
Ragebait “parola dell’anno”, e Reddit come caso di studio: un laboratorio di disinformazione che comunicatori e relatori pubblici non possono ignorare

di Giorgia Grandoni

Nota per il lettore: in questo articolo verrà alternato l'uso del maschile e del femminile sovraesteso, nel tentativo di costruire senso attraverso un linguaggio quanto più inclusivo possibile.

Negli ultimi cinque anni, nel mestiere di relatori pubblici, comunicatori e reputation & crisis manager, parole come misinformazione e disinformazione sono diventate di fatto parte del lessico di base, quasi alla stregua di “stakeholder” o “accountability”. Non è un tema nuovo, tuttavia, ad essere cambiato radicalmente è la scala del fenomeno e la velocità con cui circola: la pandemia di Covid-19 ha rappresentato uno spartiacque anche per chi si occupa di queste tematiche. Il dibattito globale sull’“infodemia” – la sovrapposizione di emergenza sanitaria ed eccesso di informazioni fuorvianti o apertamente false – ha reso evidente come la gestione dei flussi informativi sia ormai parte integrante della gestione delle crisi e delle policy pubbliche¹, circostanza confermata anche dal progetto della rubrica **Textures** di FERPI, nata da un’intuizione di Toni Muzi Falconi; anche grazie a quel lavoro minuzioso, costruito attraverso monitoraggi sistematici, letture di paper accademici e analisi di casi internazionali, diventò evidente come la disinformazione fosse diventata (assieme ai temi di *diversity & inclusion* e rendicontazione ESG) il “polo caldo” della riflessione professionale, soprattutto nei contesti anglosassoni, Stati Uniti e Regno Unito, che tradizionalmente rappresentano laboratori avanzati per il nostro settore².

Proprio in quel biennio si registrò un’esplosione di pubblicazioni su fake news, misinformazione e polarizzazione. Meta-analisi e rassegne bibliometriche mostrano infatti un incremento repentino di articoli, special issue e conferenze dedicate a questi temi a partire dal 2019–2020, spesso con un focus specifico sul ruolo delle piattaforme digitali nel catalizzare o contenere la circolazione di contenuti falsi o distorti³.

¹ WHO, 2020; Lovari, 2020

² Zhou & Zafarani, 2018; Ecker et al., 2022

³ Zhou & Zafarani, 2018; Lovari, 2020

Oggi, a distanza di cinque anni, siamo entrati in una **nuova fase**, caratterizzata da attori, meccanismi e infrastrutture di manipolazione profondamente diversi da quelli che abbiamo imparato a conoscere nel decennio precedente. Se la prima ondata di attenzione era stata innescata dai Social media “classici” e dall'infodemia pandemica, la seconda è alimentata dall'arrivo dell'**intelligenza artificiale generativa**.

Modelli capaci di produrre testi, immagini, audio e video del tutto verosimili introducono un livello ulteriore – e più insidioso – di ambiguità: la sfida di oggi non è più distinguere tra fonti attendibili e contenuti imprecisi o manipolati, ma tra **realtà e simulazioni del reale** costruite con una fedeltà tale da rendere l'esercizio di verifica estremamente più complesso anche per i professionisti dell'informazione⁴.

Gli episodi più recenti lo dimostrano con chiarezza. Immagini generate dall'IA – dall'esplosione mai avvenuta nei pressi del Pentagono fino a scene di cronaca o di politica costruite con estetica foto-realistica – sono riuscite a entrare nel circuito dell'informazione autorevole, venendo riprese da media e commentatori prima che se ne svelasse l'origine artificiale⁵. In questo contesto, anche per chi lavora nella comunicazione strategica diventa sempre più difficile distinguere, in tempo utile, ciò che è verificabile da ciò che è frutto di una **allucinazione algoritmica**, di una manipolazione intenzionale o semplicemente di una verifica approssimativa operata da fonti intermedie.

Per la nostra professione questo implica almeno due consapevolezze. Da un lato, riconoscere che – a distanza di anni – la disinformazione rappresenta la **cornice strutturale** entro cui si gioca una parte rilevante del lavoro su reputazione, fiducia e legittimità; dall'altro lato, ammettere che le categorie attraverso cui abbiamo imparato a leggere l'ecosistema mediale nell'ultimo decennio – *earned, owned, paid media* – sono sempre meno sufficienti a descrivere gli spazi, fluidi e stratificati, in cui le narrazioni oggi nascono, si deformano e si radicalizzano.

Per noi relatrici pubbliche, in quanto tessitrici di relazioni, la **credibilità dei contenuti** e l'**attendibilità dei processi comunicativi** non sono attributi astratti, ma le condizioni materiali attraverso cui si costruisce fiducia e si consolidano legami – interni ed esterni – che sostengono qualsiasi strategia complessa. In un ecosistema che evolve così rapidamente, comprendere dove e come la disinformazione si genera, si amplifica e si istituzionalizza non è più una competenza accessoria: è parte integrante della nostra “cassetta degli attrezzi”.

In relazione a ciò, in un ecosistema saturo di fake, semi-fake, contenuti manipolati e materiali generati dall'IA, la domanda sorge spontanea: come si comportano piattaforme apparentemente “di nicchia” ma straordinariamente resilienti come Reddit, agorà che – forse in virtù dell'assoluta assenza di moderazione - pare

⁴ Ecker et al., 2022

⁵ Clayton, 2023

funzionare ancora oggi come uno dei più potenti moltiplicatori di disinformazione online?

Reddit rappresenta infatti per certi versi un'eccezione strutturale nel panorama delle piattaforme digitali. A differenza di altri Social che, negli ultimi dieci anni, hanno cambiato più volte interfaccia, algoritmi, logiche di feed, modelli di business e perfino finalità dichiarate, Reddit è rimasto sorprendentemente coerente con sé stesso, dal 2005 a oggi. L'esperienza utente – basata su thread testuali, votazione “dal basso”, comunità tematiche e un'architettura volutamente essenziale – è rimasta pressoché immutata. Eppure, come mostra *Il Post* (2025), proprio questa coerenza è uno dei motivi della sua attuale centralità: Reddit è riuscito a sopravvivere all'“onda lunga” dell'intelligenza artificiale generativa perché ciò che offre non è facilmente replicabile dai modelli linguistici, ovvero un'enorme quantità di discussioni, testimonianze individuali, micro-esperienze quotidiane e contesti dialogici che l'IA fatica a generare in modo autentico.

