

<https://dottoremaeveroche.it>

Un progetto della Federazione Nazionale Medici Chirurghi e Odontoiatri (FNOMCeO)

## L'intelligenza artificiale amplifica la disinformazione?

L'articolo è stato pubblicato il 14 maggio, 2024 sul sito <https://dottoremaeveroche.it>

Probabilmente sì: i sistemi di intelligenza artificiale (IA) svolgono un ruolo fondamentale nell'alimentare il fenomeno della disinformazione che gran parte del mondo sta affrontando. Questa straordinaria innovazione tecnologica da una parte rende più semplice la produzione di contenuti falsi ma realistici e, dall'altra, facilita la diffusione su larga scala di questi stessi contenuti a un pubblico mirato<sup>1</sup>. Troll –

persone o entità che intervengono all'interno di una comunità virtuale in modo provocatorio, offensivo o insensato<sup>2</sup> – e bot – programmi che all'interno dei social media comunicano facendo credere agli utenti di interagire con persone fisiche – affollano Internet di bugie e di mezze verità<sup>3</sup>. Comprendere questi fenomeni e queste dinamiche è fondamentale per capire come si diffonde la disinformazione online.

**Dottore, lei parla di bugie e di mezze verità, ma cosa si intende con disinformazione?**

Prendiamo in prestito la definizione della Commissione europea, che ha definito la disinformazione come «un'informazione che possiamo verificare sia falsa o fuorviante, creata, presentata e diffusa a scopo di lucro o per ingannare intenzionalmente gli

utenti e che può causare un pubblico danno». Quando parliamo di «danno pubblico» intendiamo per esempio il rischio di condizionare i processi politici democratici e beni collettivi come la salute dei cittadini, la salubrità dell'ambiente o la sicurezza sociale<sup>4</sup>.

**Possiamo dire che la diffusione della disinformazione è colpa della Rete?**

La disinformazione come strumento della politica non è una novità ed esiste non solo da prima dell'avvento di Internet ma ben prima dell'invenzione della stampa. Però, la Rete ha permesso a falsità, teorie cospirative ed esagerazioni di diffondersi più velo-

cemente e più capillarmente. Diversi studi hanno mostrato che le affermazioni false si diffondono più facilmente di quelle accurate, probabilmente perché sono create proprio per attirare l'attenzione. E poi la tecnologia offre anche molte altre opportunità...

**Quali vantaggi offrirebbe la tecnologia nella diffusione della disinformazione?**

Sappiamo che moltissimi utenti postano e interagiscono in Rete anche centinaia di volte al giorno, al punto che di alcuni attivisti in Rete o influencer si parla come di «individualità postumane»<sup>6</sup>. Ma se pensiamo a un «utente medio» di Facebook o di Instagram in carne e ossa, ci viene in mente qualcuno che potrà postare contenuti in Rete solo alcune volte al giorno. Al contrario, i bot più attivi twittano o postano sui social centinaia di volte al giorno, a qual-

siasi orario, soprattutto riproponendo e rilanciando contenuti e raramente creando qualcosa di originale.

Inoltre, i progressi tecnologici nella produzione di software di editing di audio e video consentono oggi di creare filmati falsi ma sempre più realistici. In questi casi si parla di deep fake: è il caso di quei filmati in cui – per esempio – alcuni uomini politici, personaggi pubblici o dello spettacolo pronunciano discorsi che non hanno mai fatto.

**Quindi per essere definita tale la disinformazione dev'essere intenzionale?**

Sì: fraintendimenti di fatti o errori nei contenuti compiuti in buona fede non possono essere considerati alla stessa stregua della disinformazione. In questi casi possiamo parlare di cattiva informazione o – per esempio quando un contenuto viene rilanciato sui social media senza controllarne la veridicità – di misinformation, traducendo alla lettera un'espressione della lingua inglese. Sebbene non sia intenzionale, anche la misinformation può provocare gravi danni<sup>5</sup>.

**Ma le grandi aziende di tecnologia informatica non potrebbero essere alleate preziose nel contrasto alla disinformazione?**

Certamente: Google, Facebook, X (è questo il nuovo nome di Twitter) e altre note piattaforme online già utilizzano algoritmi «addestrati» con il machine learning per identificare i troll, individuare e rimuovere i falsi account bot e identificare in modo proattivo i contenuti sensibili. Secondo Facebook, il 99,5% delle rimozioni di contenuti terroristici, il 98,5% degli account falsi, il 96% delle rimozioni di nudità e attività sessuali per adulti e l'86% delle rimozioni di contenuti di violenza sono rilevati da strumenti di IA e non innescati dalla segnalazione di utenti<sup>11</sup>.



### Cosa si fa per contrastare la disinformazione alimentata dall'intelligenza artificiale?

Ci sono nel mondo molte iniziative volte a contrastare la disinformazione. Il progetto NewsGuard —che monitora e analizza i siti di informazione assegnando valutazioni in base alla loro affidabilità — supporta le aziende che sviluppano prodotti di IA generativa permettendo loro di accedere e di utilizzare i dati di NewsGuard stessa, al fine di perfezionare i modelli di produzione e diffusione di false informazioni in rete. Sempre a livello internazionale, secondo i dati pubblicati dal Duke Reporters' Lab ci sono più di 400 progetti di fact-checking attivi nati in tantissime nazioni diverse<sup>7</sup> e il numero di queste iniziati-

