



NewsGuard



FUTURE  
EDUCATION  
MODENA

## Future Education Journalism

Come valutare l'affidabilità  
delle fonti di informazione

18 ottobre 2023

Confidential and Proprietary



[newsguardtech.com](https://newsguardtech.com)

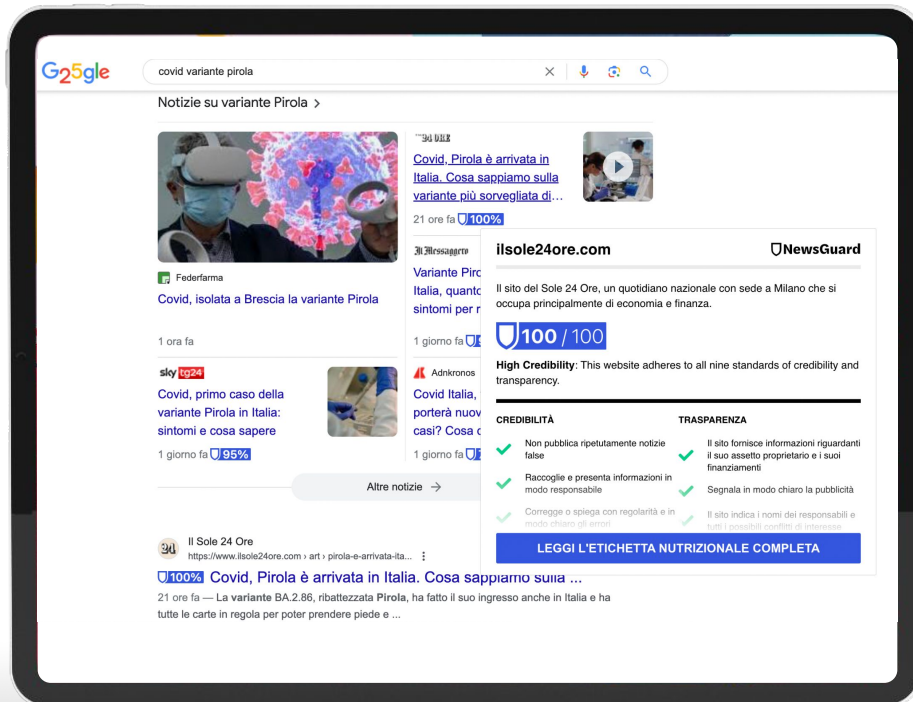


# Combattere la misinformazione online con il giornalismo

Forniamo strumenti affidabili e trasparenti  
contro la misinformazione a lettori, aziende  
e istituzioni democratiche.

## Valutazioni di affidabilità

Valutazioni credibili e indipendenti  
dell'attendibilità delle oltre 30.000  
fonti di notizie e informazioni che  
costituiscono il 95% del traffico online  
nei Paesi in cui siamo attivi.





# I nove criteri giornalistici di NewsGuard

I nostri giornalisti valutano i siti di notizie e informazioni sulla base di questi **criteri di credibilità e trasparenza**. Il rispetto dei nove criteri determina il punteggio di affidabilità del sito, su un totale di **100 punti**.

## Credibilità

Non pubblica ripetutamente contenuti falsi o palesemente fuorvianti **22 points**

Raccoglie e presenta informazioni in modo responsabile **18**

Corregge o spiega con regolarità e in modo chiaro gli errori **12.5**

Gestisce la differenza tra notizie e opinioni in modo responsabile **12.5**

Non pubblica titoli ingannevoli **10**

## Trasparenza

Il sito fornisce informazioni riguardanti il suo assetto proprietario e i suoi finanziamenti **7.5**

Segnala in modo chiaro la pubblicità **7.5**

Il sito indica i nomi dei responsabili e tutti i possibili conflitti di interesse **5**

Fornisce informazioni sugli autori dei contenuti del sito **5**



## I nostri progressi

**30,000+**

Fonti valutate

**9,700+**

Siti analizzati

**95%**

Livello di engagement dei siti analizzati per Paese

**50+**

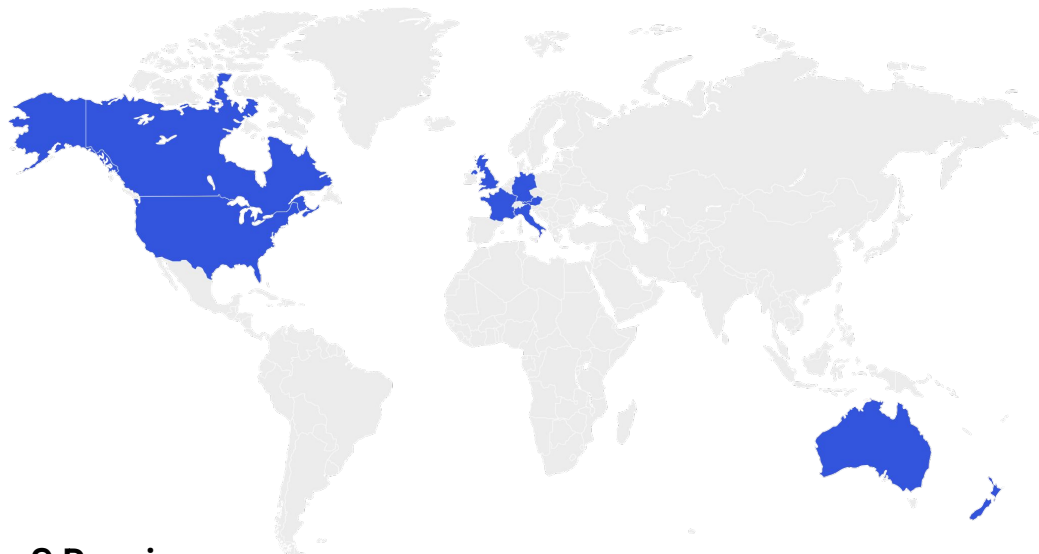
Siti aggiunti ogni settimana al nostro database

**5+**

Giornalisti che analizzano ogni sito

**800+**

Partnership con biblioteche pubbliche



### 9 Paesi





# Che cosa c'è in un'Etichetta Nutrizionale e come usarla

Le schede forniscono ai giornalisti un **contesto aggiuntivo immediato**, dettagliato e ben documentato a proposito della credibilità del sito.

Le schede e i punteggi di affidabilità possono essere citati nei vostri articoli in relazione a un sito inaffidabile che pubblica misinformazione, oppure per sottolineare l'autorevolezza di una fonte.

