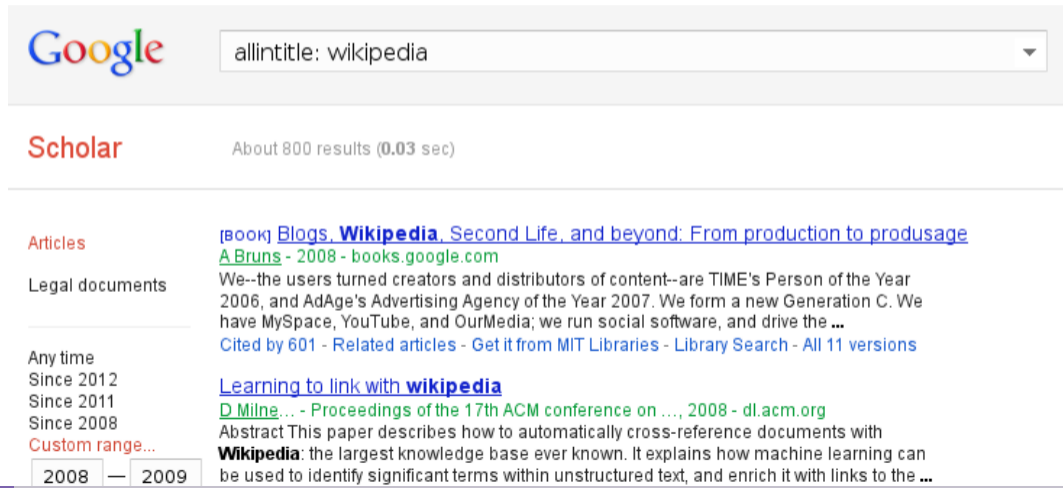


“This talk will try to [provide] a quick tour ... of the last year’s academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project.”

– From Mako’s Wikimania 2008 submission

“This talk will try to [provide] a quick tour ... of the last year’s academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project.”

– From Mako’s Wikimania 2008 submission



The image shows a screenshot of a Google Scholar search interface. At the top left is the Google logo. To its right is a search bar containing the text "allintitle: wikipedia". Below the search bar, the word "Scholar" is displayed in red, followed by the text "About 800 results (0.03 sec)".

On the left side, there are filters for "Articles" and "Legal documents". Under "Articles", there is a link to a book: "[book] [Blogs, Wikipedia, Second Life, and beyond: From production to produsage](#) [A Bruns](#) - 2008 - [books.google.com](#)".

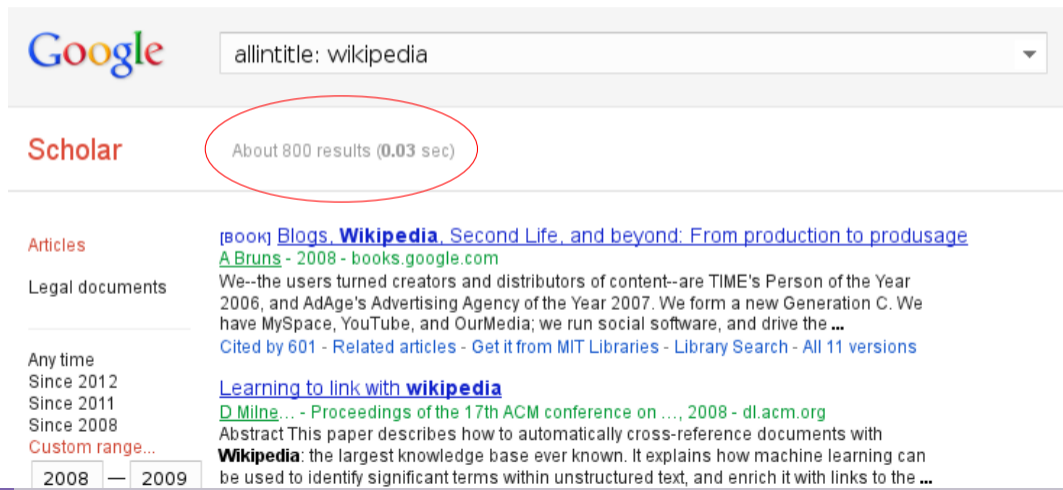
Under "Legal documents", there is a snippet of text: "We--the users turned creators and distributors of content--are TIME's Person of the Year 2006, and AdAge's Advertising Agency of the Year 2007. We form a new Generation C. We have MySpace, YouTube, and OurMedia; we run social software, and drive the ...". Below this snippet are links: "Cited by 601 - [Related articles](#) - [Get it from MIT Libraries](#) - [Library Search](#) - [All 11 versions](#)".