In altre parole: Reddit non è esploso *malgrado* l'intelligenza artificiale, ma *grazie* al fatto che l'intelligenza artificiale – per ora – non riesce a imitare pienamente ciò che produce il suo ecosistema. Anche per questo continua a essere utilizzato in modo massiccio dalle generazioni più giovani, specie nel mondo anglosassone ma non solo: è una piattaforma che non chiede di performare, ma di discutere; non richiede estetica, ma argomentazione; non valorizza l'autopromozione, ma la partecipazione collettiva.

Proprio questa combinazione – stabilità dell'architettura e persistenza dell'uso intergenerazionale – fa di Reddit un osservatorio privilegiato per capire **come circola la disinformazione e come si strutturano, nel tempo, dinamiche di polarizzazione, attacco reputazionale e manipolazione coordinata dei contenuti**.

Questo contributo non vuole quindi aggiungersi all'elenco – già consistente – degli articoli “generalisti” sulle fake news: l'intento è diverso e più circoscritto, ovvero comprendere perché in particolare Reddit, tra tutte le piattaforme Social, meriti oggi un'attenzione specifica da parte di chi si occupa di reputazione, comunicazione strategica e gestione delle crisi reputazionali, in quanto la sua natura ibrida, la sorprendente continuità strutturale nel tempo e la sua capacità di attrarre l'attenzione delle nuove generazioni — pur restando fedele a un modello quasi “pre-Social” — lo rendono un osservatorio privilegiato per studiare come le narrazioni si formano, si distorcono e si amplificano.

Nelle pagine che seguono proveremo quindi a mettere ordine tra elementi storici, empirici e concettuali per mostrare come Reddit si collochi dentro l'ecosistema informativo contemporaneo: non come semplice piattaforma, ma come uno spazio in cui la disinformazione può trovare terreno fertile e, allo stesso tempo, in cui emergono forme di resistenza comunitaria e di auto-regolazione che meritano di essere comprese con attenzione.

2. Reddit nell’evoluzione dell’internet: un “vecchio” formato, con nuova centralità

Per comprendere il ruolo di Reddit nell’ecosistema contemporaneo della disinformazione bisogna partire da un dato apparentemente semplice: Reddit non somiglia a nessuno dei Social network che dominano l’immaginario collettivo. Non ha il feed algoritmico di Facebook, né la grammatica visiva e performativa di Instagram o TikTok, né il ritmo sincopato e polarizzante di X. Reddit è, per struttura e filosofia, tutt’altra cosa: un ibrido tra forum, piattaforma di Social news e spazio di discussione pseudonima che — sorprendentemente — è cambiato pochissimo dal 2005 a oggi.

Se volessimo collocare Reddit su una mappa dell’evoluzione dell’internet, dovremmo immaginarlo come una sorta di “fossile vivente” dell’era dei “forum”, sopravvissuto all’esplosione dei Social network generalisti e ai cicli di ascesa e dal declino delle grandi piattaforme⁶. A differenza di Facebook, X o Instagram, non nasce per mettere in contatto persone che già si conoscono, ma per aggregare conversazioni attorno a interessi e temi specifici. Nato nel 2005, Reddit conserva ancora oggi la sua struttura originaria: thread testuali organizzati in board tematiche (i “*subreddit*”), un sistema di voto che premia o penalizza i contenuti, un elevato grado di anonimato o pseudonimia. Non è il feed “degli amici”: è un ecosistema di comunità.

Dal punto di vista quantitativo, parliamo di un’infrastruttura di scala industriale: centinaia di milioni di utenti attivi al mese (Perez, 2019), miliardi di visualizzazioni e oltre 2 milioni di *subreddit*, di cui più di 100.000 stabilmente attivi⁷. Ogni anno vengono pubblicati complessivamente miliardi di post e commenti, che alimentano un flusso continuo di conversazioni e interazioni.

Per comprendere perché questo formato sia così peculiare – e così rilevante quando parliamo di disinformazione – vale la pena aprire brevemente la “scatola nera” del funzionamento della piattaforma. Ogni utente, dietro un nickname, può creare o frequentare *subreddit*, ossia spazi tematici dedicati praticamente a qualsiasi argomento: dalla finanza all’attivismo politico, dai videogiochi alle malattie rare, passando per recensioni di film o serie tv. All’interno di ogni *subreddit*, gli utenti possono pubblicare post (testo, link, immagini, video) e commentare quelli altrui; ogni contenuto può essere votato positivamente (upvote) o negativamente (downvote).

La visibilità di un post dipende quindi da una combinazione di fattori: numero e bilancio dei voti, velocità con cui questi voti arrivano, livello di attività nei commenti, regole specifiche del singolo *subreddit*. I contenuti che “performano” meglio in termini di *upvote* e interazioni scalano le classifiche interne e, in alcuni casi, entrano

⁶ Il Post, 2025

⁷ FrontPageMetrics, 2020

nel flusso di r/popular, la vetrina trasversale dei contenuti più rilevanti della piattaforma. Per l’utente registrato, la homepage è una miscela di ciò che è “popolare” a livello globale e di ciò che è rilevante rispetto ai *subreddit* cui è iscritto; per chi non è registrato, l’accesso passa comunque da r/popular e da algoritmi di selezione basati su geolocalizzazione e *trending topic*.

Per chi non ha un’elevata *platform literacy* (ovvero la capacità di leggere criticamente le logiche di visibilità, ranking e moderazione) Reddit può apparire come un ambiente neutrale; in realtà, questa architettura incentiva e rende più redditizie - in termini di visibilità - **proprio le dinamiche che attori malevoli possono sfruttare**.

La letteratura più recente conferma che questo intreccio tra meccanismi di promozione della visibilità, anonimato e manipolazione, può avere effetti diretti sul clima della piattaforma: livelli più elevati di inciviltà e linguaggio ostile tendono a concentrarsi proprio nei contesti dove circolano **contenuti inaffidabili, producendo come conseguenza un aumento misurabile della polarizzazione e della conflittualità**⁸.

