

Insieme nella sfida dell'Intelligenza artificiale

Intelligenza collettiva vs Intelligenza artificiale – un nuovo patto per lo spazio pubblico

Coordinatori: Mario Spada, Massimo Giammusso

Discussant: Stefania Piantoni

Wikipedia: l'intelligenza collettiva all'opera

*Ferdinando Traversa
Wikimedia Italia*



Biennale
Spazio
Pubblico
2025

insieme/together/juntos

Il modello di funzionamento di Wikipedia: i 5 pilastri

1. Wikipedia è un'encyclopedia
2. Wikipedia ha un punto di vista neutrale
3. Wikipedia è libera
4. Wikipedia ha un codice di condotta
5. Wikipedia non ha regole fisse

Le decisioni vengono prese dalla comunità per **consenso**

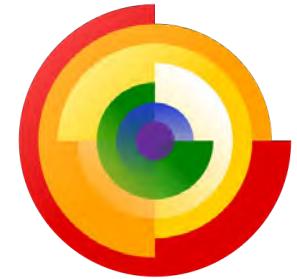
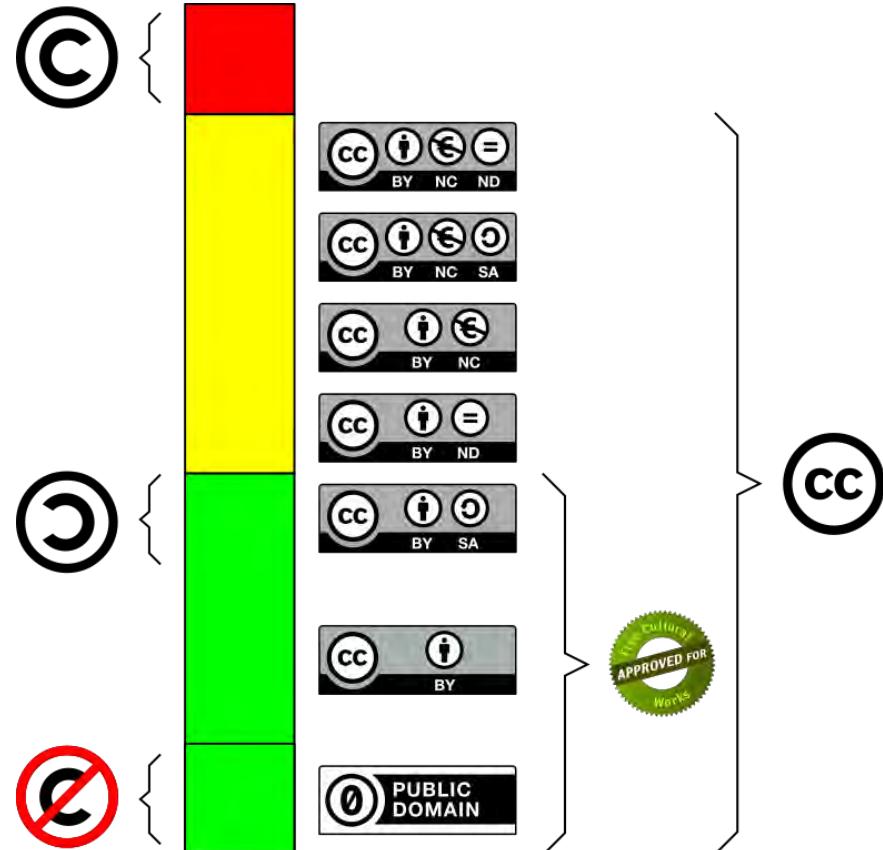


Wikipedia è un'enciclopedia

- **Fonte secondaria e terziaria**
- **Attenzione alle fonti attendibili**
- Principio di **verificabilità**
- No al **recentismo** e al **localismo**
- Concetto di **enciclopedicità**
- Autorialità, autorità VS. Autorevolezza delle fonti attendibili



Le licenze libere



Libertà essenziali

la libertà di usare e rappresentare l'opera
la libertà di studiare l'opera e impiegarne le informazioni
la libertà di redistribuire copie
la libertà di distribuire un'opera derivata

