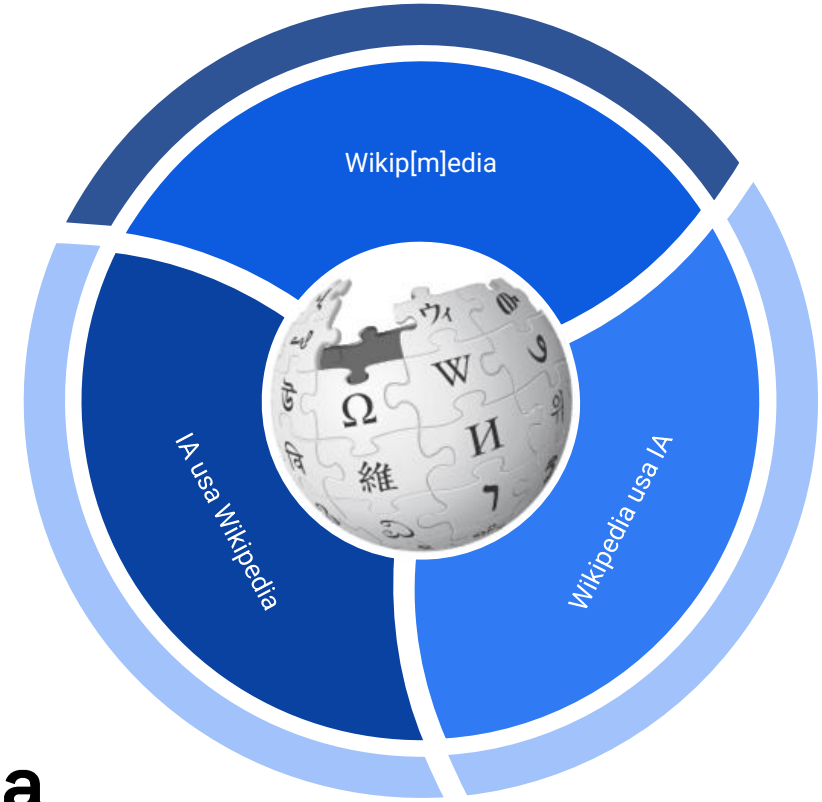


Il dataset più prezioso per l'IA è curato da volontari: come preservare la conoscenza aperta di Wikipedia?

Francisco Ardini, Marco Montanari

WIKIMEDIA ITALIA

GenOA week 2026, 03/11/2026



Wikip[m]edia

MOVIMENTO WIKIMEDIA

WIKIMEDIA FOUNDATION

Organizzazione senza fini di lucro che gestisce Wikipedia e i progetti Wikimedia, occupandosi tra le altre cose di server, marchi e fundraising

*«Solo la **conoscenza comune** salvaguarda le pari opportunità di tutti gli uomini riguardo all'accesso all'informazione, all'educazione e alla conoscenza.»*[1]

WIKIMEDIA ITALIA

Capitolo della Wikimedia Foundation. Promuove e sostiene Wikipedia, i progetti Wikimedia, OpenStreetMap e la conoscenza libera in Italia, Repubblica di San Marino e Città del Vaticano

*«[...] il miglior modo per produrre e disseminare la conoscenza è fare in modo che la maggioranza delle persone possa **liberamente prendervi parte**.»*[2]

1. https://meta.wikimedia.org/wiki/Free_knowledge_based_on_Creative_Commons_licenses/it
2. https://www.wikimedia.it/chi-siamo/

CC BY SA

PROGETTI WIKIMEDIA

Progetti liberi e collaborativi



CC BY SA

COM'È CRESCIUTA LA REPUTAZIONE DI WIKIPEDIA

2001: creation of Wikipedia

2005: «Internet encyclopaedias go **head to head**» (Nature)[1]

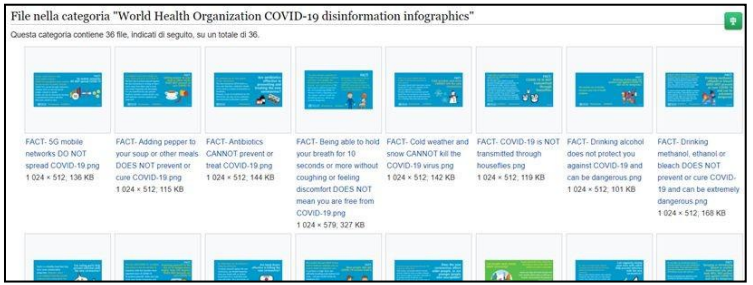
2014: «British people **trust** Wikipedia more than the news» (YouGov)[2]