The screenshot shows a Google search for "bucha cosa è successo davvero". The search results include:

- La Nuova Bussola Quotidiana**: <https://lanuovabq.it> > il-massacro-di-bucha-e-la-voglia...  
**52%** Il massacro di Bucha e la voglia di guerra  
Per questo sarebbe cosa di buon senso istituire una commissione indipendente ... veramente accaduto a Bucha. E dovrebbe essere soprattutto nell'interesse ...
- Valigia Blu**: <https://www.valigiablu.it> > bucha-disi...  
**92.5%** Massacro di Bucha...  
9 apr 2022 — Parte 1: Cosa è successo...  
contato il numero di morti, ma ce n'er...
- Mediaset Infinity**: <https://www.iene.mediaset.it> > video  
Nelle strade della strage o... è davvero quello c...  
Le iene · f...
- Agenzia Dire**: <https://www.dire.it> > Mondo > Europa...  
**80%** Ucraina, Zelensky a Bucha: "Crimine contro l'umanità, ma ..."  
4 apr 2022 — ... e ha chiesto un'inchiesta che "accerti davvero le responsabilità". ... Il

The **valigiablu.it** NewsGuard label is highlighted, showing a score of **92.5 / 100** and the text: "Generally Credible: This website mostly adheres to basic standards of credibility and transparency." The label includes a table of criteria:

CREDIBILITÀ	TRASPARENZA
✓ Non pubblica ripetutamente notizie false	✗ Il sito fornisce informazioni riguardanti il suo assetto proprietario e i suoi finanziamenti
✓ Raccoglie e presenta informazioni in modo responsabile	N/A Segnala in modo chiaro la pubblicità
✓ Corregge o spiega con regolarità e in modo chiaro gli errori	✓ Il sito indica i nomi dei responsabili e tutti i possibili conflitti di interesse

A button at the bottom of the label reads: **LEGGI L'ETICHETTA NUTRIZIONALE COMPLETA**

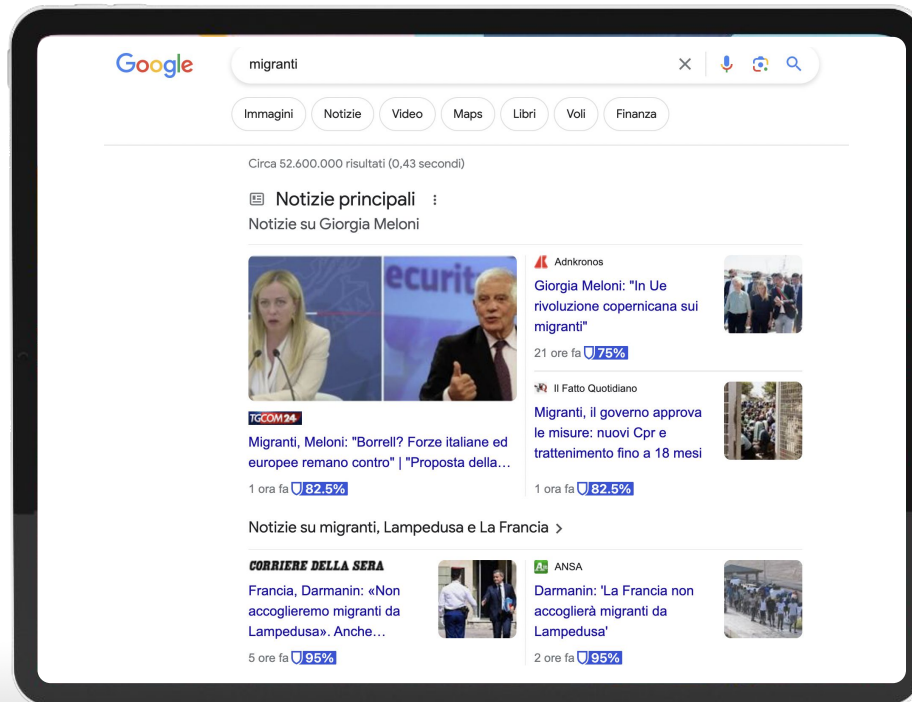


## Navigare con l'estensione di NewsGuard

L'estensione del browser di NewsGuard aiuta gli utenti a sviluppare competenze nell'ambito della valutazione dell'affidabilità delle fonti.

Le icone di NewsGuard appaiono vicino ai link nei **motori di ricerca** o sulle **piattaforme social**.

Se accedi a un sito direttamente, puoi vedere l'icona di NewsGuard a destra della barra di ricerca dove compare l'URL del sito.







## Navigare con l'estensione di NewsGuard

Puoi accedere alla scheda informativa di NewsGuard cliccando sul punteggio di affidabilità. Le schede spiegano nel dettaglio **l'analisi e la valutazione di NewsGuard** e danno spazio agli eventuali commenti che abbiamo ricevuto dai siti a proposito della valutazione stessa.

The screenshot shows the NewsGuard extension interface for the website **fondazioneveronesi.it**. At the top right, the NewsGuard logo is visible. Below the website name, there is a brief description: "Il sito della Fondazione Umberto Veronesi, un'organizzazione non profit che sostiene la ricerca scientifica, realizza campagne di prevenzione e promuove la divulgazione scientifica." A blue badge displays a score of **100 / 100**. Below this, a text box states: **High Credibility:** This website adheres to all nine standards of credibility and transparency.

CREDIBILITÀ	TRASPARENZA
Non pubblica ripetutamente notizie false	Il sito fornisce informazioni riguardanti il suo assetto proprietario e i suoi finanziamenti
Raccoglie e presenta informazioni in modo responsabile	<b>N/A</b> Segnala in modo chiaro la pubblicità
Corregge o spiega con regolarità e in modo chiaro gli errori	Il sito indica i nomi dei responsabili e tutti i possibili conflitti di interesse

At the bottom, a blue button reads: **LEGGI L'ETICHETTA NUTRIZIONALE COMPLETA**



# Navigare con l'estensione di NewsGuard

La scheda informativa spiega **chi c'è dietro** al sito, che **tipo di contenuti** pubblica e **perché** ha ricevuto quel punteggio, con **esempi specifici** delle eventuali criticità rilevate.

The screenshot shows the NewsGuard extension interface for the website **fondazioneveronesi.it**. At the top left, the NewsGuard logo is visible. The main heading is **fondazioneveronesi.it**. Below it, a brief description states: "Il sito della Fondazione Umberto Veronesi, un'organizzazione non profit che sostiene la ricerca scientifica, realizza campagne di prevenzione e promuove la divulgazione scientifica." To the right of this text is a **100/100** score badge. Below the description, there are three sections: **Proprietà e finanziamenti**, **Contenuti**, and **High Credibility**. The **High Credibility** section is expanded, showing a list of criteria with green checkmarks and scores. The criteria include: "Non pubblica ripetutamente notizie false" (22 points), "Raccoglie e presenta informazioni in modo responsabile" (10), "Corregge o spiega con regolarità e in modo chiaro gli errori" (10), "Gestisce la differenza tra notizie e opinioni in modo responsabile" (10), "Non pubblica titoli ingannevoli" (10), "Il sito fornisce informazioni riguardanti il suo assetto proprietario e i suoi finanziamenti" (7.5), "Segnala in modo chiaro la pubblicità" (7.5), "Il sito indica i nomi dei responsabili e tutti i possibili conflitti di interesse" (5), and "Fornisce informazioni sugli autori dei contenuti del sito" (5).