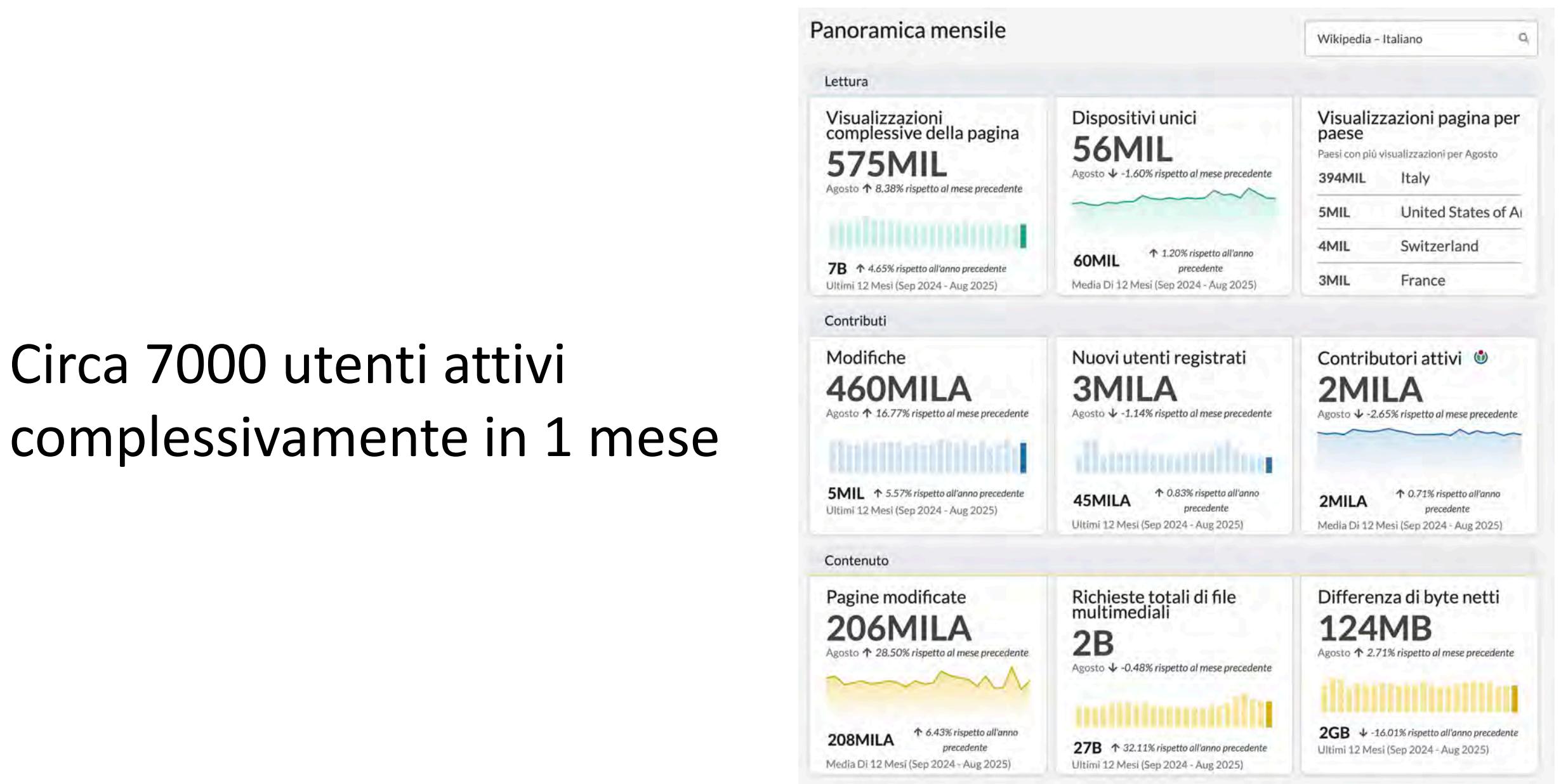
Inoltre:

Disponibilità dei dati di origine
Utilizzo di un formato libero
Nessuna restrizione tecnica
Nessun'altra restrizione o limitazione



Biennale
Spazio
Pubblico
2025

insieme/together/juntos



Circa 7000 utenti attivi complessivamente in 1 mese

Wikipedia è attendibile?

nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [news](#) > [article](#)

News | Published: 14 December 2005

Internet encyclopaedias go head to head

Jim Giles

[Nature](#) 438, 900–901 (2005) | [Cite this article](#)

92k Accesses | 1103 Citations | 982 Altmetric | [Metrics](#)

Jimmy Wales' Wikipedia comes close to Britannica in terms of the accuracy of its science entries, a *Nature* investigation finds. UPDATE: see details of how the data were collected for this article in the [supplementary information](#). UPDATE 2 (28 March 2006). The results reported in this news story and their interpretation have been disputed by [Encyclopaedia Britannica](#). *Nature* responded to these [objections](#).

Filtri attivi FILTRO NUOVO ORES CON COLORI

Probabilmente ha problemi X Molto probabilmente ha problemi X Molto probabilmente buona X Potrebbe avere problemi X Potrebbe essere in malafede X
Probabile in malafede X Molto probabilmente in malafede X Non registrato X Nuovi utenti X Principianti X Umani (non bot) X Non verificate X Versione attuale X
Tutti i contenuti X Tutte le discussioni X

Filtra le modifiche (usa il menu o cerca il nome del filtro)