2017: «**Science Is Shaped by** Wikipedia: Evidence From a Randomized Control Trial» (MIT Sloan)[3]

2017: «What can **fact-checkers** learn from Wikipedia? We asked the boss of its nonprofit owner» (Pointer)[4]

2018: «Conspiracy videos? Fake news? Enter Wikipedia, the '**good cop**' of the Internet» (Washington Post)[5]

2020: «The **World Health Organization** and Wikimedia Foundation expand access to trusted information about COVID-19 on Wikipedia» (OMS)[6]



1. https://www.nature.com/articles/438900a
2. https://yougov.co.uk/topics/politics/articles-reports/2014/08/09/more-british-people-trust-wikipedia-trust-news
3. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039505
4. https://www.poynter.org/fact-checking/2017/what-can-fact-checkers-learn-from-wikipedia-we-asked-the-boss-of-its-nonprofit-owner/
5. https://www.washingtonpost.com/outlook/conspiracy-videos-fake-news-enter-wikipedia-the-good-cop-of-the-internet/2018/04/06/ad1f018a-3835-11e8-8fd2-49fe3c675a89_story.html
6. https://www.who.int/news/item/22-10-2020-the-world-health-organization-and-wikimedia-foundation-expand-access-to-trusted-information-about-covid-19-on-wikipedia

CC BY SA

COM'È OGGI WIKIPEDIA

- 65 milioni di voci
- +340 edizioni linguistiche
- 18 modifiche ogni secondo
- 10.000 pagine visitate ogni secondo

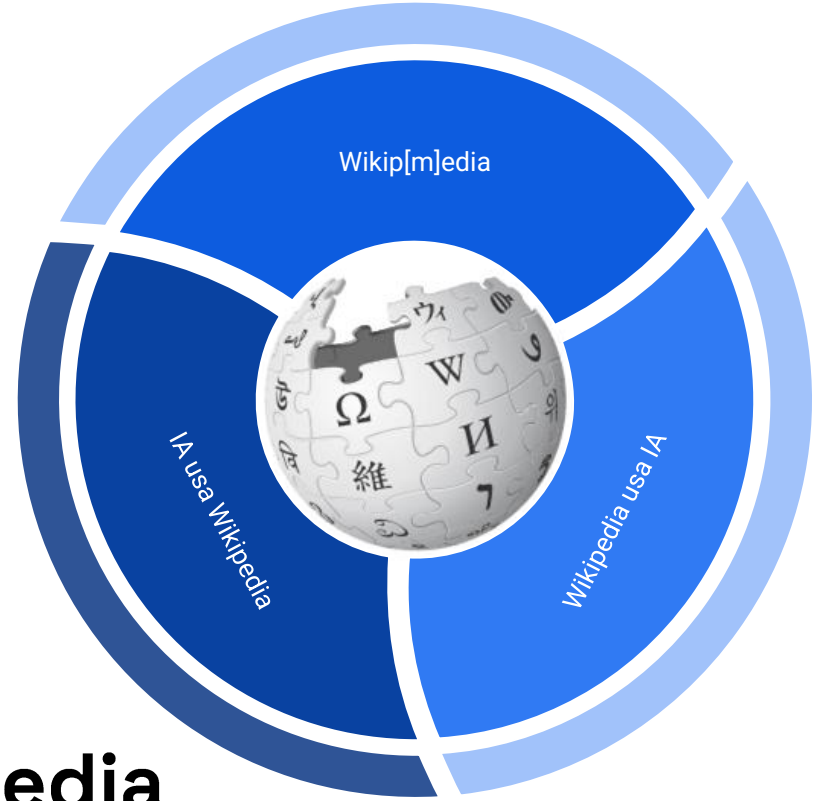
Posizione	Sito web
1	google.com
2	youtube.com
3	facebook.com
4	instagram.com
5	chatgpt.com
6	reddit.com
7	wikipedia.org