At the bottom left, there are filters for "Any time", "Since 2012", "Since 2011", "Since 2008", and "Custom range...". The "Custom range..." filter is expanded to show a range from "2008" to "2009".

Under "Legal documents", there is another link: "[Learning to link with wikipedia](#) [D Milne...](#) - [Proceedings of the 17th ACM conference on ...](#), 2008 - [dl.acm.org](#)". Below this link is a snippet of text: "Abstract This paper describes how to automatically cross-reference documents with **Wikipedia**: the largest knowledge base ever known. It explains how machine learning can be used to identify significant terms within unstructured text, and enrich it with links to the ...".

“This talk will try to [provide] a quick tour ... of the last year’s academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project.”

– From Mako’s Wikimania 2008 submission



The image shows a screenshot of a Google Scholar search interface. At the top left is the Google logo. To its right is a search bar containing the text "allintitle: wikipedia". Below the search bar, the word "Scholar" is displayed in red. A red oval highlights the text "About 800 results (0.03 sec)". Below this, there are two search results. The first result is under the heading "Articles" and is titled "[book] Blogs, Wikipedia, Second Life, and beyond: From production to produsage" by A Bruns, with a link to books.google.com. The second result is under the heading "Legal documents" and is titled "Learning to link with wikipedia" by D Milne, with a link to dl.acm.org. On the left side of the page, there are filters for "Any time", "Since 2012", "Since 2011", "Since 2008", and "Custom range...". At the bottom, there are input boxes for the years "2008" and "2009" with a minus sign between them.

Google

allintitle: wikipedia

Scholar

About 800 results (0.03 sec)

Articles

[\[book\] Blogs, Wikipedia, Second Life, and beyond: From production to produsage](#)
[A Bruns - 2008 - books.google.com](#)

Legal documents

We--the users turned creators and distributors of content--are TIME's Person of the Year 2006, and AdAge's Advertising Agency of the Year 2007. We form a new Generation C. We have MySpace, YouTube, and OurMedia; we run social software, and drive the ...
[Cited by 601 - Related articles - Get it from MIT Libraries - Library Search - All 11 versions](#)

[Learning to link with wikipedia](#)
[D Milne... - Proceedings of the 17th ACM conference on ..., 2008 - dl.acm.org](#)
Abstract This paper describes how to automatically cross-reference documents with **Wikipedia**: the largest knowledge base ever known. It explains how machine learning can be used to identify significant terms within unstructured text, and enrich it with links to the ...