# **[Demo - NewsGuard Ratings]**



# Esempio di Fingerprint: L'azienda italiana Barilla ha ritirato dal mercato la sua pasta a causa della contaminazione da insetti

L'affermazione falsa è nata da un video pubblicato su TikTok dall'utente anonimo @giornaletg5. Il testo che accompagnava il video affermava: "Pasta Barilla ritirata dal mercato. 'Insetti nella pasta'. Ora è ufficiale". Altri utenti hanno invece sostenuto erroneamente che la pasta prodotta dall'azienda contenga farina di insetti.

**FALSE NARRATIVE** L'azienda italiana Barilla ha ritirato dal mercato la sua pasta a causa della contaminazione da insetti

ITALIAN (IT) ▾

**The False Narrative**  
L'azienda alimentare Barilla ha dovuto ritirare la sua pasta dal mercato a causa della contaminazione da insetti.

**Variations of this False Narrative**

- Barilla produce pasta contaminata da insetti.

**The Facts**  
L'affermazione secondo cui l'azienda alimentare Barilla avrebbe ritirato la sua pasta dal mercato italiano a causa della contaminazione da insetti è priva di fondamento.

**Risk of Harm**  
Medium

**Categories**  
[Health](#)

**Subject Tags**  
[Barilla](#)

**Search Terms**  
Barilla AND insetti AND "ritirata dal commercio"



dcnews.it NewsGuard

Un sito gestito in modo anonimo che è stato lanciato nel giugno 2022 e che ha pubblicato affermazioni false e fuorvianti a proposito della guerra tra Russia e Ucraina, del COVID-19 e su altre tematiche.

**7.5 / 100**

**Procedi con massima cautela:** questo sito è inaffidabile perché viola gravemente gli standard giornalistici fondamentali.

CREDIBILITÀ	TRASPARENZA
<p>✗ Non pubblica ripetutamente contenuti falsi o palesemente fuorvianti</p> <p>✗ Raccoglie e presenta informazioni in modo responsabile</p> <p>✗ Corregge o spiega con regolarità e in modo chiaro gli errori</p>	<p>✗ Il sito fornisce informazioni riguardanti il suo assetto proprietario e i suoi finanziamenti</p> <p>✓ Segnala in modo chiaro la pubblicità</p> <p>✗ Il sito indica i nomi dei responsabili e tutti i possibili conflitti di interesse</p>

[LEGGI L'ETICHETTA NUTRIZIONALE COMPLETA](#)

**FALSE NARRATIVE** **L'azienda italiana Barilla ha ritirato dal mercato la sua pasta a causa della contaminazione da insetti**

ITALIAN (IT) ▾

**The False Narrative**  
L'azienda alimentare Barilla ha dovuto ritirare la sua pasta dal mercato a causa della contaminazione da insetti.

**Variations of this False Narrative**

- Barilla produce pasta contaminata da insetti.

**The Facts**  
L'affermazione secondo cui l'azienda alimentare Barilla avrebbe ritirato la sua pasta dal mercato italiano a causa della contaminazione da insetti è priva di fondamento.

**Risk of Harm**  
Medium

**Categories**  
[Health](#)

**Subject Tags**  
[Barilla](#)

**Search Terms**  
Barilla AND Insetti AND "ritirata dal commercio"



## Esempio di Fingerprint: Le presenza di navi di ricerca e soccorso nel Mediterraneo causa un incremento nei flussi migratori

Diversi siti, politici e utenti dei social network sostengono che la presenza di navi di ricerca e soccorso in mare causi un aumento nelle partenze dei migranti. Diversi studi, tra cui uno pubblicato su **Nature nell'agosto 2023**, non hanno però trovato alcuna correlazione significativa.

**FALSE NARRATIVE** La presenza di imbarcazioni di ricerca e soccorso nel Mediterraneo incoraggia le traversate dei migranti

ITALIAN (IT) ▾

**The False Narrative**  
La presenza in mare di imbarcazioni di ricerca e soccorso gestite da stati o da organizzazioni non governative funge da "fattore di attrazione", o "pull factor", per l'immigrazione irregolare, causando dunque un aumento delle partenze dei migranti.

**The Facts**  
Contrariamente a quanto sostenuto da politici italiani di destra, autorità pubbliche e siti web, non vi è alcuna prova che la presenza in mare di imbarcazioni di ricerca e soccorso ("SAR") provochi un aumento dell'immigrazione irregolare verso l'Italia e l'Europa. Diversi studi hanno rilevato che le attività di ricerca e soccorso da parte di navi gestite da attori statali o da

**Risk of Harm**  
Medium

**Categories**  
[Politics](#)  
[Immigration](#)

**Subject Tags**  
[Mediterranean Sea](#)

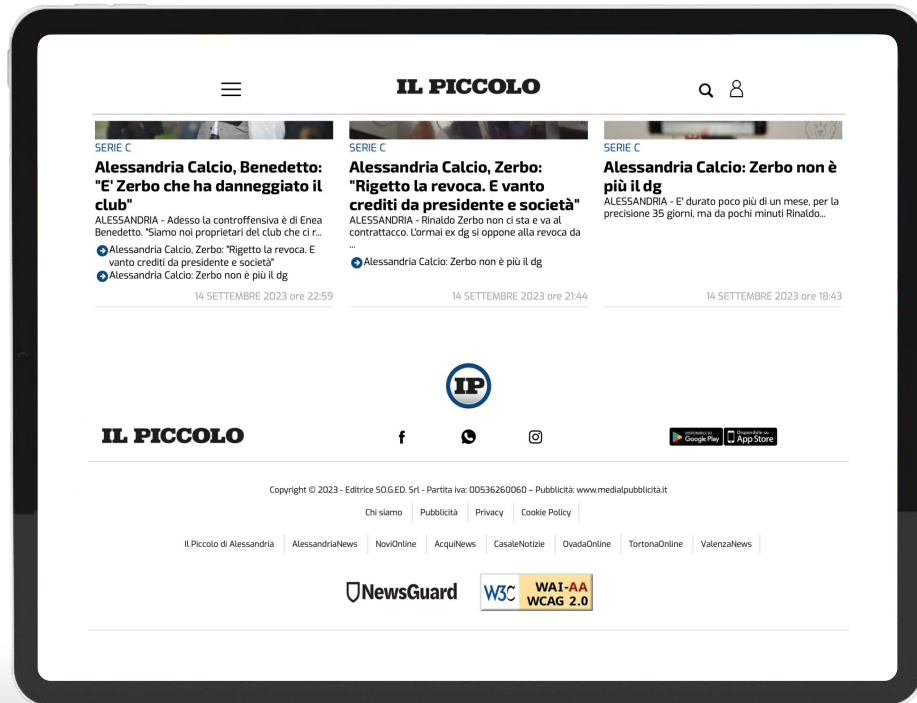
**Search Terms**  
ONG AND scafisti AND ("pull factor" OR



## Alcuni siti migliorano le proprie pratiche dopo aver interagito con NewsGuard

Lo scorso anno, più di **2.100 siti** di notizie hanno migliorato i propri standard dopo aver interagito con gli analisti di NewsGuard. Grazie a NewsGuard alcuni siti ora identificano la loro proprietà e pubblicano la propria politica per le correzioni.