Aggiornamenti in tempo reale

180 modifiche, 25 giorni

6 apr 2025

11:20 Clemente Grimaldi (diff 1 cron) .. (+108) .. 93.33.48.142 (discussione) (→Vita Privata) (rollback) (Etichette: Modifica visuale)
11:20 Alessandro Fiorenzi (diff 1 cron) .. (-131) .. 158.47.227.139 (discussione) (→Palmarés) (rollback) (Etichette: Modifica da mobile, Modifica da web per mobile)
11:19 La bomba (film 1999) (diff 1 cron) .. (+11) .. 2001:B07:6468:D463:8C72:2B5B:DAB2:10E7 (discussione) (ha scritto meglio) (rollback) (Etichette: Modifica visuale, Modifica da mobile, Modifica da web per mobile)
11:15 Strade provinciali della provincia di Foggia (diff 1 cron) .. (-1) .. 93.67.196.64 (discussione) (→SP 51 - SP 60) (rollback) (Etichette: Modifica visuale)
11:15 Gambia (diff 1 cron) .. (-412) .. 2A02:B027:6012:8F43:A465:7389:A7D3:A797 (discussione) (rollback) (Etichette: Modifica visuale, Modifica da mobile, Modifica da web per mobile)
11:14 Abbazia di Chiaravalle (diff 1 cron) .. (-3) .. 2.229.210.187 (discussione) (→Il tiburio) (rollback) (Etichette: Modifica visuale)
11:13 E tu.. (diff 1 cron) .. (+16) .. 79.23.212.63 (discussione) (rollback) (Etichette: Modifica da mobile, Modifica da web per mobile)
11:12 Party Rock Anthem (diff 1 cron) .. (-57) .. 5.91.20.228 (discussione) (rollback) (Etichette: Modifica da mobile, Modifica da web per mobile)
11:11 Notte di Valpurga (diff 1 cron) .. (+22) .. 37.161.57.36 (discussione) (una breve annuncio all'inizio) (rollback) (Etichette: Modifica visuale)
11:10 Ferrari 250 LM (diff 1 cron) .. (-2) .. 94.35.113.147 (discussione) (ho cambiato "ritenuta" con "ritiene") (rollback) (Etichette: Modifica visuale)
11:10 Rivolta di Chilembwe (diff 1 cron) .. (0) .. 79.16.108.234 (discussione) (La traduzione da "domande" a "pretese") (rollback) (Etichette: Modifica visuale)
11:09 Locomotiva FS D.445 (diff 1 cron) .. (+1) .. 2A02:B125:8011:21D1:DCA8:6BFF:FED1:9E59 (discussione) (rollback) (Etichette: Modifica visuale, Modifica da mobile, Modifica da web per mobile)
11:08 Igor Protti (diff 1 cron) .. (+126) .. 2001:B07:6468:1E3D:D447:4E40:4E63:5D56 (discussione) (→Giocatore) (rollback)
11:06 Schiavonee (diff 1 cron) .. (+67) .. 194.230.145.217 (discussione) (Aggiunto informazioni su capoluoghi vicini) (rollback) (Etichette: Modifica visuale, Modifica da mobile, Modifica da web per mobile)
11:06 San Marino Song Contest 2025 (diff 1 cron) .. (-23) .. 79.22.209.72 (discussione) (rollback) (Etichette: Modifica da mobile, Modifica da web per mobile)
11:02 Eurovision Song Contest 2025 (diff 1 cron) .. (-129) .. 79.22.209.72 (discussione) (→Stai partecipanti: Non è il primo di della storia a partecipare al concorso semplicemente (a differenza di altri) i cantanti non saranno annunciati (vedasi come esempio i Secret Garden) (rollback) (Etichette: Ripristino manuale, Modifica da mobile, Modifica da web per mobile)
11:01 Legami (diff 1 cron) .. (+39) .. 80.117.87.123 (discussione) (rollback) (Etichette: Sequenze di caratteri ripetuti da parte di un nuovo utente o IP, Modifica visuale)
11:00 Halloween (diff 1 cron) .. (+18) .. 87.16.56.214 (discussione) (rollback)
10:58 Jurassic World - Teoria del caos (diff 1 cron) .. (-5) .. 2.36.105.51 (discussione) (→Principali) (rollback) (Etichette: Modifica visuale)
10:57 Ci hai rotto papà (diff 1 cron) .. (+55) .. 2A02:B127:8F04:51DF:8C69:1AF0:C99B:E595 (discussione) (rollback) (Etichette: Modifica visuale, Edit Check (reference))

Elenco di abbreviazioni:
D Modifica effettuata su Wikidata
r Questa modifica potrebbe avere problemi e dovrebbe essere rivista (ulteriori informazioni)
N Creazione di una nuova pagina (vedi le [nuove pagine](#))
m Modifica minore
b Modifica effettuata da un bot
i Modifica non ancora verificata
#123 Variazione della pagina in numero di byte
P Pagina temporaneamente osservata



Biennale
Spazio
Pubblico
2025

insieme/together/iuntos

L'intelligenza artificiale su Wikipedia

[Wikipedia:Intelligenza artificiale](#)

15 lingue ▾

Pagina di servizio [Discussione](#) [Leggi](#) [Modifica wikitesto](#) [Cronologia](#) [☆](#)

Wikidata: [Wikipedia:Intelligenza Artificiale \(Q120644401\)](#), Nessuna descrizione fornita

Alias: Nessuno

 **Il succo di questa pagina:** Evita di usare strumenti che facciano uso di intelligenza artificiale (IA) per generare nuovo contenuto originale, per generare note, o per scrivere da zero commenti. Fonti che facciano uso di testi prodotti da IA non sono attendibili. I risultati ottenuti da IA non sono utilizzabili come prova a sostegno di un'argomentazione in una discussione. Nei casi in cui l'uso di IA è ammesso, controlla sempre il risultato prima di pubblicare.