Questo meccanismo – apparentemente semplice – ha due implicazioni decisive. La prima è che la visibilità è la posta in gioco principale: ogni attore che desideri spingere una narrazione, un link, un frame interpretativo ha un incentivo forte a intervenire su voti e commenti per alterare artificialmente la percezione di “rilevanza” del contenuto⁹. La seconda è che il legame tra identità offline e profilo online è molto labile: il fatto che l’unità base sia il *nickname* e non la persona facilita il ricorso a *sockpuppet*, ovvero account coordinati, *bot* e forme di manipolazione – a volte malevola - organizzata. In letteratura queste strategie sono descritte con un lessico ormai consolidato: *vote brigading* (coordinare *upvote* e *downvote* per spingere o affossare un contenuto), *astroturfing* (simulare consenso “spontaneo” dove in realtà c’è una regia organizzata), *shill posting* (post retribuiti o interessati mascherati da opinioni neutre), *karma-whoring* (ricerca sistematica di visibilità e punteggio a prescindere dalla qualità informativa intrinseca) e uso estensivo di *bot* per pubblicare, commentare e votare in modo automatico. È un repertorio di pratiche che sfrutta fino in fondo la centralità della visibilità nel sistema di funzionamento di Reddit.

A livello di governance, Reddit si regge su un’architettura multilivello: da un lato gli amministratori della piattaforma, che definiscono le regole generali (*content policy*, regole contro l’odio e le molestie, *policy* sulle interferenze coordinate, etc); dall’altro migliaia di moderatori volontari, che applicano queste regole all’interno dei singoli *subreddit* e spesso introducono norme aggiuntive, locali, su ciò che è accettabile o meno nelle rispettive comunità¹⁰. In buona sostanza, è una forma di auto-governo

⁸ Hanley & Durumeric, 2025

⁹ Carman et al., 2018

¹⁰ Chandrasekharan et al., 2018

distribuito, nella quale le regole formali e le norme implicite si intrecciano in modo non sempre trasparente.

Non a caso, come osservava già Bergstrom più di dieci anni fa, Reddit “*lotta con le aspettative della sua stessa community*” e oscilla costantemente tra anonimato e responsabilità, tra mito della libertà assoluta e necessità di porre limiti all'abuso nella piattaforma¹¹.

Per lungo tempo la piattaforma è stata raccontata – anche dai suoi stessi leader – come una sorta di baluardo della libertà di espressione, al punto da rivendicare pubblicamente, nei primi anni, l'idea di “*non rimuovere nulla*”. In realtà, dietro questa narrativa, Reddit ha sempre moderato, bannato ed editato contenuti ritenuti inappropriati “*sulla base delle nostre convinzioni*”, come ammette lo stesso CEO Steve Huffman¹² in un intervento pubblico del 2015¹³.

La tensione tra mito e pratica diventa evidente nella stagione delle grandi controversie tra il 2014 e il 2016: il divieto della pornografia non consensuale, il ban di alcuni *subreddit* accusati di molestie coordinate *off-line*, il caso Pao¹⁴ e il blackout di numerosi *subreddit* in segno di protesta per il licenziamento della coordinatrice degli AMA, Victoria Taylor¹⁵ ¹⁶, sono tutti episodi che mostrano quanto il potere dei moderatori volontari – in grado, con un blackout mirato, di bloccare intere sezioni del sito – si intrecci con le scelte strategiche del management, in un equilibrio sempre instabile tra *community*, inserzionisti e opinione pubblica.

Nello stesso periodo, Reddit inizia anche a trasformarsi più esplicitamente in infrastruttura commerciale: passa dal codice *open-source* a un modello chiuso, consente l'ingresso di investitori istituzionali e, in seguito, di partecipazioni contestate da parte di attori come la cinese Tencent, che alimentano sospetti di interferenze politiche e di condizionamento nelle scelte di moderazione¹⁷. È in questo contesto che la piattaforma viene progressivamente associata non solo a comunità **eccentriche o tossiche**, ma a episodi concreti di disinformazione politica e interferenza nelle elezioni, in particolare quelle statunitensi del 2016¹⁸.

Gli studi sulle campagne di influenza legate all'elezione di Donald Trump mostrano come vari *subreddit* politici abbiano visto aumentare in modo esponenziale la circolazione di link verso fonti controverse e contenuti polarizzanti, con un

¹¹ Bergstrom, 2011

¹² imprenditore, dirigente d'azienda e attivista statunitense, conosciuto per essere il cofondatore e CEO di Reddit

¹³ Huffman, 2015; Hill, 2012

¹⁴ Il “caso Pao” fa riferimento alla controversia legale avviata da Ellen Pao contro la società di venture capital Kleiner Perkins per discriminazione di genere (causa conclusa nel 2015 con esito a favore dell'azienda) e alle successive proteste sorte su Reddit durante il suo incarico come CEO ad interim della piattaforma.
<https://www.wired.it/economia/business/2015/07/13/ellen-pao-reddit-addio/>

¹⁵ Nel 2015 numerosi *subreddit* andarono in blackout — sospendendo le attività o rendendosi privati — per protestare contro il licenziamento improvviso di Victoria Taylor, coordinatrice delle sessioni “Ask Me Anything” (AMA). L'evento mise in luce tensioni fra la dirigenza di Reddit e le sue comunità, contribuendo a un ampio dibattito sulla gestione del sito.

¹⁶ Kienzle, 2016; Matias, 2016

¹⁷ Freedom House, 2019; Constine, 2019; Liao, 2019

¹⁸ Nithyanand, Schaffner & Gill, 2017; Hautala, 2017; Breland, 2017

incremento al tempo stesso dei livelli di inciviltà comunicativa e di coordinamento tra account che si presentavano come “semplici cittadini” ma agivano invece in modo orchestrato¹⁹. In parallelo, analisi giornalistiche e rapporti indipendenti evidenziano il ruolo di Reddit come “banco di prova” dove testare e socializzare contenuti fuorvianti prima di rilanciarli su piattaforme più esposte come Facebook e Twitter, spesso al riparo dai riflettori mediatici e dalle inchieste regolatorie che si concentravano su altri player²⁰.