Any time

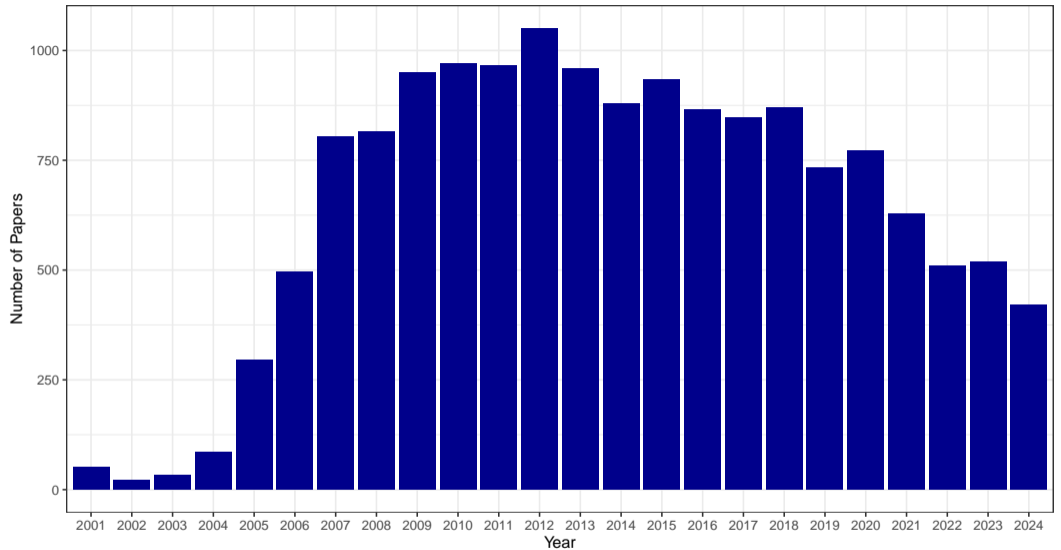
Since 2012

Since 2011

Since 2008

Custom range...

2008 — 2009



Number of items, per year, with the term “wikipedia” in the title.

(Source: Google scholar results. Accessed: 2025-08-07)

- 200+ tweets from @WikiResearch account on Twitter/X (covering research papers, events, blog posts etc.)
- 92 recent publications covered in the issues of the [Wikimedia Research Newsletter](#) from August 2024 to July 2025 (and hundreds more on our to-do list!)
- 46 extended abstracts presented at the Wiki Workshop 2025 in May 2025

Themes and Papers



This presentation has multiple issues. Please help [improve it](#) by asking questions and making comments along the way.

- This presentation is [horribly biased](#), as it describes the articles that seemed **interesting to me**.
(July 2012)
- The [comprehensiveness](#) of this presentation is [impossible](#). Please read the [Wikimedia Research Newsletter](#) to get a more complete view.
(July 2012)

In selecting papers, our goal was to choose work that:

- Represents [important themes](#) from Wikipedia in the last year.
- Is likely to be of [interest](#) to Wikimedians.
- Research by people who are [not at Wikimania](#).
- ...with a bias towards [peer-reviewed](#) publications

Theme 1. Effects of generative AI on Wikimedia projects

Theme 1. Effects of generative AI on Wikimedia projects

Lyu, Liang, James Siderius, Hannah Li, Daron Acemoglu, Daniel Huttenlocher, and Asuman Ozdaglar. 2025. "Wikipedia Contributions in the Wake of ChatGPT." *Companion Proceedings of the ACM on Web Conference 2025* (New York, NY, USA), WWW '25, May 23, 1176–79. <https://doi.org/10.1145/3701716.3715543>.



Search Wikipedia

OceanGate

Contents [hide](#)

(Top)

[Background](#)

[History](#)

2009–2013: Founding

[Article](#) [Talk](#)

From Wikipedia, the free encyclopedia

This article is about the submersible company. For the bus company, see OceanGate Bus Lines.

OceanGate Inc. is an American privately owned company be



OceanGate Inc.

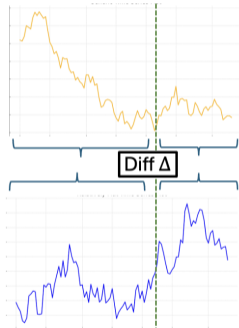
OceanGate Inc. is a privately held company specializing in the operation of manned submersibles for deep-sea exploration