- **Le allucinazioni**
- **L'AI si nutre di Wikipedia**
- **Il model collapse**
- **L'umano al centro**

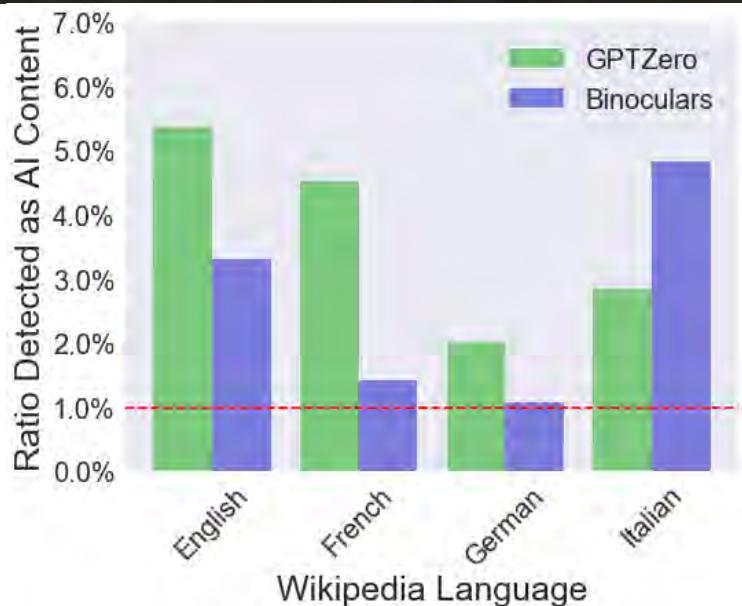
The Rise of AI-Generated Content in Wikipedia

Creston Brooks Samuel Eggert Denis Peskoff

Princeton University cabrooks, sameggert, dp28961@princeton.edu

Abstract

The rise of AI-generated content in popular information sources raises significant concerns about accountability, accuracy, and bias amplification. Beyond directly impacting consumers, the widespread presence of this content poses questions for the long-term viability of training language models on vast internet sweeps. We use GPTZero, a proprietary AI detector, and Binoculars, an open-source alternative, to establish lower bounds on the presence of AI-generated content in recently created Wikipedia pages. Both detectors reveal a marked increase in AI-generated content in recent pages compared to those from before the release of GPT-3.5. With thresholds calibrated to achieve a 1% false positive rate on pre-GPT-3.5 articles, detectors flag over 5% of newly created English Wikipedia articles as AI-generated, with lower percentages for German, French, and Italian articles. Flagged Wikipedia articles are typically of lower quality and are often self-promotional or partial towards a specific viewpoint on controversial topics.



G15. LLM-generated pages without human review

This applies to any page that exhibits one or more of the following signs which indicate that the page could only plausibly have been generated by [large language models \(LLM\)](#)^[15] and would have been removed by any [reasonable human review](#)^[16]

- **Communication intended for the user:** This may include collaborative communication (e.g., "Here is your Wikipedia article on..."), knowledge-cutoff disclaimers (e.g., "Up to my last training update ..."), self-insertion (e.g., "as a large language model"), and unfilled [phrasal templates](#) (e.g., "Smith was born on [Birth Date]" rather than the person's actual birth date).
- **Implausible non-existent references:** This may include external links that are [dead on arrival](#), [ISBNs](#) with invalid [checksums](#), and unresolved [DOIs](#). Since humans can make typos and links may suffer from [link rot](#), a single example should not be considered definitive. Editors should use additional methods to verify whether a reference truly does not exist.
- **Nonsensical citations:** This may include citations of incorrect temporality (e.g., a source from 2020 being cited for a 2022 event), DOIs that resolve to completely unrelated content (e.g., a paper on a beetle species being cited for a computer science article), and citations that attribute the wrong author or publication.

Shortcut
WP:G15

Wikipedia:Signs of AI writing

Pagina di servizio [Discussione](#)

[Leggi](#) [Modifica sorgente](#) [Cronologia](#) [★](#)

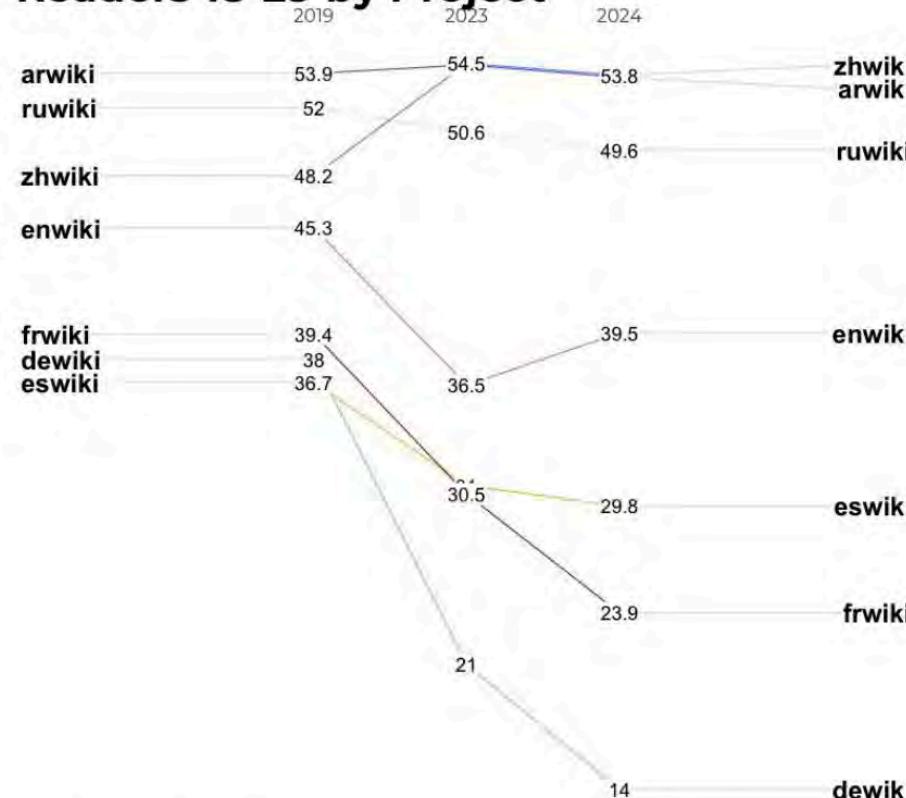
Da Wikipedia.