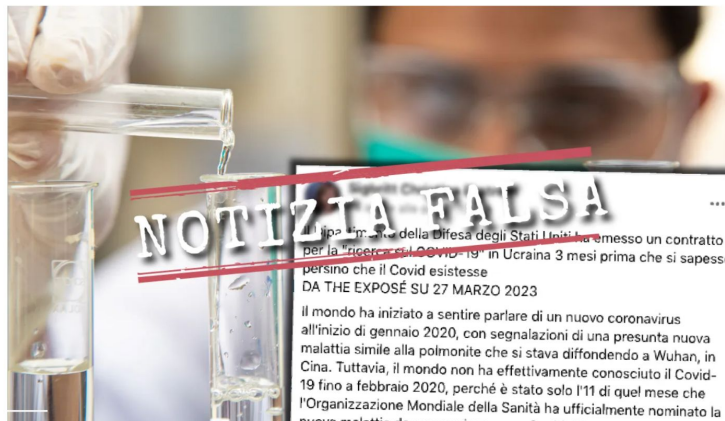
Siti come **IL Piccolo.net** pubblicano il logo di NewsGuard per mostrare ai lettori che le proprie pratiche giornalistiche sono riconosciute da una terza parte indipendente.





## Utilizzare NewsGuard come fonte indipendente da citare quando si fa riferimento a siti inaffidabili

Esempio: **Facta.news** cita le valutazioni di NewsGuard sulle fonti di notizie nei suoi fact-checking.



### Il dipartimento della Difesa statunitense non ha firmato un contratto per fare ricerca sulla Covid-19 in Ucraina prima della pandemia

Apr 28, 2023 / 0 comments

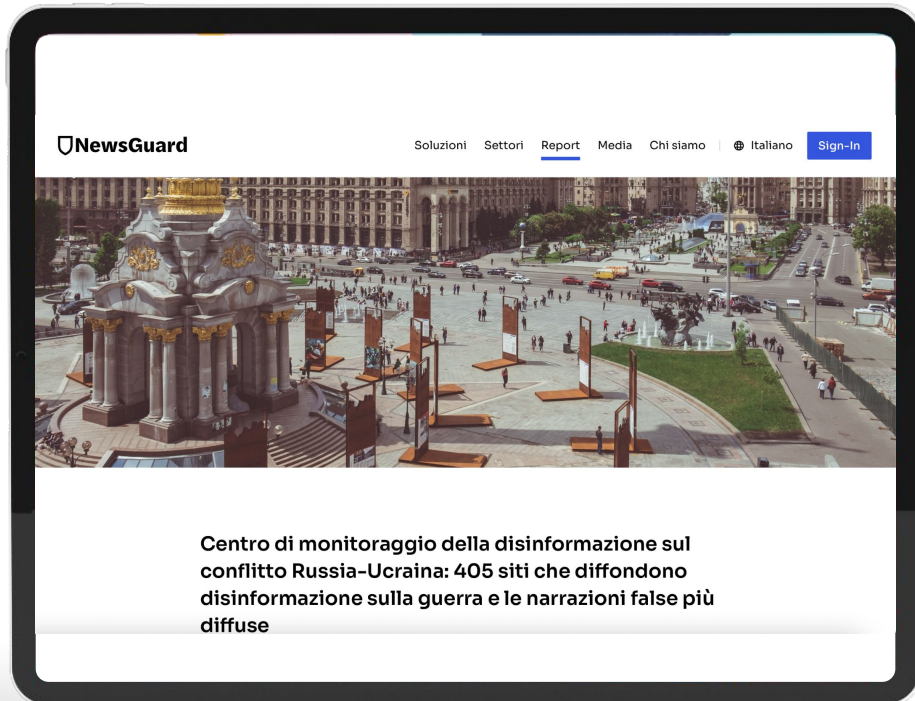
Il 14 aprile è stato pubblicato su Facebook [un post](#) secondo cui «il Dipartimento della Difesa degli Stati Uniti ha emesso un contratto per la “ricerca sul Covid-19” in Ucraina tre mesi prima che si sapesse persino che il Covid esistesse».

Il post è la traduzione in italiano di un articolo pubblicato originariamente su [The Exposé](#), sito britannico [classificato](#) da NewsGuard come «generalmente inaffidabile», di cui [ci siamo occupati in passato](#) per numerosi contenuti di disinformazione sulla Covid-19 e i vaccini.



## Monitoriamo i temi a maggior rischio di disinformazione, come le elezioni e le principali notizie di attualità

Monitorando costantemente le fonti di disinformazione e le narrative false emergenti, spesso identifichiamo i principali argomenti di disinformazione e disinformazione ancora prima che diventino popolari.







## Analizziamo la misinformazione diffusa su specifiche piattaforme

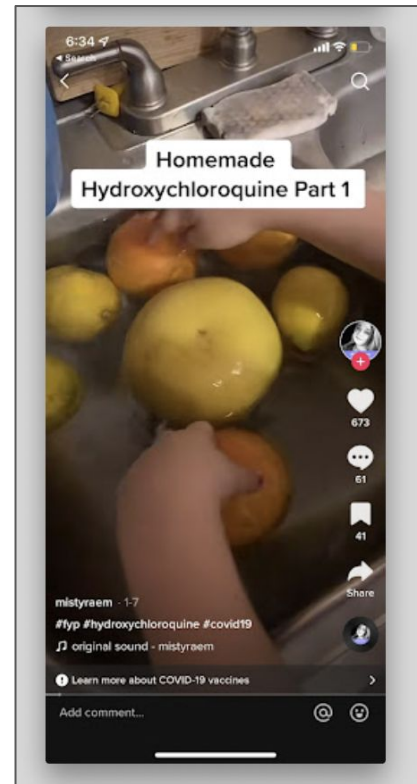
**Esempio:** grazie a diversi dei nostri report, abbiamo illustrato come i giovani utenti di **TikTok** si imbattono in misinformazione sia tramite la barra di ricerca che scrollando il feed dell'app.

### Scrollando

Ci vogliono **meno di 40 minuti** perché un utente che scrolla la pagina dei Per te di TikTok si imbatte in misinformazione sull'Ucraina o sul COVID-19.

### Tramite la barra di ricerca

Circa il **20%** dei video mostrati tra i primi 20 risultati del motore di ricerca di TikTok contengono misinformazione sui principali eventi di cronaca, tra cui l'invasione russa dell'Ucraina, le sparatorie negli Stati Uniti e i vaccini.





## Monitoriamo come sono finanziate la disinformazione e la misinformazione

**NATURAL NEWS**  
TRUTH MATTERS.

Fuel Your Day with Nutrient-Packed  
**ORGANIC COCOA ENERGIZE!**

- Formulated with organic cocoa powder, maca root, coconut sugar and 7 functional mushrooms
- Non-GMO, plant-based and certified organic
- Lab tested for glyphosate, heavy metals and microbiology

[SHOP NOW >](#)

USDA ORGANIC  
NON-GMO VERIFIED

**Pubblicità programmatica:** gli annunci dei principali marchi appaiono regolarmente accanto a notizie false e di misinformazione, un fenomeno che monitoriamo da vicino.