La risposta di Reddit a questa stagione è stata ambivalente. Da un lato, nel 2019 e 2020 la piattaforma pubblica i primi rapporti di sicurezza sulla “manipolazione dei contenuti”, dichiarando di aver rimosso decine di milioni di post e commenti, sanzionato milioni di account e bannato quasi 7.000 *subreddit* tossici²¹ in seguito all'introduzione di nuove policy e al rafforzamento degli strumenti di moderazione automatica e umana²². I primi *Security Report* restituiscono con chiarezza questa scala d'intervento massivo: in soli sei mesi del 2019 Reddit registra oltre 10 milioni di segnalazioni di manipolazione dei contenuti, più di 30 milioni di rimozioni amministrative e milioni di sanzioni agli account; nel 2020 le rimozioni arrivano a decine di milioni di contenuti e a quasi 7.000 *subreddit* bannati. Non è più solo un problema di “mele marce”: è un ecosistema strutturalmente esposto a tentativi sistematici di manipolazione²³.

Dall'altro lato, proprio questi interventi alimentano il sospetto che Reddit stia cedendo alle pressioni degli inserzionisti e degli attori politici, “ripulendo” selettivamente il sito per renderlo più presentabile, ma senza affrontare in modo strutturale le dinamiche profonde di polarizzazione e manipolazione che attraversano la piattaforma²⁴.

In questo scenario, la definizione di Reddit come “vecchio formato con nuova centralità” trova un ulteriore livello di lettura: non si tratta solo di una piattaforma che ha conservato l'estetica testuale e modulare dell'internet dei forum, ma di un ambiente che ha mantenuto – quasi intatti – meccanismi di visibilità, incentivi e culture d'uso pensati in un'altra fase storica della rete. Secondo un'analisi recente de Il Post sarebbe proprio questa coerenza architetture, unita alla capacità di rinnovare il proprio modello di business attraverso la pubblicità, l'abbonamento

¹⁹ Nithyanand, Schaffner & Gill, 2017

²⁰ Breland, 2017; Potter, 2021

²¹ Nel dibattito contemporaneo, il termine “tossicità” non viene utilizzato soltanto nel suo significato colloquiale di comunicazione sgradevole o conflittuale. Nella letteratura di Human–Computer Interaction e di Computational Social Science, “tossicità” indica un insieme specifico e misurabile di comportamenti linguistici che ostacolano la partecipazione, innescano ostilità o interrompono la conversazione. È un concetto che comprende forme di inciviltà, aggressività verbale, attacchi personali e commenti ostili, ed è associato alla degradazione della qualità del dialogo pubblico. Nella maggior parte degli studi empirici recenti, la “tossicità” viene valutata tramite modelli automatici di classificazione linguistica — in particolare la metrica *SEVERE_TOXICITY* della Google Jigsaw Perspective API — che definisce come tossici i messaggi esplicitamente aggressivi, insultanti o intimidatori, tali da scoraggiare l'interazione dei partecipanti. In questo senso, la tossicità è una proprietà *sistemica* delle conversazioni online, non una semplice caratteristica individuale: funge da indicatore del deterioramento dei contesti comunicativi e del loro progressivo scivolamento verso l'ostilità e la polarizzazione.

²² Reddit, Security Reports 2019–2020; Copland & Davis, 2020

²³ Reddit Transparency Report 2019; Reddit Transparency Report 2020; Copland & Davis, 2020

²⁴ Marsh, 2019; Freedom House, 2019

premium e l'integrazione sempre più stretta con l'ecosistema dell'intelligenza artificiale (anche come “miniera” di dati per l'addestramento dei modelli), a permettere a Reddit di attrarre investimenti e attenzione, nonostante un *look and feel* che potrebbe apparire “datato” rispetto ai Social più visuali²⁵.

Questa combinazione di fattori – struttura da forum, centralità dei subreddit tematici, anonimato/pseudonimia, governance ibrida tra amministratori e moderatori, incentivi alla visibilità e storia di ambivalenza nella moderazione – rende Reddit un terreno particolarmente interessante per chi si occupa di disinformazione. Da un lato, è un ambiente che facilita la creazione di community iper-verticali, dove gruppi di appassionati o di attivisti possono costruire conversazioni sofisticate, mutualismo informativo e forme di auto-regolazione virtuosa; dall'altro, proprio le stesse *affordance* possono essere sfruttate da **attori malevoli che mirano a testare frame, amplificare contenuti divisivi, coordinare campagne di ragebait o attacco reputazionale a basso costo e con elevata plausibilità**.

Per chi fa il nostro mestiere, Reddit appare quindi come una sorta di laboratorio permanente: un osservatorio privilegiato per studiare, in forma quasi “pura”, le dinamiche di **costruzione collettiva del consenso e del dissenso**, e al tempo stesso un moltiplicatore potenziale di disinformazione capace di condizionare, in modo indiretto ma molto concreto, le narrazioni che poi arrivano sui media *mainstream* e nelle arene decisionali. È su questo doppio registro – Reddit come spazio di sperimentazione comunitaria e come infrastruttura di rischio reputazionale – che si innesteranno le sezioni successive di questo contributo.

3. Reddit come bacino e moltiplicatore di disinformazione: cosa dicono le ricerche

Una volta chiarita l'architettura della piattaforma e le sue peculiarità, diventa possibile osservare Reddit non più soltanto come un “contenitore neutrale di discussioni”, bensì come un ambiente in cui determinate configurazioni tecniche e culturali **rendono la disinformazione particolarmente fertile**. Le ricerche degli ultimi anni – sia qualitative sia quantitative – convergono infatti su un punto essenziale: **Reddit non si limita a ospitare flussi di contenuti inaffidabili, ma tende ad amplificarne gli effetti, a trasformarli in catalizzatori di inciviltà, polarizzazione e conflitto**²⁶. È come se alcune parti dell'ecosistema funzionassero da “camere di risonanza” in cui il rumore informativo si addensa, si radicalizza e, in alcuni casi, prende la forma di campagne coordinate capaci di influenzare conversazioni ben oltre i confini della piattaforma.