Shortcuts
WP: AISIGNS
WP: AITELLS
WP: LLMSIGNS

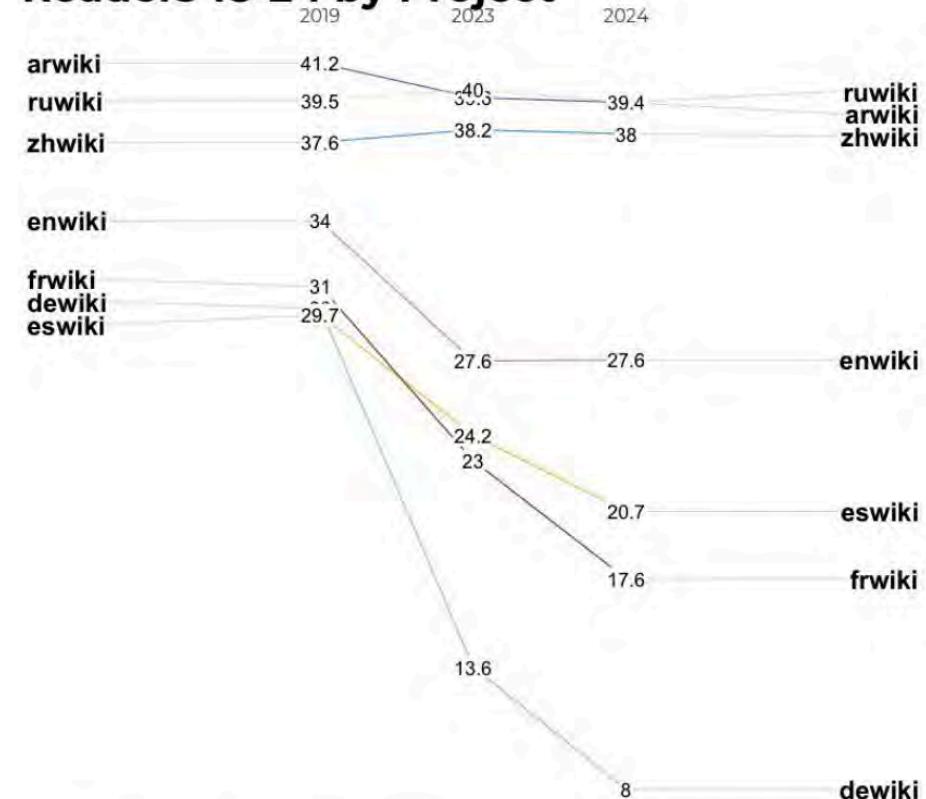
Wikipedia e l'AI: i lettori

Readers 18-29 by Project

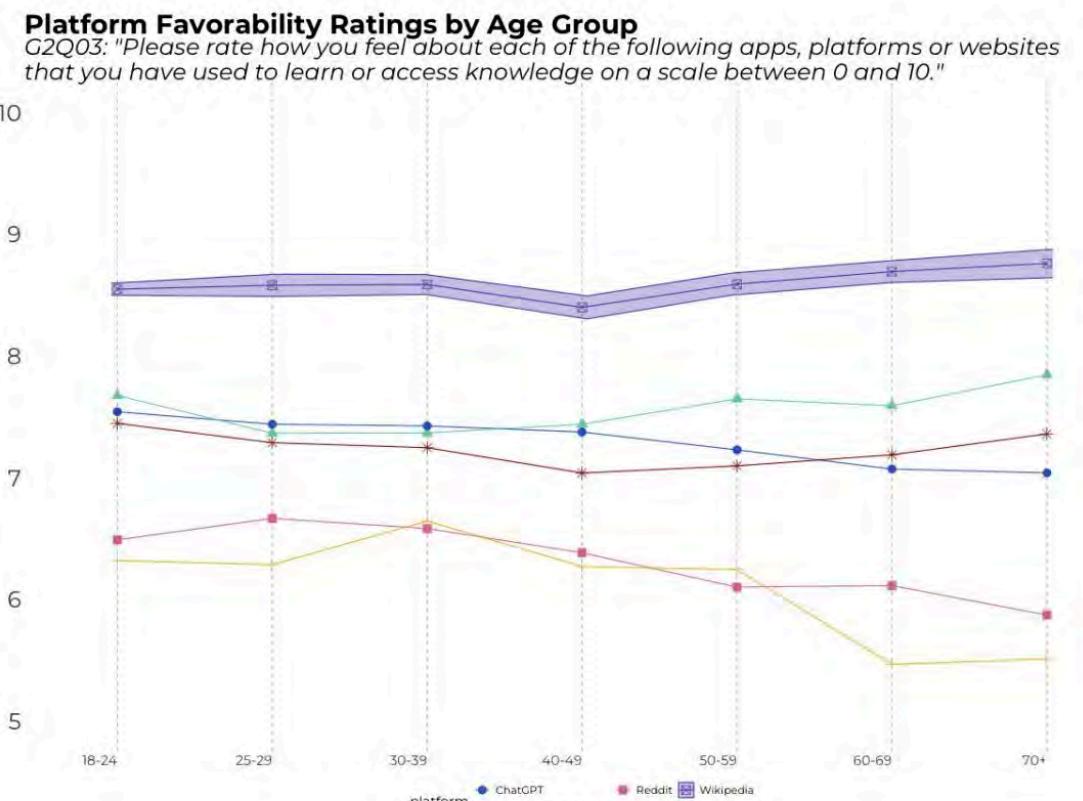