ve sta crescendo moltissimo negli ultimi anni. Più in generale, preoccupa la difficoltà di arginare grazie all'attività di persone fisiche la disinformazione alimentata dall'IA. Fino a oggi il fact-checking si è basato principalmente sull'analisi "manuale" messa in atto da persone fisiche per verificare l'attendibilità delle informazioni. Sebbene l'opera di demistificazione e correzione di notizie false o infondate — il cosiddetto debunking — sia preziosa e possa ancora avere giustificazione e utilità<sup>8</sup>, dato che il volume stesso della disinformazione continua a crescere il fact-checking svolto da persone fisiche è

considerato sempre meno capace di filtrare e valutare l'enorme mole di informazione pubblicata online<sup>9</sup>. Inoltre, restano le difficoltà nel raggiungere le persone più esposte o vulnerabili alla disinformazione. Una prospettiva convincente — praticata anche dalla FNOMCeO con il progetto *Dot-tore ma è vero che* e di recente sottolineata dal filosofo statunitense Lee McIntyre<sup>10</sup> — è la necessità di intensificare la produzione e la disseminazione istituzionale di contenuti corretti, investendo soprattutto nei buoni influencer, "messaggeri" di informazione corretta e basata su evidenze.

### Come utenti possiamo fare qualcosa?

Chi studia il cambiamento dei modi con i quali ci informiamo raccomanda di attingere a una varietà di fonti informative, di riflettere in modo critico quando leggiamo qualcosa sui social media e di pensarci due volte prima di ripubblicare contenuti virali. Altrimenti, sostengono diversi ricercatori, la disinformazione continuerà a circolare e gli utenti continueranno a diffonderla<sup>9</sup>.

Riguardo l'IA, la prudenza nei confronti di un'innovazione non significa respingerla o rifiutarla. Piuttosto, vuol dire impegnarsi in un dialogo critico e riflessivo su come e perché queste tecnologie sono pensate e costruite<sup>14</sup>.

### Quindi l'intelligenza artificiale potrà aiutare a contrastare la disinformazione?

Le soluzioni di IA si stanno già rivelando efficaci nel rilevare e rimuovere contenuti illegali, dubbi e indesiderati online. Le tecniche di IA hanno avuto successo anche nel monitoraggio e nell'identificazione di account fasulli: in questo caso si parla di tecniche come bot-spotting e bot-labeling<sup>12</sup>. In altre parole, si "etichettano" gli account fasulli, così che gli utenti possano inquadrare correttamente i contenuti con cui si confrontano e giudicarne autonomamente la veridicità.

Nel prossimo futuro è molto probabile che assisteremo a un duello tra l'IA sfruttata in modo sempre più sofisticato per alterare la realtà e l'IA utilizzata per identificare e bloccare i contenuti falsi o fuorvianti<sup>13</sup>.

Ad ogni modo, già oggi alcuni strumenti di IA aiutano le redazioni delle riviste scienti-

fiche a controllare la qualità e l'attendibilità dei contributi che ricevono per la pubblicazione. Per esempio, il software SciScore è capace di verificare i metodi seguiti dagli autori di uno studio clinico o epidemiologico, nonché la complessiva trasparenza di metodi e conclusioni del lavoro. RobotReviewer è invece uno strumento sviluppato con il machine learning che aiuta gli utenti a verificare le informazioni alla base di una ricerca scientifica, ai modi in cui uno studio è stato condotto e riguardo possibili rischi di condizionamenti. Un software come Proofig è molto utilizzato dalle redazioni scientifiche per identificare le immagini contraffatte, duplicate o manipolate che sempre più spesso sono contenute in manoscritti sottoposti alla valutazione editoriale per la pubblicazione.

### Bibliografia

- Bontridder N, Poulet Y. The role of artificial intelligence in disinformation. *Data & Policy* 2021; 3: e32
- Troll. Vocabolario online Treccani. Disponibile su: [https://lc.cx/RBlco\\_](https://lc.cx/RBlco_) [ultimo accesso 8 maggio 2024].
- Swenson A, Chan K. Election disinformation takes a big leap with AI being used to deceive worldwide. *Associated Press* 2024; 14 marzo.
- European Commission. Tackling online disinformation. Disponibile su: <https://lc.cx/3htL3Y> [ultimo accesso 7 maggio 2024].
- Puglisi F. Misinformazione e disinformazione medica sui social media. *Oncotwitting* 2022; 4 aprile
- Bennato D. *La società del XXI secolo. Persone, dati, tecnologie*. Roma-Bari: Laterza, 2024.
- Duke reporters LAB. <https://reporterslab.org/fact-checking/> [ultimo accesso 8 maggio 2024].
- Buquicchio C, Pulcinelli C, Romersi D. *La comunicazione nelle emergenze sanitarie*. Roma: Il Pensiero Scientifico Editore, 2023.
- Kertysova K. Artificial intelligence and disinformation: how AI changes the way disinformation is produced, disseminated, and can be countered. *Security Human Rights* 2018; 29: 55-81.
- McIntyre L. *On disinformation. How to fight for truth and protect democracy*. Cambridge, MA: MIT Press, 2023.
- Marsden C, Meyer T. *Regulating disinformation with artificial intelligence: effects of disinformation initiatives on freedom of expression and media pluralism*. European Parliament 2019; marzo.
- Wild J, Godart C. Spotting bots, cyborgs and inauthentic activity. *Verification handbook: for disinformation and media manipulation*. European Journalism Centre. Disponibile su: <https://lc.cx/UB2K56> [ultimo accesso 8 maggio 2024].
- Rosenbach E, Mansted K. *Can democracy survive in the information age?* Belfer Center for Science and International Affairs 2018; 30 Oct.
- Cabitza F. *Deus in machina. L'uso umano delle nuove macchine, tra dipendenza e responsabilità*. In: Floridi L, Cabitza F. *Intelligenza artificiale. L'uso delle nuove macchine*. Firenze-Milano: Giunti, 2021.