Posizione	Sito web
1	google.com
2	youtube.com
3	facebook.com
4	amazon.it
5	wikipedia.org

https://it.semrush.com/



CC BY SA



IA usa Wikipedia

I LLM SI AFFIDANO A WIKIPEDIA

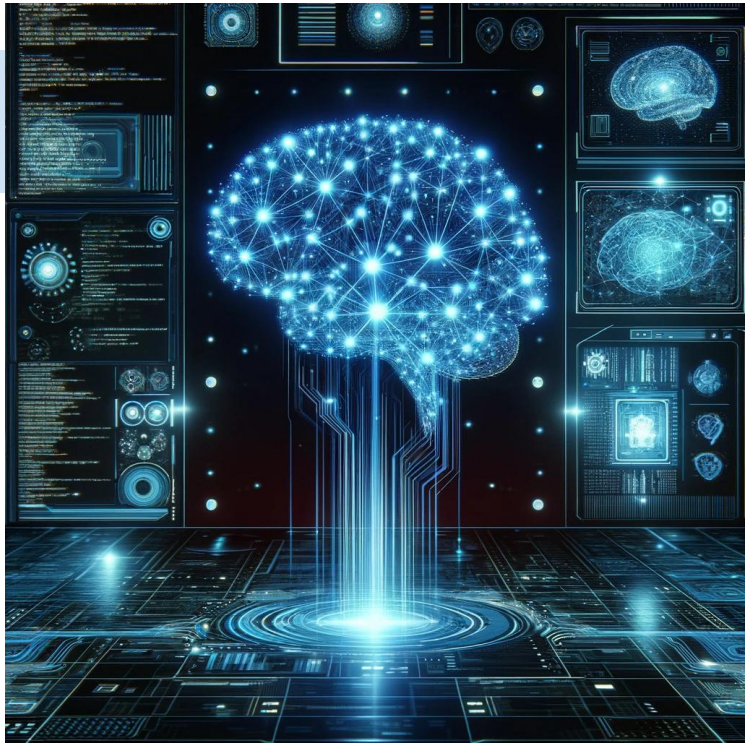
L'addestramento dell'IA si affida in gran parte a Wikipedia per contenuti di alta qualità, accurati, variegati, verificabili.

"**Wikipedia is a standard source of high-quality text** for language modeling. In addition to being a source of high quality, clean English text, it is also valuable as it is written in expository prose, and spans many domains."

The Pile: An 800GB Dataset of Diverse Text for Language Modeling, arXiv:2101.00027

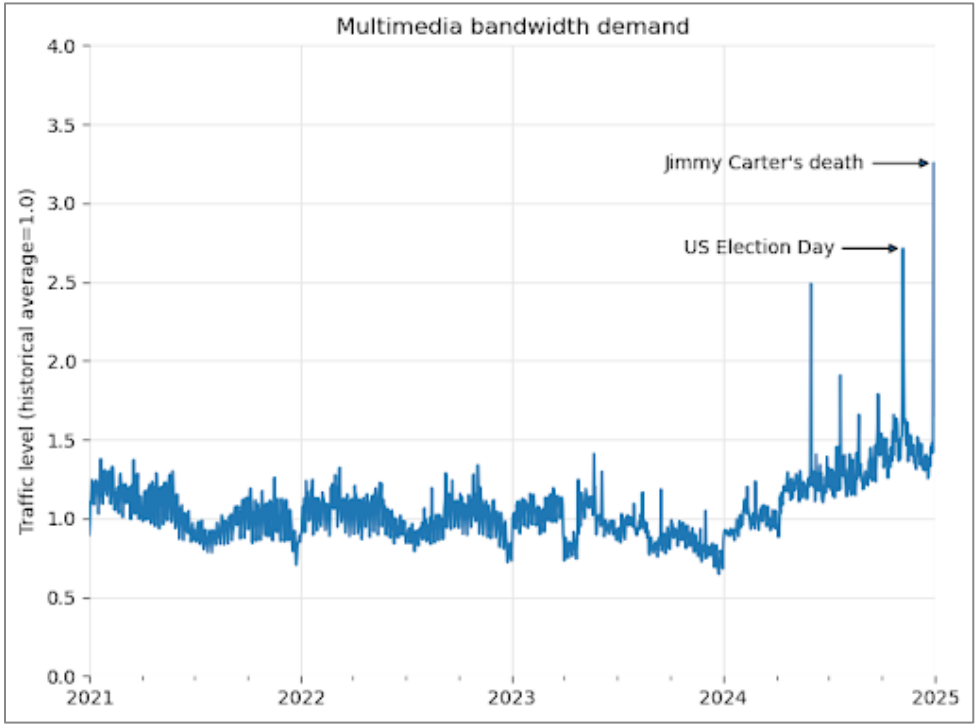
"**Wikipedia presents the highest quality text**, before Books and OpenWeb.
[...] Removing Books, CommonCrawl and OpenWeb lead to the greatest decrease in toxicity metrics. **Removing Wikipedia had a strong increase in toxicity generation**.