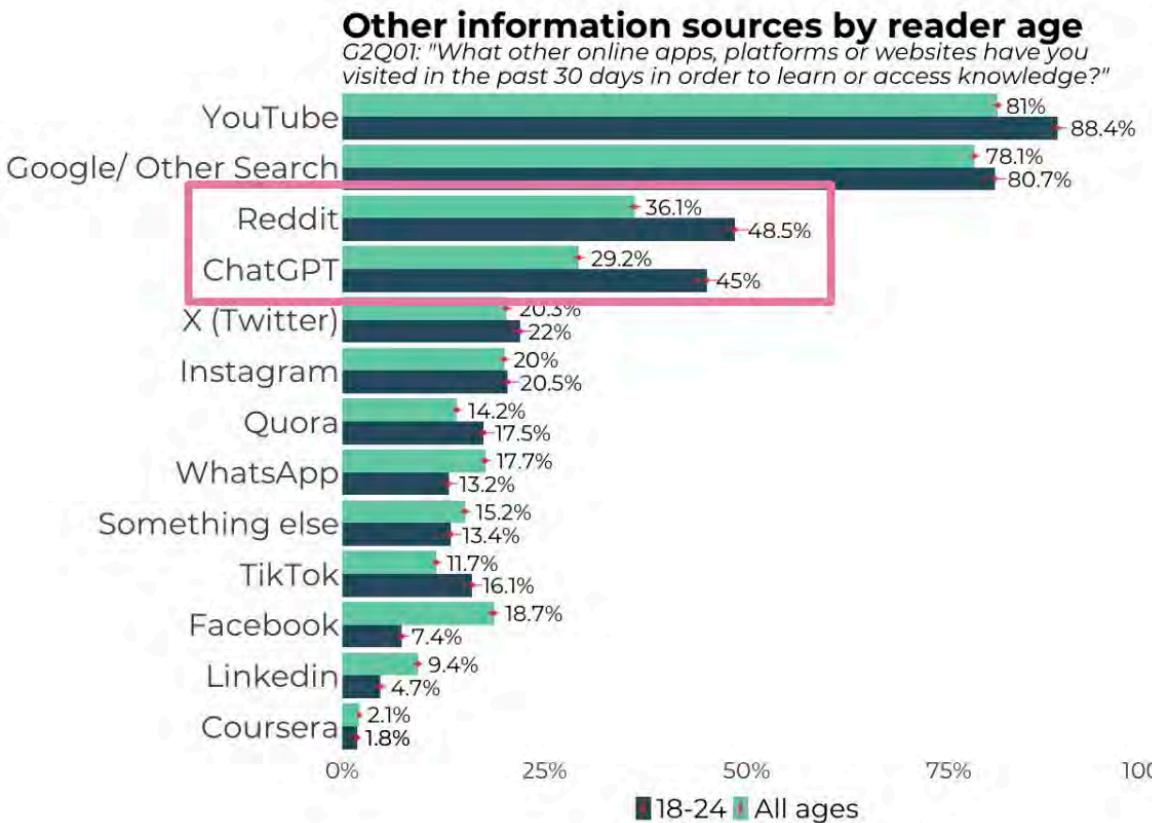


2023 Global Readers Survey data weighted by OS family, Session length, Referrer class, and Geography
2024 Global Readers Survey data weighted by Geography