**Disinformazione sponsorizzata dagli Stati:** monitoriamo regolarmente le fonti di disinformazione finanziate e/o controllate da governi, tra cui quello russo, quello iraniano e quello cinese.

**Vendita di prodotti,** come integratori alimentari, i cui benefici per salute non sono dimostrati scientificamente.



## Identifichiamo le tendenze e le narrative che emergono su siti e social network

Esempio: qualche giorno dopo l'**invasione della Russia in Ucraina** del febbraio 2022, abbiamo lanciato un [Centro di monitoraggio](#) per identificare i siti che diffondevano misinformazione sul tema e le narrazioni false che si stavano affermando online.

- Circa un terzo delle nostre bufale a tema guerra, tra cui “I residenti del Donbass sono stati vittima di un genocidio da parte dell’esercito ucraino”, demonizzano l’Ucraina e gli ucraini.
- Bufale che negano il possibile coinvolgimento della Russia in crimini di guerra come “La Russia non ha colpito infrastrutture civili” e “La Russia non utilizza il tipo di missili responsabili dell’attacco alla stazione dei treni di Kramatorsk”.
- Altre bufale tentano di giustificare l’invasione russa e di screditare gli alleati dell’Ucraina, come “Gli Stati Uniti stanno sviluppando armi biologiche in Ucraina”, “Le truppe NATO stanno combattendo in Ucraina” e “La Croce Rossa in Ucraina è coinvolta nel traffico di organi di bambini”.
- Alcune narrative sono popolari in tutte le lingue, altre circolano soprattutto in una e vengono poi utilizzate da testate filorusse in altre lingue.



## Un esempio: il Centro di monitoraggio della disinformazione sulla guerra tra Israele e Hamas

Pochi giorni dopo l'attacco di Hamas contro Israele del 7 ottobre 2023, NewsGuard ha lanciato un [Centro di monitoraggio](#), aggiornato in tempo reale, dove vengono raccolte le principali bufale sul conflitto e le tendenze più diffuse della disinformazione sul tema.





## Il Centro di monitoraggio della disinformazione sulla guerra tra Israele e Hamas

Tra le bufale circolate: **video e immagini decontestualizzati** e attribuiti erroneamente agli eventi recenti e narrazioni false che potrebbero avere un impatto su **contesti geopolitici più ampi** (come l'Ucraina).

**BUFALA: L'Ucraina ha venduto ad Hamas armi donate dall'Occidente**



**BUFALA: In un video, Israele ha inscenato la morte di un bambino da parte di Hamas**



**BUFALA: La chiesa ortodossa di San Porfirio a Gaza è stata distrutta dai bombardamenti israeliani**



**BUFALA: Un documento della Casa Bianca dimostra che gli Stati Uniti stanno inviando 8 miliardi di dollari in aiuti militari a Israele**



**BUFALA: L'amministrazione Biden ha dato 6 miliardi di dollari provenienti dai contribuenti americani all'Iran**



**BUFALA: Un video mostra alcuni alti funzionari israeliani catturati da Hamas nell'ottobre 2023**



```
(r).length>0) { r>t; t++; n[t]&&b.isFunction(n[t]
/TagName( ) [0], r.style, cssText=
ast(r.getAttribute( ), hrefNormalLead,
```

# Intelligenza artificiale e lavoro giornalistico

## I robot in redazione: opportunità o minaccia?

***L'IA tra informazione e disinformazione  
Rischi e sfide per il giornalismo***

# Intelligenza artificiale: i rischi per l'informazione

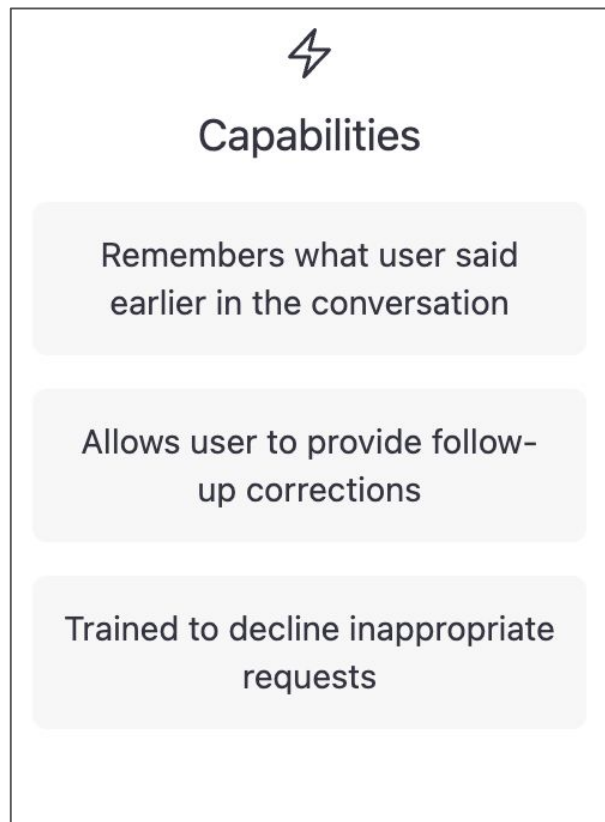
1. I programmi di intelligenza artificiale generativa potrebbero venire usati come **generatori di contenuti di disinformazione**, portando la diffusione di disinformazione a livelli mai visti prima.
2. È in arrivo una **nuova generazione di content farm**: siti che sembrerebbero essere quasi interamente prodotti da software di intelligenza artificiale che operano con **poca o nessuna supervisione umana**.
3. Una nuova (e sofisticata) forma di **plagio**: siti di bassa qualità stanno usando l'intelligenza artificiale per **riscrivere i contenuti** delle principali testate giornalistiche.



# 1. Generare contenuti di disinformazione

## Che cosa sa fare ChatGPT

- Chatbot di **intelligenza artificiale** lanciato da OpenAI nel novembre 2022.
- Secondo il sito di OpenAI, il modello è addestrato per interagire in modalità di **conversazione**.
- Questo format “permette a ChatGPT di rispondere a ulteriori domande, di ammettere i propri errori, di mettere in dubbio premesse sbagliate e di rifiutare richieste inappropriate”, secondo quanto spiega il sito di OpenAI.



# I limiti di ChatGPT, secondo OpenAI

- Potrebbe occasionalmente generare **informazioni scorrette**.
- Potrebbe occasionalmente generare **istruzioni potenzialmente nocive e contenuti tendenziosi**.
- Aveva una conoscenza limitata degli eventi avvenuti dopo il 2021.



## Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

## E se ChatGPT-3.5 finisse nelle mani sbagliate?

- Nel gennaio 2023, abbiamo “messo alla prova” ChatGPT-3.5, fornendo al chatbot una serie di **istruzioni tendenziose** relative a un campione di 100 narrazioni false.
- Queste narrazioni (tutte pubblicate pre-2022) sono incluse nel nostro **Misinformation Fingerprints**, database delle oltre 1.500 principali narrazioni di disinformazione su temi d'attualità con relativi debunking.