Il lavoro di Potter (2021), ricostruendo in profondità i momenti più controversi nella storia di Reddit, mostra come questa infrastruttura sia stata più volte utilizzata come

²⁵ Il Post, 2025

²⁶ Potter, 2021; Hanley & Durumeric, 2025

vero e proprio *staging ground*: un terreno di prova nel quale attori organizzati – talvolta politici, talvolta para-istituzionali, o perché no all’interno di dinamiche distorsive della concorrenza tra brand – hanno potuto testare frame narrativi, misurare reazioni, simulare consenso e perfezionare contenuti prima di esportarli su piattaforme con una visibilità maggiore. L’elezione presidenziale statunitense del 2016 rappresenta, sotto questo profilo, un caso emblematico: i dati riportati da Nithyanand, Schaffner e Gill (2017) mostrano come i subreddit politici abbiano subito un’accelerazione improvvisa nella circolazione di link provenienti da *controversial sources*, con esposizioni che nei *subreddit* repubblicani aumentano del 600% dopo l’inizio delle primarie e del 1.600% dopo la Convention repubblicana. Non si tratta solo di un aumento solo quantitativo: nello stesso periodo esplode anche il livello di inciviltà nei commenti, insieme alla crescita di account apparentemente “comuni” ma attivi in forme di coordinamento difficilmente spiegabili come comportamento organico. Il risultato è un ambiente che, nel giro di pochi mesi, passa dall’essere uno spazio di confronto spesso acceso ma relativamente spontaneo, a diventare uno dei principali laboratori di sperimentazione della comunicazione politica polarizzante e a tratti tossica.

Ciò che emerge dai documenti del *Digital Forensic Research Lab*²⁷ conferma questa lettura. Reddit risulta coinvolto in operazioni di disinformazione multiplatforma condotte attraverso account falsi, documenti manipolati e tattiche di occultamento riconducibili ad attori “persistenti, sofisticati e ben finanziati”²⁸. È significativo che queste operazioni utilizzino Reddit nella fase iniziale del ciclo di manipolazione: è il luogo in cui si prepara il materiale, si osservano le reazioni delle community e si calibra la narrativa da diffondere poi altrove. Se un contenuto riesce a ottenere trazione – attraverso *upvote* coordinati, commenti indignati, *thread* ad alta densità emotiva – allora può essere rilanciato su Facebook, X, YouTube e, in ultimo, persino nei media tradizionali.

Su queste dinamiche qualitative si innesta il contributo quantitativo di Hanley & Durumeric (2025), che consente di misurare in modo rigoroso ciò che avviene quando link a fonti inaffidabili vengono inserite nei *thread*. L’evidenza è inequivocabile: i *thread* che rimandano a domini classificati come *unreliable* presentano **livelli di inciviltà più elevati** rispetto ai *thread* che linkano a fonti affidabili, con un incremento relativo della tossicità pari al 25% a livello aggregato e superiore al 32% se si osserva il fenomeno all’interno dei singoli *subreddit*. È un effetto non solo robusto, ma profondamente rivelatore: **significa che la presenza di contenuti inaffidabili agisce da fattore scatenante, altera lo stile discorsivo di una community e spinge verso forme comunicative più ostili, più aggressive, più scoraggianti per chi – potenzialmente – vorrebbe partecipare con spirito collaborativo.**

²⁷ Atlantic Council, 2019

²⁸ DFRLab, 2019

Ma l'aspetto forse più rilevante riguarda la *distribuzione* della disinformazione. Non è infatti un fenomeno uniforme: le fonti inaffidabili rappresentano una minoranza dei contenuti complessivi, ma sono fortemente concentrate in un numero ristretto di *subreddit* insulari, politicamente orientati e caratterizzati da livelli già elevati di tossicità. Sono, in altre parole, delle vere e proprie “zone ad alta densità” in cui le dinamiche di polarizzazione affettiva diventano più intense. L'analisi tramite *Exponential Random Graph Models* mostra infatti che, in presenza di link inaffidabili, la probabilità che utenti con orientamenti politici opposti interagiscano in modo tossico aumenta in modo significativo (OR=1.38). La disinformazione, dunque, non si limita “a circolare”: produce attrito, amplifica la frattura tra gruppi, irrigidisce le percezioni reciproche e innesca un conflitto simbolico che tende ad auto-alimentarsi.

Un ulteriore elemento riguarda il modo in cui questi contenuti generano partecipazione. Hanley & Durumeric mostrano che, nelle community più tossiche, i post che rimandano a fonti inaffidabili attirano un volume maggiore di commenti nelle fasi iniziali della discussione. Si tratta, tuttavia, di un engagement “esplosivo ma superficiale”: **molte interazioni concentrate in poco tempo, spesso connotate da ostilità e scarsa profondità argomentativa**. Anche questo conferma quanto la letteratura osserva da anni: l'indignazione è un potente motore di attenzione, e la disinformazione sa sfruttare perfettamente questa dinamica, attivando reazioni rapide e polarizzate che, nel giro di poche ore, possono alterare in modo significativo la percezione di un tema o di un soggetto.

Se si mettono insieme questi elementi – l'uso di Reddit come banco di prova per operazioni coordinate; la concentrazione della disinformazione in comunità insulari e polarizzate; l'aumento sistematico di inciviltà e ostilità nei thread con fonti inaffidabili; l'intensificazione della polarizzazione affettiva; e infine la capacità della piattaforma di innescare forme di engagement rapido e tossico – emerge un quadro coerente: Reddit non è soltanto uno spazio in cui la disinformazione *circola*, ma un ambiente nel quale essa può *radicarsi, crescere e produrre effetti amplificati*, sia a livello discorsivo sia a livello relazionale.

Per noi professionisti della comunicazione strategica questo comporta una doppia responsabilità. Da un lato, significa comprendere le condizioni materiali e simboliche che rendono Reddit un habitat così favorevole alla distorsione informativa; dall'altro lato, implica riconoscere che alcune delle narrazioni che incontriamo nei media *mainstream* – e che a volte diventano problemi reputazionali veri e propri – hanno origine esattamente in questi spazi: in *thread* che nascono come discussioni di nicchia, si trasformano in catalizzatori di ostilità e, se intercettati da attori organizzati, possono diventare parte di una strategia di influenza a più livelli. Ed è proprio in questa intersezione – tra spontaneità e manipolazione, tra cultura della partecipazione e sfruttamento delle sue vulnerabilità – che si gioca una parte decisiva della gestione contemporanea della reputazione e delle crisi online.

4. Ragebait e danni reputazionali

Solo pochi giorni fa, l'Oxford University Press ha scelto **ragebait** come *Word of the Year 2025*, sancendo a livello simbolico qualcosa che chi lavora nella comunicazione conosce da tempo: **l'esistenza di contenuti progettati non per informare, ma per suscitare rabbia, indignazione, risentimento, con l'obiettivo di massimizzare click, commenti e condivisioni**. Lo stesso Oxford definisce il *ragebait* come un tipo di contenuto sensazionalistico costruito per provocare reazioni negative intense e aumentare l'engagement, più che per trasmettere informazioni accurate²⁹.