A Pretrainer's Guide to Training Data, arXiv:2305.13169v2



CC BY SA

I LLM POSSONO RALLENTARE WIKIPEDIA



https://diff.wikimedia.org/2025/04/01/how-crawlers-impact-the-operations-of-the-wikimedia-projects/

Aumenta il traffico di crawler e sistemi automatici

Più consumo di banda → peggiori prestazioni e costi maggiori

"Our content is free, our infrastructure is not"

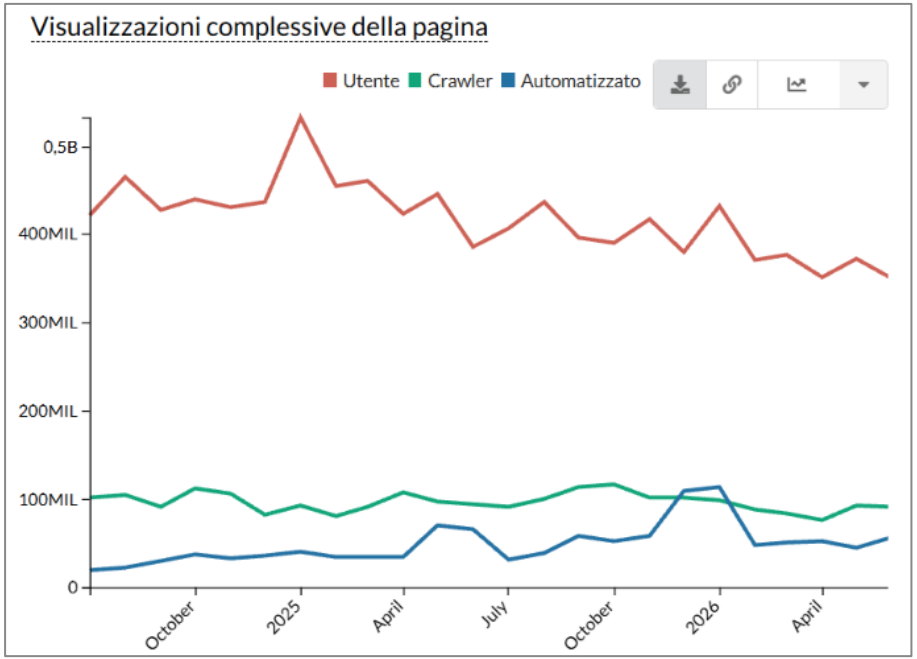
Come risolvere?

- Accesso prioritario per lettori
- Bloccare traffico abusivo
- Canali differenziati (Wikipedia Enterprise)

https://diff.wikimedia.org/2026/03/26/quo-vadis-crawlers-progress-and-whats-next-on-safeguarding-our-infrastructure/

CC BY SA

I LLM CAUSANO UN CALO DELLE VISITE SU WIKIPEDIA



https://stats.wikimedia.org/#/it.wikipedia.org/reading/total-page-views/normal|line|2024-06-21~2026-06-18|agent~user*spider*automated|monthly