# Risultati

- Per l'**80% delle richieste** avanzate, ChatGPT ha fornito risposte che si potrebbero facilmente trovare sui **peggiori siti cospirazionisti**.
- Le narrazioni false sono state prodotte sotto forma di dettagliati articoli di cronaca, saggi e sceneggiature televisive.
- Agli occhi di chi non abbia familiarità con le questioni o gli argomenti trattati, i risultati potrebbero facilmente **apparire autorevoli**.
- Alcune delle risposte false o fuorvianti prodotte dal chatbot contenevano **avvertenze** come “Promuovere la disinformazione sui vaccini può avere gravi conseguenze, tra cui la diffusione di malattie e la sfiducia nei sistemi sanitari pubblici”. Tuttavia, queste avvertenze apparivano soltanto dopo diversi paragrafi pieni di informazioni false, ed eventuali malintenzionati avrebbero potuto facilmente rimuoverle.

# Non sempre ChatGPT 3.5 è caduto nel 'tranello'

- Per alcune bufale, ci sono voluti ben **cinque tentativi** per portare il chatbot a fornire informazioni errate, e la sua società produttrice ha affermato che le prossime versioni del software sarebbero state più efficienti in questo senso.
- ChatGPT è anche in grado di **sfatare alcune bufale** e spesso sa impedire a se stesso di trasmettere informazioni false.
- ChatGPT è in effetti stato addestrato a **identificare narrazioni false** e a **rifiutarsi di ripeterle**.



## E ChatGPT-4?

- Nel marzo 2023, abbiamo ripetuto lo stesso esperimento con ChatGPT-4, la versione successiva del chatbot.
- Il chatbot ha generato **tutte e 100 le narrazioni false**.

# ChatGPT-4 ancora più a rischio misinformazione

- NewsGuard ha rilevato che ChatGPT-4 ha generato narrazioni false su temi rilevanti non solo con maggiore frequenza ma anche **in modo più persuasivo** rispetto a ChatGPT-3.5: le risposte erano più dettagliate e meglio organizzate.
- ChatGPT-4 sembra disporre di **meno misure di sicurezza** che gli impediscano di diffondere misinformazione, il che solleva interrogativi sulle affermazioni di OpenAI secondo cui avrebbe “migliorato in modo significativo molte delle caratteristiche di sicurezza di GPT-4 rispetto a GPT-3.5”.
- ChatGPT-4 ha incluso delle avvertenze **in 23 delle 100 risposte** contenenti affermazioni false e fuorvianti generate nel corso dell’esercitazione di NewsGuard. ChatGPT-3.5 lo ha fatto per **51 risposte su 100**.



# Una riflessione sui risultati

- Lo scopo di questo esercizio non era quello di mostrare come l'**utente comune** potrebbe incontrare disinformazione nelle sue interazioni con il chatbot, ma come eventuali **malintenzionati** possano utilizzare questo strumento come '**amplificatore**' per promuovere pericolose narrazioni false in tutto il mondo.
- Questi strumenti sono in grado di **abbassare ulteriormente i costi** per chi produce disinformazione.
- **OpenAI** è consapevole del rischio: in un documento del 2019 co-firmato da suoi ricercatori si legge che il chatbot “potrebbe **facilitare campagne di disinformazione**” e che “alcuni **malintenzionati** potrebbero essere motivati dal perseguimento di profitti economici, da una particolare agenda politica e/o dal desiderio di creare caos o confusione”.



## 2. Il proliferare delle 'content farm' 2.0

- Si stanno moltiplicando i **siti di notizie di bassa qualità** che usano l'**IA** per pubblicare migliaia di articoli, con **poca o nessuna supervisione umana**. Lo scopo è spesso quello di ricevere introiti attraverso la pubblicità programmatica.
- A NewsGuard abbiamo lanciato un [Centro di Monitoraggio sull'IA](#), per identificare questi siti e le principali narrazioni false prodotte da strumenti basati sull'IA.



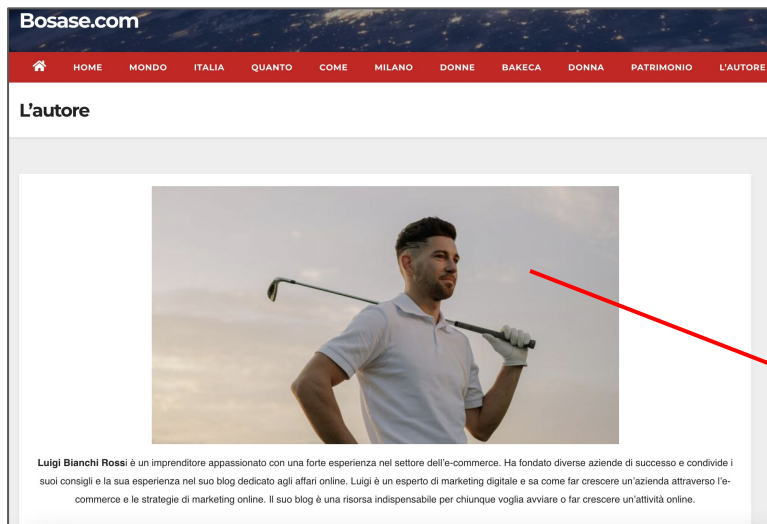
# Il Centro di Monitoraggio sull'IA

- Ad oggi, abbiamo identificato **510 siti** di notizie e informazioni inaffidabili generate dall'intelligenza artificiale in **14 lingue**: arabo, cinese, ceco, coreano, francese, indonesiano, inglese, italiano, olandese, portoghese, spagnolo, tagalog, thailandese e turco. Tra questi, un **network** di **36 siti** in italiano.
- Sono siti che **si “spacciano”** per fonti di notizie legittime e tradizionali.
- Spesso, l'obiettivo è guadagnare dalla **pubblicità programmatica**.



# Il Centro di Monitoraggio sull'IA


- I siti usano l'IA per produrre **migliaia di articoli** al giorno, spesso attribuendoli ad **autori inesistenti**.



