

Communicating Science Online with Wikipedia in the LLM Era

Mike Peel









Postdoc, Imperial College London
Volunteer for Wikimedia

27 March 2026

Wikipedia and Wikimedia

- Wikipedia's been around since 2001: anyone under 25 has never known a life without it!
- Also: Wikidata, Wikimedia Commons, ..., Wikimedia organisations
- Donor-supported: millions of individual small donors (not beholden to any billionaires or government funding)

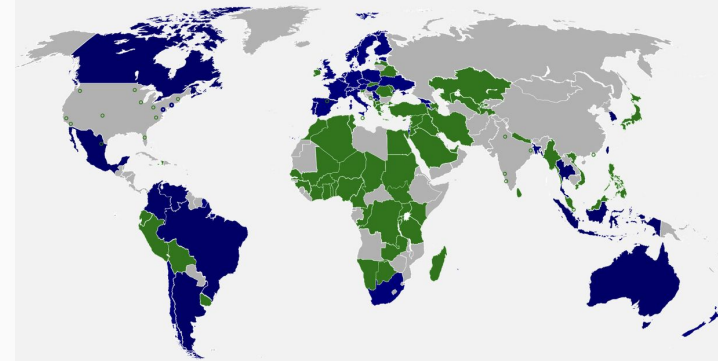
Content projects specialized by linguistic edition

 Wikipedia The free encyclopedia	 Wiktionary Free dictionary and thesaurus	 Wikinews Free-content news
 Wikivoyage Free travel guide	 Wikiquote Collection of quotations	 Wikiversity Free learning resources
 Wikisource Free-content library	 Wikibooks Free textbooks and manuals	

Multilingual content projects

 Wikimedia Commons Free media repository	 Wikidata Free knowledge base	 Wikispecies Free directory of species
 Wikimedia Incubator For language versions in development	 Wikifunctions Free code repository	

Map of geographic based affiliates



As of 24 July 2025

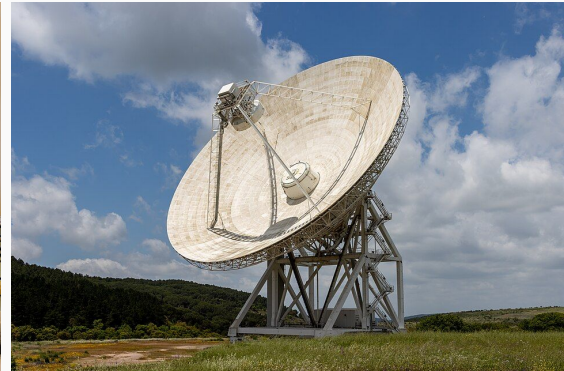
■ Wikimedia chapters (blue)
■ Wikimedia user groups with a geographic focus (green)

"Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing."

https://en.wikipedia.org/wiki/Wikipedia:Prime_objective

Using Wikimedia for science communication

- Wikipedia used to be the ~first result on a search engine (now second behind AI responses...)
- Important to get information right on Wikipedia
- Anyone can edit
- ... but referenced content rules, along with neutral point of view
- ... and be careful of conflicts of interests
- Uploading photos, releasing content under CC-BY-SA, also helps!
 - (The photos used to illustrate this event's webpage demonstrate this well...!)



Content reuse by LLMs

- Because Wikipedia is free, it's one of the first things that LLMs learn from
- (CC-BY-SA: LLMs forget the attribution and licence information...)
- Two types of uses: database dumps (static copies) and live scraping
- Getting content right on Wikipedia -> LLMs will (hopefully) also get it right
- (Interestingly, Wikidata was set up to be computer-parsable, but LLMs are trained on Wikipedia content...!)
- Wikimedia Commons media files also being increasingly used/scraped by AIs...

Potential impacts on Wikipedia...

- AI scraping uses a lot of resources...! Around half of visits to Wikipedia are by AI now (and they are the 'expensive' visits to the main servers).
- Content is free to use, server resources are not...
- Wikimedia Enterprise provides a fast, paid, API service for big reusers (<https://enterprise.wikimedia.com/>)
- API/scraping requests now being limited - check your tools!
 - Need to include user agents / log in to have higher rate limits
- Recruiting new editors becomes more difficult
 - If people don't know they are getting content from Wikipedia, how would they know to edit Wikipedia to correct/update/etc.?



WIKIMEDIA
ENTERPRISE

Generating encyclopaedia articles with AI

- Wikipedia is human-written - that won't be changing
- Neutral Point of View really important on Wikipedia
- Grokipedia basically bulk-copied Wikipedia articles and started AI-updating them to have a right-wing bias ... that's not going well.
- People have tried using AI to write Wikipedia articles ... can provide a good first step to avoid a blank page, but typically needs significant rewriting and referencing to meet Wikipedia standards
- LLM articles recently banned on English Wikipedia, except for suggesting copyediting and translating articles from other language Wikipedias:
https://en.wikipedia.org/wiki/Wikipedia:Writing_articles_with_large_language_models

- Use AI tools to support editors, not replace them
- (Not actually new: machine learning has been used for many years for anti-vandalism work!)
- E.g., check articles being drafted for 'peacock' words, and suggest to the editor to reword them (pointing to Wikipedia policies)
- Make suggestions of edits/automated tooling to help with repetitive tasks

https://meta.wikimedia.org/wiki/Strategy/Multigenerational/Artificial_intelligence_for_editors

Why contributing content to Wikipedia is still important

- If you want to make sure LLMs get things right ... making sure Wikipedia is right is a good first step!
- Companies use all data/images regardless of copyright ... but Wikipedia doesn't - still important to use CC-BY-SA or similar!
- It's a global community of like-minded people who deeply care about getting facts right - natural allies (very similar to academia mindsets!)
- The best way to learn is to teach ... encouraging students to write Wikipedia articles is a great and impactful intro to science writing!

- Collaborative initiative focusing on open knowledge and AI
 - Ongoing series of talks and research activities
 - Led by UoL, in collaboration with Wikimedia Affiliates
 - Upcoming symposium:
 - 28 May: pre-symposium training day (intro to Wikidata etc.)
 - 29 May: Main event, Senate House, London, and online
- Free registration, sign up at:
- <https://ai-bridges.org/2026/02/13/symposium/>
- For more info, see: <https://ai-bridges.org/>



Lots of open questions:

- Want to get knowledge to readers
Does it matter if that's through AI or the website?
- People used to say 'don't trust Wikipedia, read other sources'
Now they say 'don't trust AI, use Wikipedia'
- Need new editors to ensure that Wikipedia is multi-generational
If AI means readers don't convert to editors, what happens?
- Need funding to keep the projects going
Will donors still see fundraising banners? Or fundraise otherwise.

However, Wikipedia isn't going anywhere!

"May you live in interesting times"

(not a Chinese curse?)

https://en.wikipedia.org/wiki/May_you_live_in_interesting_times

Thanks for listening!
Questions?

[[User:Mike Peel]

<https://www.mikepeel.net/>
email@mikepeel.net