Su Reddit, inutile sottolinearlo, questo meccanismo trova un terreno d'elezione, particolarmente favorevole. Molti *thread* nascono già come *ragebait*: titoli iperbolici, frame moralmente binari (“scandalo”, “vergogna”, “gente da arrestare”) e narrazioni che sollecitano sentimenti di ingiustizia, tradimento o minaccia. Spesso il contenuto di partenza è parziale, decontestualizzato o apertamente falso; ciò che conta non è la qualità informativa, ma la capacità di innescare un'ondata di reazioni viscerali.

L'architettura della piattaforma – upvote/downvote, schede *hot* e *controversial*, visibilità legata alla rapidità dei commenti – **premia esattamente questo tipo di dinamica**, e i dati di Hanley e Durumeric sull'aumento sistematico di commenti tossici nei *thread* che linkano a domini inaffidabili ci ricordano che, laddove entra in gioco la disinformazione, la probabilità che un contenuto si trasformi in *ragebait* cresce in modo significativo³⁰.

In questa prospettiva, alcune campagne di disinformazione su Reddit assomigliano molto a operazioni di **black PR**. Per *black PR* intendiamo quell'insieme di pratiche comunicative che mirano a danneggiare deliberatamente la reputazione di un soggetto – persona, azienda, istituzione – attraverso la costruzione e la diffusione di narrazioni negative, spesso basate su insinuazioni, mezze verità o notizie decontestualizzate o nei casi limite completamente inventate. Il *ragebait* funziona come vettore ideale di queste operazioni: trasforma sospetti e illazioni in “pozzi gravitazionali” di attenzione, canalizza la rabbia latente di comunità già polarizzate e legittima – con il linguaggio della “denuncia dal basso” – cornici profondamente delegittimanti.

I casi aziendali degli ultimi anni mostrano con chiarezza questa sequenza. Prendiamo il caso **Mattress Firm** (2018), la più grande catena di negozi di materassi negli Stati Uniti. Tutto parte da un post su Reddit che si domanda come sia possibile che esistano così tanti punti vendita, spesso vicinissimi fra loro, suggerendo – senza

²⁹ Oxford Languages, 2025; Wikipedia, 2025

³⁰ Hanley & Durumeric, 2025

alcuna prova – che l'azienda possa essere **una copertura per attività di riciclaggio di denaro**. Il commento diventa virale, rimbalza tra diversi subreddit, viene discusso, memato, amplificato; la teoria del "money laundering" viene ripresa da altri social e attira l'attenzione dei media generalisti³¹.

Qui il ragebait è evidente: un framing complottista ("nessuno compra così tanti materassi, quindi deve esserci qualcosa di losco") che intercetta il sospetto diffuso verso le grandi catene retail, lo estremizza e lo rende irresistibile da commentare. Per l'azienda, il risultato è un problema reputazionale tutt'altro che teorico: il CEO è costretto a rilasciare dichiarazioni ufficiali per smentire le illazioni, spiegare la logica delle acquisizioni e della rete di negozi, rispondere ai giornalisti su una storia nata da un singolo thread. In termini di dinamica, lo schema è chiaro:

1. nasce un post altamente emotivo in un contesto di nicchia;
2. la discussione si polarizza e si autoalimenta;
3. il contenuto migra verso altre piattaforme e media;
4. il sospetto si sgancia dal contesto originario e diventa narrazione autonoma ("forse davvero c'è qualcosa di strano in quella catena"), che l'azienda deve faticosamente disinnescare.

Ancora più istruttivo, per chi si occupa di crisi, è il caso **Wayfair** (2020)³². In piena stagione QAnon, un utente del subreddit r/conspiracy nota alcuni mobili industriali dal prezzo elevato sul sito Wayfair e osserva che i prodotti hanno nomi propri che ricordano quelli di alcune ragazze scomparse. Da qui il salto: l'ipotesi – del tutto infondata – che Wayfair sia in realtà **un canale per il traffico di minori**. Nel giro di poche ore la teoria si diffonde su Reddit, poi trabocca su Twitter, TikTok, Facebook e nell'ecosistema QAnon, diventando hashtag, meme, video virali³³.

Anche in questo caso, gli elementi tipici del *ragebait* ci sono tutti: l'accusa è moralmente di massimo impatto (pedofilia, traffico di bambini), la narrativa è semplice e intuitiva ("costi folli + nomi di ragazze = qualcosa di inconfessabile"), le prove sono nel contempo debolissime ma emotivamente potenti (screenshot di schede prodotto, collage di immagini, analogie visive). Reddit funge da camera di risonanza iniziale in cui la rabbia e l'orrore vengono coltivati, legittimati, resi condivisibili; il frame "azienda = complice di un crimine indicibile" è così forte da sopravvivere, **anche dopo i fact-checking di BBC, USA Today, Reuters e Snopes, che smontano punto per punto ogni collegamento con reali casi di scomparsa**³⁴.

Per Wayfair, la gestione della crisi è complessa: l'azienda è costretta a ritirare temporaneamente i prodotti incriminati, a emettere comunicati dettagliati, a

³¹ Lee 2018; Levin 2018; Greene, 2018; Business Insider, 2018

³² Lokmanoglu et al. 2025; Basanta 2020

³³ Spring, 2020; BBC News, 2020; Brown, 2020

³⁴ Ibidem;

spiegare perché alcuni prezzi fossero effettivamente molto alti, e da cosa derivassero i nomi dei prodotti. **Ma, come spesso accade nei casi di black PR o di teorie cospirative a forte componente emotiva, la smentita corretta fatica a raggiungere la stessa profondità del frame iniziale, e il marchio resta per lungo tempo associato, nell’immaginario di una parte dell’opinione pubblica, a una vicenda tanto spettacolare quanto inesistente.**

In entrambi gli esempi, e in molti altri casi analoghi, alcuni rilevati anche in Italia, assistiamo allo stesso cortocircuito: un contenuto strutturato come *ragebait* – forse nato da un sospetto “organico”, forse alimentato da attori interessati – sfrutta le *affordance* di Reddit per trasformarsi in un vortice di rabbia condivisa. Nel passaggio da discussione di nicchia a narrazione di massa, la componente fattuale si assottiglia, mentre si sedimentano etichette e associazioni emotive che colpiscono direttamente la reputazione del *brand*. Per chi si occupa di comunicazione strategica e in particolare di crisis communication, questo significa riconoscere che Reddit non è solo un luogo da monitorare: è uno degli spazi in cui, oggi, si progettano, si testano e si amplificano le forme più sofisticate di *ragebait*, spesso indistinguibili – agli occhi del pubblico – dalle campagne di *black PR* propriamente dette.