1) 2024 → 2025
   Calo delle visite dell'**8%**

2) Da giugno 2026
   Traffico umano sotto il **52%** del totale

Meno persone visitano il sito, meno persone proveranno a contribuire

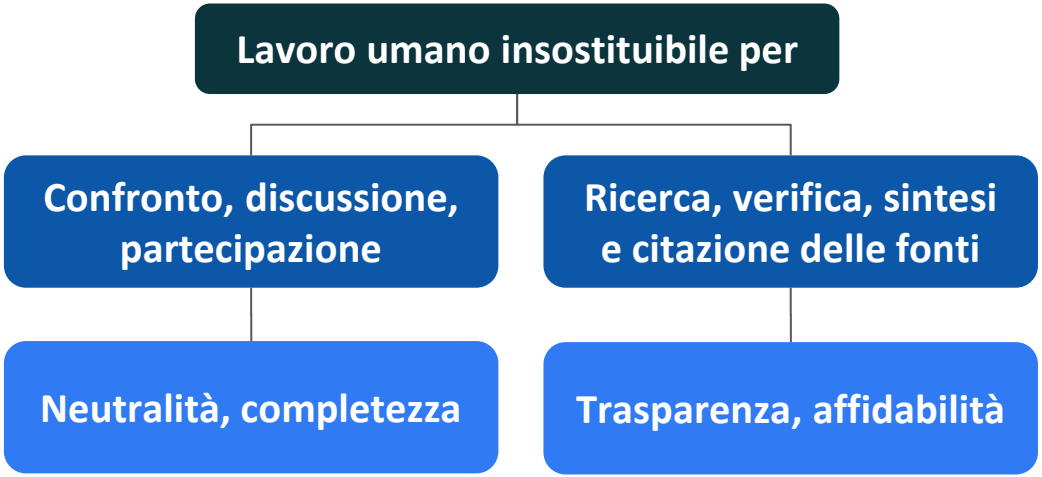
3) Gennaio 2024 → Luglio 2026
   Calo dei contributori occasionali dell'**11%**

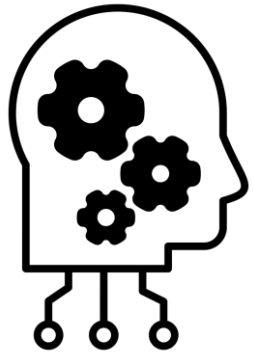
1) https://diff.wikimedia.org/2025/10/17/new-user-trends-on-wikipedia/
2) https://stats.wikimedia.org

CC BY SA

CAMBIA LA GENERAZIONE DELLA CONOSCENZA?

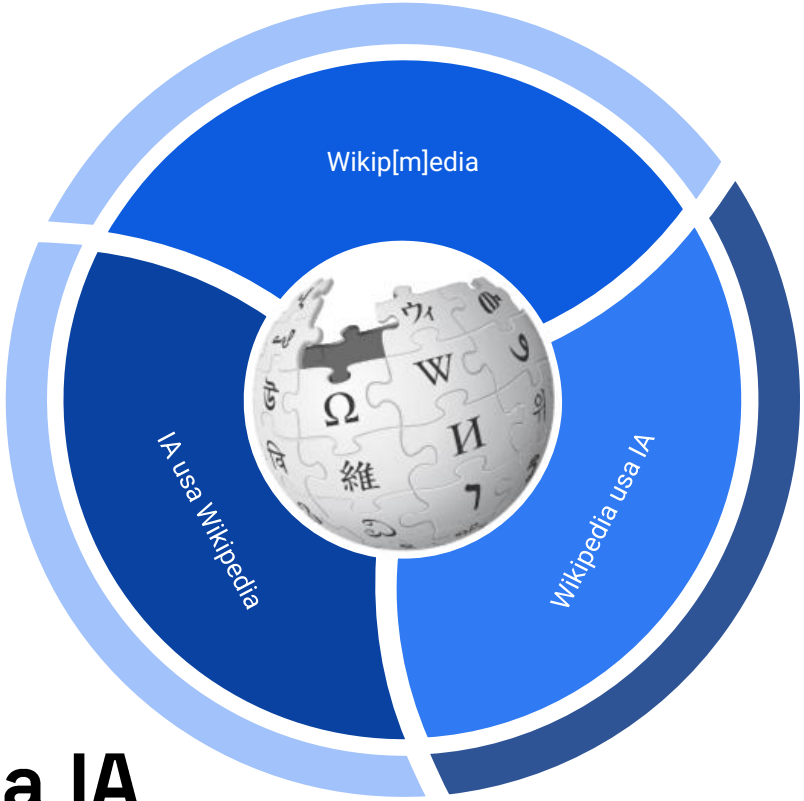
La conoscenza non è solo il risultato: è anche il **processo** che lo rende **affidabile**.





L'analisi e la critica non possono sottrarsi al contributo umano. Una IA dà una sola risposta, su Wikipedia la conoscenza si negozia con confronto e fonti.