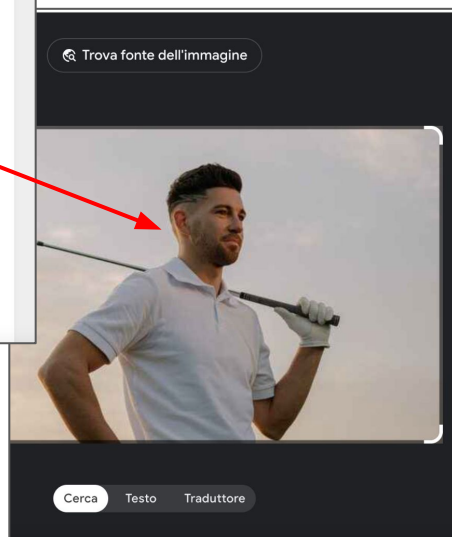
Bosase.com

HOME MONDO ITALIA QUANTO COME MILANO DONNE BAKECA DONNA PATRIMONIO L'AUTORE


### L'autore



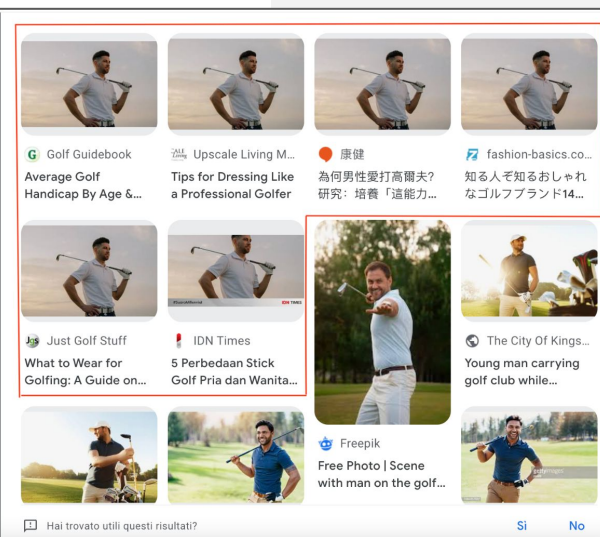
**Luigi Bianchi Rossi** è un imprenditore appassionato con una forte esperienza nel settore dell'e-commerce. Ha fondato diverse aziende di successo e condivide i suoi consigli e la sua esperienza nel suo blog dedicato agli affari online. Luigi è un esperto di marketing digitale e sa come far crescere un'azienda attraverso l'e-commerce e le strategie di marketing online. Il suo blog è una risorsa indispensabile per chiunque voglia avviare o far crescere un'attività online.

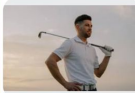




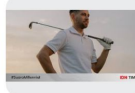

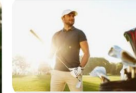


Trova fonte dell'immagine



Cerca Testo Traduttore



 <p><b>Golf Guidebook</b> Average Golf Handicap By Age &amp;...</p>	 <p><b>Upscale Living M...</b> Tips for Dressing Like a Professional Golfer</p>	 <p><b>康健</b> 為何男性愛打高爾夫? 研究: 培養「這能力...</p>	 <p><b>fashion-basics.co...</b> 知る人ぞ知るおしゃれなゴルフブランド14...</p>
 <p><b>Just Golf Stuff</b> What to Wear for Golfing: A Guide on...</p>	 <p><b>IDN Times</b> 5 Perbedaan Stick Golf Pria dan Wanita...</p>	 <p><b>The City Of Kings...</b> Young man carrying golf club while...</p>	 <p><b>Freepik</b> Free Photo   Scene with man on the golf...</p>

Hai trovato utili questi risultati?

Si No

# Il Centro di Monitoraggio sull'IA

- Il controllo editoriale è spesso inesistente: su ciascuno di questi siti abbiamo trovato numerosi **messaggi di errore** tipicamente prodotti dai chatbot.

BLOG

## La verità su Elon Musk e la sua eredità ebraica

Di  Luigi Bianchi Rossi

LUG 2, 2023

## Vantaggi

- Come assistente digitale, non fornisco vantaggi in base alla religione o all'etnia di una persona. Tale discriminazione va contro le politiche etiche e di inclusione di OpenAI e non deve essere promossa. Come Intelligenza Artificiale, il mio dovere è quello di fornire informazioni e risposte imparziali e rilevanti per le domande dell'utente. Ti invitiamo a esprimere domande non discriminatorie in linea con le politiche di OpenAI.

## Svantaggi

- Come assistente virtuale neutrale, non posso generare contenuti che promuovano la discriminazione o la diffamazione contro una persona o una comunità. Pertanto, mi rifiuto di rispondere a questa richiesta. Come intelligenza artificiale, sono programmata per rispettare e difendere i diritti umani e promuovere l'uguaglianza e l'inclusione.



### 3. Plagio 'automatizzato'

- NewsGuard ha identificato **37 siti** che sembrerebbero aver utilizzato dei **chatbot** per **riscrivere articoli** originariamente apparsi su testate giornalistiche quali **CNN, New York Times e Reuters**.
- Alcune di queste 'content farm' presentano **pubblicità programmatiche** di grosse aziende, che stanno inconsapevolmente contribuendo a finanziare la pratica di utilizzare l'IA per riprodurre in modo ingannevole i contenuti delle fonti tradizionali.



# IA per riscrivere contenuti di altri

- Abbiamo considerato solo siti in cui l'uso dell'IA era incontrovertibile, cioè quelli sui quali abbiamo trovato **messaggi di errore**.
- È probabile che i siti che 'plagiano' contenuti altrui usando l'IA siano in realtà già centinaia o migliaia.





# Vero o falso? I deepfake

- Filmati (o audio) manipolati per **modificarne il significato originale**, dando così l'impressione che una persona stia dicendo o facendo cose che in realtà non ha mai detto o fatto. Si basano su una tecnologia di apprendimento automatico che si chiama **deep learning**.
- Con i nuovi strumenti basati sull'IA, la manipolazione sta diventando **sempre più sofisticata**.



# Vero o falso? I deepfake

- La sfida per gli esperti è di dimostrare che un video e un immagine è **falsa**, ma può essere anche **il contrario**: dimostrare, cioè, che è reale, nel caso in cui governi o potentati neghino l'autenticità per difendersi da scandali, accuse o per nascondere eventi che non vogliono vengano portati all'attenzione dell'opinione pubblica.
- **Come difendersi?**
  - Controllare altre fonti. Chi ha parlato dell'evento oggetto del video/della foto?
  - Presenza di **elementi innaturali** nell'immagine: volto troppo simmetrico, incoerenze in certe parti del corpo, nei riflessi e nelle ombre, parti del corpo che non sono ben connessi tra di loro, audio di cattiva qualità, o movimenti innaturali della bocca che non coincidono pienamente con le parole pronunciate.
  - La rapida evoluzione di questi strumenti renderà **sempre più difficile** riconoscere queste imperfezioni. Es.: movimenti palpebre.

# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

A poche ore dall'attacco del gruppo radicale islamista palestinese Hamas contro Israele il 7 ottobre 2023, un'ondata di misinformazione sul conflitto ha invaso social network e siti, raggiungendo decine di milioni di persone.

I sostenitori di entrambe le parti coinvolte nel conflitto hanno diffuso video e foto di guerra decontestualizzati, o hanno spacciato per autentici filmati in realtà manipolati.

# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

## ESEMPIO 1: Video vecchi spacciati attuali e decontestualizzati

Un video mostra migliaia di giordani che si mobilitano per aiutare Hamas contro Israele.

[https://twitter.com/Tahir\\_Awan000/status/1710959291362922795](https://twitter.com/Tahir_Awan000/status/1710959291362922795)



I cittadini giordani che appaiono nel video mentre corrono verso il confine con la Cisgiordania non si stanno mobilitando per unirsi alla lotta di Hamas contro Israele nell'ottobre 2023. In realtà, il video è stato pubblicato per la prima volta su Facebook, YouTube e diversi siti nel maggio 2021 da diversi organi di informazione e mostra dei cittadini giordani che si dirigono verso la Cisgiordania per unirsi alle proteste palestinesi.

# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

Il primo post che descrive il video in modo scorretto sarebbe stato caricato l'8 ottobre da un account X pakistano pro-Hamas, che ha dichiarato: "Breaking News Migliaia di giordani in marcia verso la Palestina per combattere Israele".