Readers 18-24 by Project

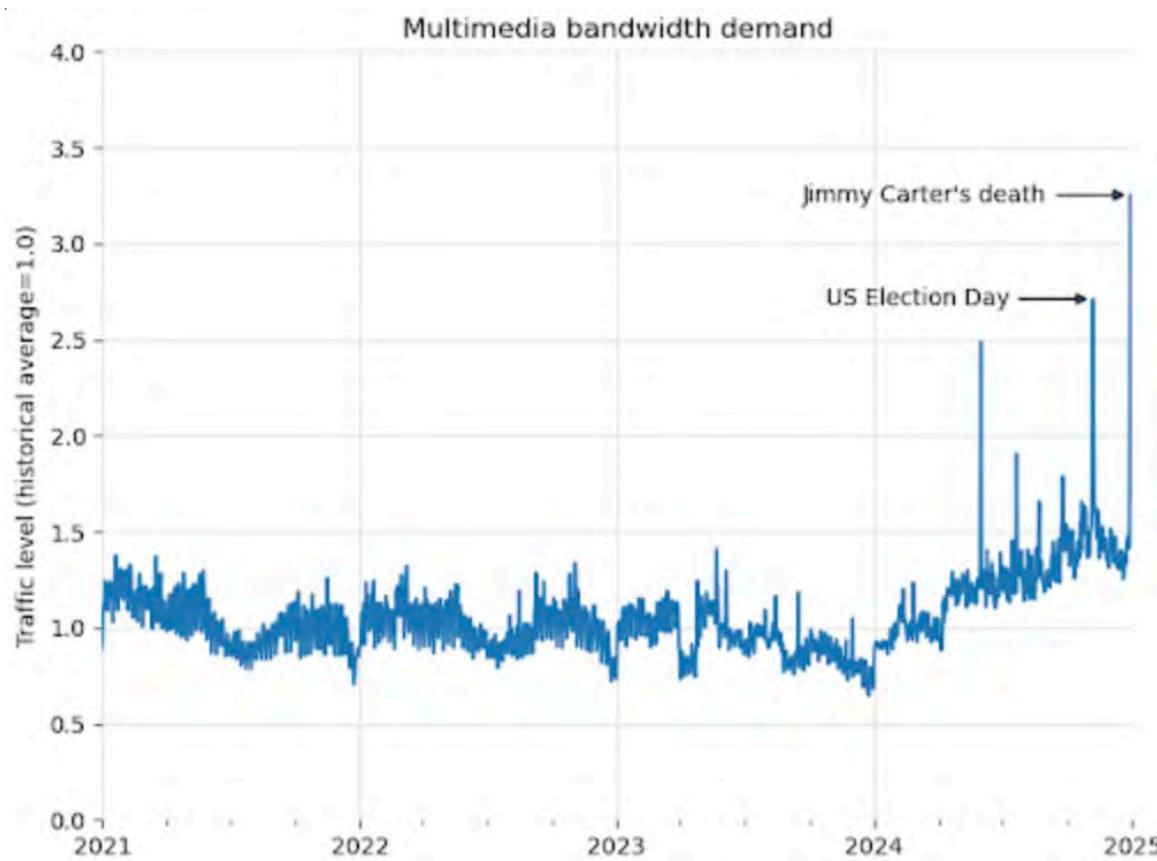


2023 Global Readers Survey data weighted by OS family, Session length, Referrer class, and Geography
2024 Global Readers Survey data weighted by Geography



2024 Global Readers Survey
Responses weighted by Geography and Project Share of Total Traffic

Our content is free, our infrastructure is not



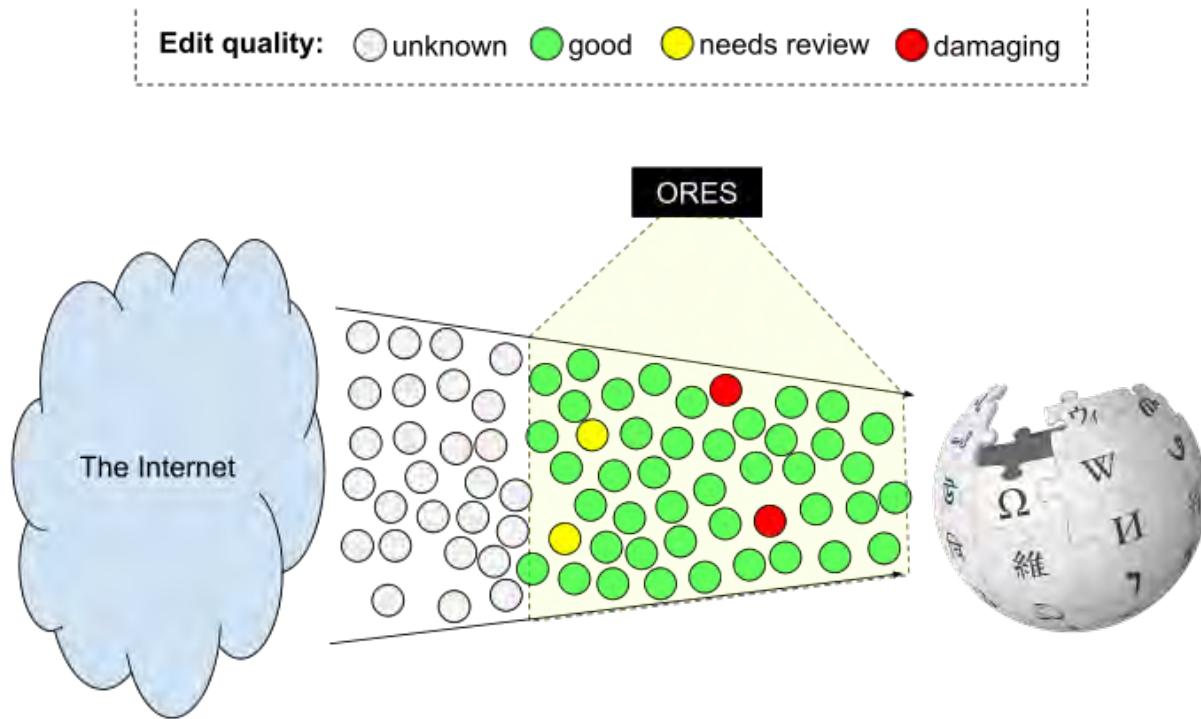
65% del nostro traffico più costoso
viene dai bot (che costituiscono
solo il 35% del traffico)