CC BY SA



Wikipedia usa IA

LA SCELTA WMF: IA COME SUPPORTO

Aprile 2025: la Wikimedia Foundation decide di usare l'IA generativa. Non per scrivere Wikipedia, ma per **alleggerire il lavoro di chi la fa.**

GESTORI
moderatori, patroller, amministratori

Flussi di lavoro assistiti: meno compiti ripetitivi nel controllo delle modifiche

EDITOR
chi scrive, esperti e nuovi arrivati

Informazioni più facili da trovare, traduzione assistita, mentoring per chi inizia

SVILUPPATORI
chi costruisce strumenti e infrastruttura

Modelli aperti ospitati da Wikimedia, esposti via API a bot e tool

I principi dichiarati

Persone al centro | **Modelli aperti** | **Trasparenza** | **Multilinguismo**

Albon, Zia (30/04/2025), Our new AI strategy puts Wikipedia's humans first | Strategy/Multigenerational/Artificial intelligence for editors

CC BY SA

GESTORI: MENO FATICA, STESSO CONTROLLO

Moderatori, patroller, admin

Dalla strategia WMF

- *Flussi di lavoro assistiti dall'IA per moderatori e patroller*
- *Automatizzare i compiti noiosi, a tutela dell'integrità della conoscenza*

Il modello segnala, la comunità decide soglie e regole.

Esempi concreti

Revert Risk (Lift Wing)

Stima la probabilità che una modifica venga annullata. Erede di ORES (2015), funziona in decine di lingue.

Automoderator

Annulla le modifiche ad altissimo rischio. Ogni comunità sceglie se attivarlo e con quale soglia.

ClueBot NG

Bot antivandalismo a rete neurale, sviluppato da volontari e attivo dal 2010.

Strategia IA WMF | Automoderator | Usi dell'IA su Meta

CC BY SA

EDITOR: PIÙ TEMPO PER IL GIUDIZIO

Chi scrive, esperti e nuovi arrivati

Dalla strategia WMF

- *Rendere le informazioni più facili da trovare, per lasciare tempo a discussione e consenso*
- *Tradurre e adattare i temi comuni, per dare spazio alle prospettive locali*
- *Accogliere i nuovi volontari con un mentoring guidato*

Meno tempo sul “come”, più tempo sul “cosa”.

Esempi concreti

Content Translation + MinT

Traduzione automatica con modelli aperti. Il testo va sempre rivisto da chi traduce.

Modifiche suggerite (Add a link)

Un modello propone collegamenti tra voci: il nuovo utente accetta o scarta.

Edit Check

Suggerimenti mentre si scrive: fonte mancante, tono poco neutrale, testo incollato.

Strategia IA WMF | Usi dell’IA su Meta

CC BY SA

SVILUPPATORI: UN'INFRASTRUTTURA APERTA

Chi costruisce tool e infrastruttura

Dalla strategia WMF

- *Usare l'IA per abbattere le barriere tecniche*
- *Preferire modelli open source o open weight*
- *Essere trasparenti su modelli e dati*

I modelli sono un bene comune, come i contenuti.

Esempi concreti

Lift Wing
La piattaforma WMF che ospita i modelli di machine learning e li espone via API a bot e tool.

Model card pubbliche
Ogni modello è documentato su Meta: dati di addestramento, limiti, prestazioni.

Wikidata Embedding Project
Ricerca semantica su Wikidata, pensata per chi sviluppa applicazioni di IA (2025).

Strategia IA WMF | Machine learning models (Meta) | Lift Wing

CC BY SA

IL LIMITE LO DECIDE LA COMUNITÀ

GIUGNO 2025

Simple Article Summaries

La WMF prova riassunti generati dall'IA in cima alle voci.

Dopo le critiche degli editor, il test viene sospeso in pochi giorni.

Cosa ci insegna

1. Supporto dietro le quinte: sì.
   Testo generato davanti ai lettori: no, senza consenso.
2. Ogni progetto linguistico scrive le proprie regole sull'uso dell'IA.
3. La Foundation propone, le comunità decidono.

Wikimedia pauses AI summaries trial after backlash (Mediaweek) | Policy IA per progetto (Meta) | Panoramica strumenti

CC BY SA

IN CONCLUSIONE

1. **L’IA assiste, non sostituisce.** La responsabilità resta delle persone.
2. **Le fonti restano verificabili,** perché tra persone possiamo discutere, correggere e motivare le scelte.
3. **Il giudizio critico non si delega.** L’IA libera tempo per esercitarlo.

> “
> ***Wikipedia is the last best place on the internet***
> ”
>
> *Wired*, 2020

Come facciamo a far dialogare le due realtà?

CC BY SA

GRAZIE PER L'ATTENZIONE

Domande?



WIKIMEDIA
ITALIA

Contatti

https://www.wikimedia.it/
info@wikimedia.it
segreteria@wikimedia.it

CC BY SA