendo, dunque, il coinvolgimento di tutti gli *stakeholder*[121].

Il Tavolo Tecnico, quindi, si conferma la sede più idonea a sviluppare una gamma di soluzioni che dovranno essere approfondite e discusse specificamente all'interno dei diversi Gruppi di lavoro, secondo un'agenda di temi rilevanti che, a partire dalle evidenze emerse in questo documento, l'Autorità si propone di sottoporre all'attenzione dei membri del Tavolo. Al tempo stesso, il Tavolo dovrà rivelarci, accanto alle opportunità, i limiti dell'autoregolazione e quali misure, ancorché minime, di regolazione possono esser necessarie, non tanto sotto il profilo della regolazione del contenuto – tema che si porrebbe in aperto contrasto con la libertà di espressione – quanto piuttosto sotto il profilo delle tecniche di selezione algoritmiche, di *framing* e organizzazione editoriale, di trasparenza e consapevolezza per gli utenti, di accesso diretto – da parte di un soggetto terzo e indipendente - ai dati profilati e rivelati dai comportamenti di utenti e inserzionisti, nonché delle dinamiche e delle tecniche di disseminazione, contagio, diffusione. Diventa peraltro fondamentale valutare gli impatti delle misure di autoregolazione, anche attraverso misurazioni terze e indipendenti degli impatti.

[121] In tale direzione, peraltro, si è mossa anche la Commissione europea, sia mediante l'adozione di un Codice di condotta a luglio 2018, sia mediante le altre iniziative che hanno fatto seguito alla Comunicazione 2018/236 - *Tackling online disinformation: a European Approach.*

Al riguardo, sin d'ora è possibile individuare i tre tipi principali di attività da portare avanti nel prosieguo dei lavori:

- Sviluppo di forme di **cooperazione** continuativa di tipo tecnico-scientifico, anche con università e centri di ricerca, per migliorare la conoscenza dei fenomeni di disinformazione *online* e misurare gli impatti delle misure adottate o in corso di adozione;

- Iniziative di **divulgazione e scambio di esperienze** all'interno dei membri del Tavolo, come già sperimentato con gli eventi organizzati dall'Autorità, sui temi della disinformazione commerciale[122] e sulle soluzioni di mercato al problema della disinformazione[123], tenendo conto anche del versante del mercato che si riferisce alle strategie degli inserzionisti; .

- **Predisposizione di soluzioni operative**, quali ad esempio la preparazione di linee guida, la progettazione e implementazione di campagne di informazione indirizzate agli utenti, l'individuazione di modalità di opt-in e opt-out degli utenti rispetto ai default informativi, l'elaborazione di programmi e buone prassi di media *literacy*, lo sviluppo di nuove soluzioni di *fact-checking*, la promozione di nuove attività ad alto carattere innovativo condivise tra tutti gli operatori del mercato dell'informazione *online*, *etc*.

In particolare, relativamente all'ultimo punto – sul piano operativo – alla luce dell'esperienza maturata nell'ambito del Tavolo Tecnico con il lavoro svolto nel corso della campagna elettorale precedente[124], l'Autorità si farà promotrice di linee guida e codici di condotta condivisi con i membri del Tavolo, sviluppate anche in vista delle prossime elezioni, ovvero di nuove attività condivise tra gli operatori, sotto l'egida dell'Autorità, tra le quali in particolare:

- **Il costante monitoraggio dei fenomeni di disinformazione e di hate speech *online*** (Gruppo A - Metodologie di classificazione e rilevazione dei fenomeni di disinformazione *online*).

- **L'individuazione di forme di trasparenza del sistema della pubblicità *online*** (Gruppo B – Definizione di sistemi di monitoraggio dei flussi economici pubblicitari, da fonti nazionali ed estere, volti al finanziamento dei contenuti *fake*).

- **La creazione di una piattaforma di coordinamento delle attività autonome di *fact-checking*** e definizione di standard giornalistici in materia di trasparenza, etica e qualità della struttura editoriale (Gruppo C – *Fact-checking*: organizzazione, tecniche, strumenti ed effetti).

- **Le iniziative per la *media literacy e il contrasto ai fenomeni di hate speech online*** (Gruppo D - Media e *digital literacy*: promuovere la cultura mediatica e digitale fornendo ai cittadini strumenti per un uso consapevole e critico dei (social e non) media), anche attraverso linee guida, spot e altre iniziative di comunicazione

- **L'introduzione di nuovi strumenti di *trasparenza ed empowerment* del consumatore** (Gruppo E – Progettazione e realizzazione di campagne informative su disinformazione rivolte ai consumatori), anche attraverso linee guida, spot e altre iniziative di comunicazione.

[122] Gli effetti della disinformazione commerciale sulle scelte dei consumatori, tenutosi l'11 giugno 2018.
[123] Disinformazione e soluzioni di mercato: i casi Qwant e Wikipedia", tenutosi lunedì 25 giugno 2018.
[124] Cfr. Linee guida per la parità di accesso alle piattaforme online durante la campagna elettorale per le elezioni politiche 2018