5. Cosa significa tutto questo per chi fa relazioni pubbliche e comunicazione

Tutto ciò che abbiamo osservato finora non è un fenomeno marginale, né un’anomalia legata a una singola piattaforma: è parte del nuovo paesaggio entro cui si muove chi lavora nella comunicazione strategica. Per noi professioniste, Reddit rappresenta oggi un radar precoce: un luogo in cui le narrative ostili spesso emergono, si consolidano e si sperimentano prima di raggiungere TikTok, X, Instagram o, nei casi peggiori, i media tradizionali.

È spesso qui che compaiono i primi segnali deboli di una crisi reputazionale:

- frame disinformativi che iniziano a prendere forma;
- interpretazioni tossiche che sedimentano sotto la soglia dell’attenzione mainstream;
- “storie sul brand” che non hanno ancora una vita pubblica ma stanno già costruendo immaginario.

Monitorare Reddit significa intercettare ciò che ancora non esiste ufficialmente, ma sta già acquisendo trazione in comunità che anticipano — talvolta di settimane o mesi — ciò che diventerà conversazione pubblica.

Per farlo, servono competenze nuove, alcune tecniche, altre culturali. La prima è la *platform literacy*: non una generica familiarità con il digitale, ma la capacità di leggere i meccanismi interni della piattaforma, comprenderne la cultura,

interpretare moderazione e reputazione locale dei singoli *subreddit*. La seconda è la capacità di trattare la tossicità non come semplice rumore, bensì come segnale: un indicatore precoce della polarizzazione che può attrarre, e amplificare, narrazioni ostili.

Accanto alle competenze umane entra in campo l'intelligenza artificiale. Da un lato, AI e modelli linguistici sono sempre più integrati negli strumenti di moderazione automatica; dall'altro lato, diventano risorse fondamentali per noi professionisti, che possiamo usarli per:

- social listening avanzato;
- classificazione automatica di grandi volumi di contenuti;
- rilevazione anticipata di thread anomali;
- identificazione di pattern ricorrenti nelle conversazioni.

È una tensione inevitabile: la stessa tecnologia che aiuta a riconoscere la disinformazione può essere usata per produrla, automatizzarla, scalarla. Ma proprio per questo il nostro ruolo acquista centralità. La responsabilità non è quella di essere guardiani moralistici della rete, ma di proteggere la qualità delle relazioni — il vero capitale del nostro lavoro — in un ambiente dove la fiducia è fragile e il rumore può travolgere anche i messaggi più solidi.

Ed è qui che riprendono consistenza i tre fili che hanno orientato questo contributo: la disinformazione come cornice strutturale, non episodica, del nostro lavoro; Reddit come laboratorio e moltiplicatore di narrazioni, non solo come piattaforma Social; la responsabilità professionale come pratica quotidiana, non come mera dichiarazione di principio.

L'appello che ne scaturisce è un invito a immaginare una forma più evoluta e matura della nostra professione. Integrare nel lavoro quotidiano l'analisi delle piattaforme non mainstream, investire nella comprensione delle culture digitali, formarsi su come funzionano davvero le community: sono questi i passi che ci permetterebbero di tutelare lo spazio pubblico informativo e, insieme, la dignità del nostro mestiere.

Perché, in fondo, chi fa relazioni pubbliche lavora da sempre sul fragile terreno della fiducia. E in un ecosistema rumoroso, polarizzato e spesso manipolabile, contribuire a preservare la qualità delle conversazioni — e la verità delle relazioni — non è un compito accessorio: è la parte più alta e necessaria del nostro lavoro.

Postilla dell'autrice. Il processo di scrittura di questo contributo non è stato delegato all'intelligenza artificiale generativa, che pure è stata utilizzata come supporto nella redazione dell'articolo, in particolare per la ricerca delle fonti. Nell'intento di ribadire un principio che considero parte integrante della formazione contemporanea — soprattutto per le studentesse e gli studenti che si preparano a entrare in una professione in cui gli strumenti cambiano con rapidità — ci tengo a sottolineare che l'intelligenza artificiale, quando è compresa e utilizzata con consapevolezza, può diventare un acceleratore di analisi e un'estensione della capacità di sintesi, non in sostituzione della capacità di valutazione e giudizio umani, ma al fine di accompagnarli al meglio, senza mai abdicare alla propria autonomia critica e di pensiero.

Bibliografia

- Appadurai, A.; *Modernity at large: Cultural dimensions of globalization*; University of Minnesota Press; 1996.
- Basanta, T.; *What Companies Should Take Away from the Wayfair Conspiracy Theory and Messaging Gone Wrong*; FIU News; 2020.
- BBC News; *Wayfair: The false conspiracy about a furniture firm and child trafficking*; BBC News; 2020.
- Bergstrom, K.; *"Don't feed the troll": shutting down debate about community expectations on Reddit.com*; First Monday; 2011.
- Boyd, d.; *You think you want media literacy... do you?*; keynote e paper; Data & Society; 2018.
- Breland, A.; *Warner sees Reddit as potential target for Russian influence*; The Hill; 2017.
- Brown, M.; *Fact check: Home goods retailer Wayfair is not involved in child sex trafficking*; USA Today; 2020.
- Business Insider; *Mattress Firm CEO denies Reddit money-laundering conspiracy theory*; Business Insider; 2018.
- Carman, M., Koerber, M., Li, J., Choo, K. R., Ashman, H.; *Manipulating visibility of political and apolitical threads on Reddit via score boosting*; IEEE TrustCom/BigDataSE Conference Proceedings; 2018.
- Carrigan, M.; *Platform literacy: The new critical skill for a datafied society*; blog accademico / volume collettaneo da definire; 2019.
- Chandrasekharan, E. et al.; *The Internet's hidden rules: an empirical study of Reddit norm violations at micro, meso, and macro scales*; Proceedings of the ACM on Human-Computer Interaction, CSCW; 2018.
- Clayton, A.; *Fake AI-generated image of explosion near Pentagon spreads on social media*; The Guardian; 2023.
- Constine, J.; *Reddit confirms \$300M Series D led by China's Tencent at \$3B value*; TechCrunch; 2019.
- Copland, C.; Davis, M.; *Deplatforming extremism: Reddit's 2020 subreddit bans*; *Information, Communication & Society*; 2020.
- Copland, S., Davis, J.; *Reddit removes millions of pro-Trump posts. But advertisers, not values, rule the day*; The Conversation; 2020.
- Digital Forensic Research Lab (DFRLab); *Operation "Secondary Infektion": A Russian disinformation campaign*; Atlantic Council; 2019.