Wikipedia e il rischio disinformazione



Mechanism	Description	Type	Wikipedia's Vulnerability
Bots	Software used to automatize the spread of messages, generating the idea that of a lot people is given an specific opinion or interest about a topic	Technical	Low
Sock-puppets	Multiple Online identities used for purposes of deception.	Social	Medium
Web Brigades	A set of users coordinated to introduce fake content by exploiting the weakness of communities and systems.	Social	High
Click farms	Where a large group of low-paid workers are hired to perform some micro-tasks to deceive online systems.	Social	Medium
Deepfake	AI a technique for human image synthesis that can be used to create fake videos of celebrities or notable people.	Technical	Medium
Data Voids	Exploiting missing data to manipulate search results	Social	Medium
Circular reporting	A situation where a piece of information appears to come from multiple independent sources, but in reality comes from only one source.	Social	High

Wikipedia e l'AI: il contrasto al «vandalismo»



Quality	Importance						Total
	Top	High	Mid	Low	???	Total	
★ FA	1,562	2,464	2,382	1,896	180	8,484	
★ FL	179	667	762	700	106	2,414	
● A	360	680	786	579	83	2,488	
⊕ GA	3,180	7,217	14,562	19,093	1,734	45,786	
B	16,646	32,340	53,380	66,330	20,997	189,693	
C	16,621	53,351	133,083	299,487	85,459	588,001	
Start	18,449	91,958	412,517	1,586,521	398,413	2,507,858	
Stub	4,212	31,620	278,052	2,780,170	760,723	3,854,777	
List	4,784	16,908	53,189	189,210	66,933	331,024	
Assessed	65,993	237,205	948,713	4,943,986	1,334,628	7,530,525	
Unassessed	125	499	1,265	18,974	430,309	451,172	
Total	66,118	237,704	949,978	4,962,960	1,764,937	7,981,697	

Reverts by Automoderator

Across all Wikimedia projects where Automoderator is enabled,

Total reverts

18.8k

reverts

Time to revert (average)

5.94

seconds

User talk page notifications

696

messages

Reverted Automoderator reverts

Note: If an Automoderator revert was reverted back, it is considered as a potential false positive if it was reverted by different user from whose edit was initially reverted, and not by an anonymous or temporary user. This metric is not perfect, and should only be used as an estimate.

Total reverts
reverted

2.07k

reverts

Average time for
Automoderator t...

1,591

minutes

Potential false
positive reverts

702

reverts

Average time for
potential false...

68.33

hours

Reverted reverts: User types

User type	All reverted reverts	Potential false positives
anonymous	1.64k	328
other_registered	297	297
newcomer	117	73
temporary	16	4
Totals	2.07k	702

Potential false positive rate

3.6%

of all reverts

▶ Aggiornamenti in tempo reale

🕒 Visualizza le modifiche dal 12:08 6...

Evidenzia risultati

Filtri

Previsioni sulla qualità del contributo

[Come funzionano?](#)

Molto probabilmente buona

Molto preciso nel trovare quasi tutte le modifiche senza problemi.

Potrebbe avere problemi

Trova la maggior parte delle modifiche dannose, ma con minor precisione.

Probabilmente ha problemi

Con una precisione media, trova una frazione intermedia di modifiche problematiche.

Molto probabilmente ha problemi

Molto accurato nel trovare le modifiche dannose più evidenti.

Previsioni delle intenzioni dell'utente

[Come funzionano?](#)

Molto probabilmente in buona fede

Molto preciso nel trovare quasi tutte le modifiche in buona fede.

Potrebbe essere in malafede

Trova la maggior parte delle modifiche in malafede, ma con una precisione inferiore.

Probabile in malafede

Con una precisione media, trova una frazione intermedia di modifiche in malafede.

Molto probabilmente in mala fede

Molto accurato nel trovare le modifiche in malafede più evidenti.



Biennale
Spazio
Pubblico
2025

insieme/together/iuntos

RICHARD COOKE

BUSINESS FEB 17, 2020 6:00 AM

Wikipedia Is the Last Best Place on the Internet

Wikipedia...

People used to think the crowdsourced encyclopedia represented all that was wrong with the web. Now it's a beacon of so much that's right.

... non ha algoritmi di raccomandazione che creano *filter bubble* e mirano alla massima monetizzazione e user retention

... non profila gli utenti perché non ha pubblicità personalizzata e **tutela la loro privacy**

... si basa ancora sulla **collaborazione spontanea tra persone** (*l'intelligenza collettiva*)