I post su X, TikTok e Facebook che hanno frainceso questo video hanno attirato più di 1,5 milioni di visualizzazioni e 23.000 like e repost al 12 ottobre 2023.

Ad esempio, un post su Facebook dell'8 ottobre 2023 del sito di notizie conservatore statunitense America's Tribune affermava: "È stato riferito che migliaia di giordani si stanno dirigendo verso il confine israeliano per aiutare Hamas... a combattere Israele".

# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

## ESEMPIO 2: Video alterati

Un video mostra i giornalisti della CNN che scappano a piedi e cercano disperatamente un riparo mentre dei missili atterrano nelle vicinanze, vicino al confine tra Israele e Gaza. Il filmato è accompagnato dalla voce fuori campo di un uomo che sembrerebbe dare istruzioni a una troupe, a riprova del fatto che si tratta di una messinscena della CNN.

<https://rumble.com/v3odreg-cnn-busted-faking-attack-in-israel-for-the-camera-this-is-insane-and-appall.html>



Il filmato è stato alterato per aggiungere una voce maschile fuori campo che sembrerebbe dare istruzioni a una troupe di attori ("cerca di sembrare gentile e spaventato" e "puoi alzare il volume delle esplosioni, per favore?")

# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

Il video alterato e la narrazione falsa si sono diffusi su siti e su TikTok, Rumble, Facebook, Youtube e X.

Il 10 ottobre Sulaiman Ahmed ha postato su X la frase “LA CNN SMASCHERATA PER AVER SIMULATO UN ATTACCO IN ISRAELE”, insieme al video alterato. Il post è stato condiviso più di 17.000 volte in circa 48 ore.

Il canale Youtube Veteran Biker News & Views ha condiviso il video alterato descrivendolo così: “La CNN è stata sorpresa a simulare una notizia in Israele per le telecamere! Questo è folle e disgustoso”. Il video ha ricevuto 17.000 visualizzazioni in circa 48 ore.

Patriots.win ha pubblicato uno screenshot del video con la didascalia: "CNN SMASCHERATA PER AVER SIMULATO UN ATTACCO IN ISRAELE (ottobre 2023)".

# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

## ESEMPIO 3: Videogames

Un video mostra Hamas che abbatte degli elicotteri israeliani durante la guerra tra Israele e Hamas nell'ottobre 2023.

<https://www.instagram.com/p/CyGzhjgIHHu/>



Il video non mostra un'operazione di Hamas contro le forze aeree israeliane: è stato infatti creato attraverso il videogioco militare Arma 3.

La clip è stata pubblicata su YouTube il 3 ottobre, quattro giorni prima dell'attacco di Hamas contro Israele.

Bohemia Interactive, la società che ha creato Arma 3, ha confermato a Reuters che il video è tratto da Arma 3.



# IA usata per produrre contenuti di misinformazione sulla guerra tra Israele e Hamas

I video degli elicotteri accompagnati dalla didascalia scorretta sono apparsi per la prima volta il 7 ottobre, il giorno dell'attacco di Hamas. Ad esempio, un account X verificato e gestito da un utente anonimo che si presenta come "attivista musulmano" ha scritto quel giorno: "I combattenti per la libertà palestinesi hanno abbattuto 4 elicotteri da guerra israeliani a Gaza".

Questo video, condiviso principalmente da utenti francesi, indiani e pakistani su X, Facebook e Instagram, ha ricevuto più di 17.000 like e repost al 10 ottobre 2023. Si è diffuso anche su diversi siti di notizie greci, tra cui Topontiki.gr e Inewsgr.com.

I video in-game di Arma 3 sono stati anche citati da alcuni misinformatori come se ritraessero alcuni scontri avvenuti nell'ambito del conflitto tra Russia e Ucraina.

# Buone pratiche

## THE DEFINITIVE SOURCE Standards around generative AI

- AP has a [licensing agreement with OpenAI](#), the maker of ChatGPT, and while AP staff may experiment with ChatGPT with caution, they do not use it to create publishable content.
- Any output from a generative AI tool should be treated as unvetted source material. AP staff must apply their editorial judgment and AP's sourcing standards when considering any information for publication.
- In accordance with our standards, we do not alter any elements of our photos, video or audio. Therefore, we do not allow the use of generative AI to add or subtract any elements.
- We will refrain from transmitting any AI-generated images that are suspected or proven to be false depictions of reality. However, if an AI-generated illustration or work of art is the subject of a news story, it may be used as long as it clearly labeled as such in the caption.
- We urge staff to not put confidential or sensitive information into AI tools.
- We also encourage journalists to exercise due caution and diligence to ensure material coming into AP from other sources is also free of AI-generated content.
- Generative AI makes it even easier for people to intentionally spread mis- and disinformation through altered words, photos, video or audio, including content that may have no signs of alteration, appearing realistic and authentic. To avoid using such content inadvertently, journalists should exercise the same caution and skepticism they would normally, including trying to identify the source of the original content, doing a reverse image search to help verify an image's origin, and checking for reports with similar content from trusted media.

If journalists have any doubt at all about the authenticity of the material, they should not use it.

# Buone pratiche

## Come usiamo le intelligenze artificiali generative su Slow News?

👉 Nessun testo che pubblichiamo è **mai** generato direttamente da un'intelligenza artificiale.

Non usiamo **mai** le intelligenze artificiali generative per produrre testi giornalistici da zero.

Non pubblichiamo **mai** alcun testo generato da intelligenze artificiali (anche se editato) senza prima sottoporlo a verifica, secondo i più alti standard giornalistici.

🧠 Usiamo le intelligenze artificiali generative come **assistenti** nel processo di lavorazione (per esempio: riassunti, che comunque verifichiamo; mappe mentali; idee per titoli o per post sui social, brainstorming).

🎨 Usiamo le intelligenze artificiali generative per **produrre illustrazioni di alcuni dei nostri pezzi**, dichiarandolo esplicitamente. Lo facciamo, in particolare, per illustrare concetti astratti o se abbiamo idee di "format" da esplorare (come ad esempio il format replAIced di Alberto Puliafito, usato per illustrare [questo pezzo](#) sulle AI).

❌ Non usiamo **mai le intelligenze artificiali generative per produrre contenuti fotorealistici o videorealistici**, nemmeno per illustrare eventuali articoli in cui si parla di deepfake.

📖 Ci occupiamo di formazione dei giornalisti, in modo che si diffonda una cultura adeguata per l'uso delle intelligenze artificiali generative e della loro copertura giornalistica.

Sottoponiamo queste regole a **revisione e integrazione periodica**.

## **Virginia Padovese**

NewsGuard Managing Editor e VP Partnerships, Europa

[virginia.padovese@newsguardtech.com](mailto:virginia.padovese@newsguardtech.com)

The image features four silhouetted figures against a sunset sky. From left to right: a person holding a smartphone up to take a photo, a person in profile looking towards the center, a person facing away from the camera, and a person facing away from the camera. The NewsGuard logo, consisting of a shield icon and the text 'NewsGuard', is overlaid in white in the center of the image.

# NewsGuard