- Draper, N.; Turow, J.; The corporate cultivation of digital resignation; *New Media & Society*; 2019.
- Ecker, U.; Lewandowsky, S.; Cook, J. et al.; *The psychological drivers of misinformation belief and its resistance to correction*; *Nature Reviews Psychology*; 2022.
- Freedom House; *China Media Bulletin: Tencent complicity, surveillance upgrades, Reddit manipulation* (No. 134); Freedom House; 2019.
- Freedom House; *Freedom on the Net: The Crisis of Social Media*; Freedom House; 2019.
- FrontPageMetrics; *Reddit subreddit statistics and growth*; FrontPageMetrics.com; 2020.
- Greene, D.; *The wild Mattress Firm money-laundering theory, explained*; *The Washington Post*; 2018.
- Hanley, H. W. A.; Durumeric, Z.; *Sub-Standards and Mal-Practices: Misinformation's Role in Insular, Polarized, and Toxic Interactions on Reddit*; *Proceedings of the ACM on Human-Computer Interaction*, vol. 9, n. 7, articolo CSCW480; 2025.
- Hautala, L.; *Reddit was a misinformation hotspot in 2016 election, study says*; *CNET News*; 2017.
- Hill, K.; *Reddit co-founder Alexis Ohanian's rosy outlook on the future of politics*; *Forbes*; 2012.
- <https://theconversation.com/reddit-removes-millions-of-pro-trump-posts-but-advertisers-not-values-rule-the-day-141703>
- Huffman, S.; *Reddit's content policy and the evolution of self-governance*; *Reddit blog post*; 2015.
- *Il Post*; *Reddit è rimasto uguale mentre intorno Internet peggiorava*; *Il Post*; 2025.
- Kienzle, J.; *The technological factors of Reddit: communication and identity on relational networks*; University of Nebraska–Lincoln, Communication Studies Theses; 2016.
- Lee, M.; *A Reddit Thread Accusing Mattress Firm of Money Laundering Went Viral — Here's Why People Are So Convinced*; *Business Insider*; 2018.
- Levin, A.; *Mattress Firm Files for Bankruptcy, Could Close Up to 700 Stores*; *Fox Business*; 2018.
- Liao, S.; *Reddit gets a \$150 million investment from Tencent and users are posting memes to mock the deal*; *The Verge*; 2019.
- Liao, S.; *Tencent's investment in Reddit raises questions over content moderation*; *The Verge*; 2019.
- Lokmanoglu, A.; Walter, D.; Ophir, Y.; Fabregat, E.; *Some Assembly Required: Unpacking the Content and Spread of Wayfair Conspiracy Theory on Reddit and Twitter*; Boston University College of Communication; 2025.
- Lovari, A.; *Spreading (dis)trust: COVID-19 misinformation and government communication in Italy*; *Journal of Science Communication*; 2020.
- Marsh, P.; *Reddit users revolt after reports of Chinese tech giant Tencent's possible investment*; *ABC News (Australia)*; 2019.
- Matias, J. N.; *Going dark: social factors in collective action against platform operators in the Reddit blackout*; *CHI Conference Proceedings*; 2016.
- Newman, N.; *Journalism, media, and technology trends and predictions 2019*; Reuters Institute; 2019.
- Nithyanand, R., Schaffner, B., Gill, P.; *Online political discourse in the Trump era*; *arXiv / research report*; 2017.
- Nithyanand, R.; Schaffner, B.; Gill, P.; *Online political discussions and the spread of controversial sources on Reddit*; in *Proceedings of the 2017 Workshop on Politics, Elections and Data (ACM)*; 2017.
- Oxford Languages; *Oxford Word of the Year 2025: Rage bait*; Oxford Languages; 2025
- Perez, S.; *Reddit now has 430 million active users*; *TechCrunch*; 2019.
- Potter, M.; *Bad actors never sleep: content manipulation on Reddit*; *Continuum – Journal of Media & Cultural Studies*; 2021.
- Reddit Inc. (2019). *Reddit Transparency Report 2019*.
<https://www.redditinc.com/policies/transparency-report-2019>

- Reddit Inc. (2020). *Reddit Transparency Report 2020*.
<https://www.redditinc.com/policies/transparency-report-2020>
- Reddit Inc.; *Reddit Security Report – June 18, 2020*; r/announcements (Reddit); 2020.
- Reddit Inc.; *Reddit Security Report – October 30, 2019*; r/announcements (Reddit); 2019.
- Stewart, L.G.; Arif, A.; Starbird, K.; Examining trolls and polarization with a retweet network; in *Proceedings of ACM CHI*; 2018.
- Turner, B.S.; *Vulnerability and human rights*; Pennsylvania State University Press; 2006.
- u/KeyserSosa; *Reddit security report: April–September 2019*; post su r/redditsecurity; 2019.
- u/worstnerd; *Changes to Reddit's content policy and enforcement*; post su r/redditsecurity; 2019–2020.
- UNESCO; *Guidance for generative AI in education and research*; UNESCO; 2023.
- Unger, R.M.; *The knowledge economy*; Verso; 2019.
- Wikipedia; *Rage bait*; Wikipedia.org; 2025.
- World Health Organization; *Managing the COVID-19 infodemic: promoting healthy behaviours and mitigating the harm from misinformation and disinformation*; WHO; 2020.
- World Health Organization; *Munich Security Conference: “We're not just fighting an epidemic; we're fighting an infodemic”*; WHO; 2020.
- Zhou, X.; Zafarani, R.; *A survey of research on fake news and detection*; *Social Network Analysis and Mining / arXiv*; 2018.
- Zhou, X.; Zafarani, R.; *Fake news: survey, detection, and mitigation*; *ACM Computing Surveys*; 2018.