OpenAI's
Embedding
Model

Similar

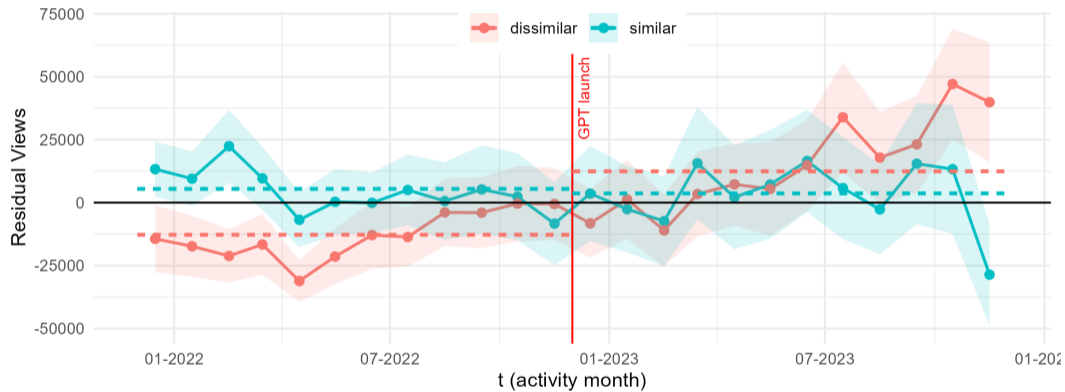
Dissimilar

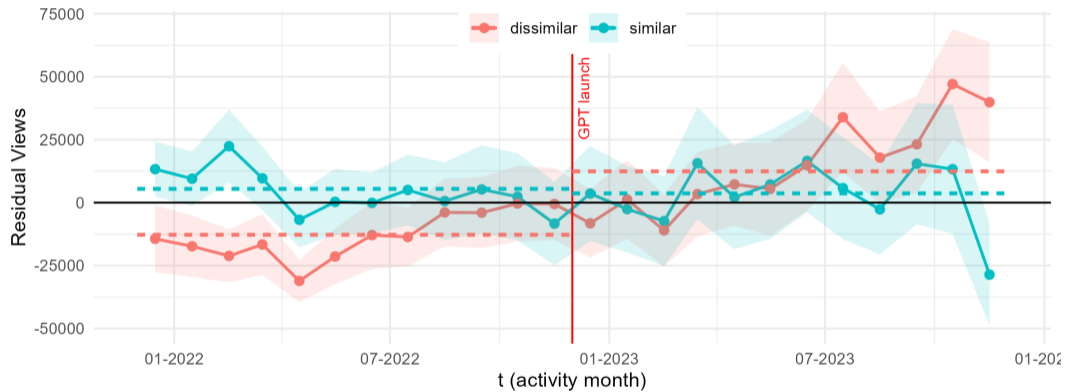
Example Edits (demonstration purposes only)



November 2022
Intro of ChatGPT

Diff-in-Diff
 $\Delta_S - \Delta_D$

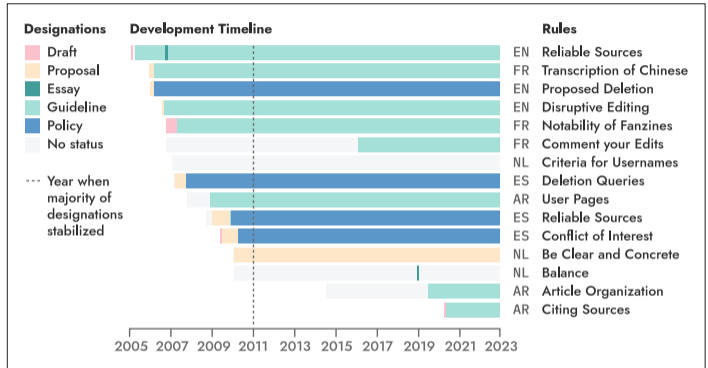




Theme 2. Governance

Jankowski, Steve, Claudio Celis Bueno, Ouejdane Sabbah, and Jakko Kemper. 2025. "Templates and Sovereignty: Wikipedia's Policy Development and the Reflection of Community Consensus." *New Media & Society*, May 21, 14614448251336436. <https://doi.org/10.1177/14614448251336436>.

- Full edit histories of 15 important policies across five Wikipedia language editions (EN, ES, AR, FR, NL)
- Discourse analysis and qualitative content analysis

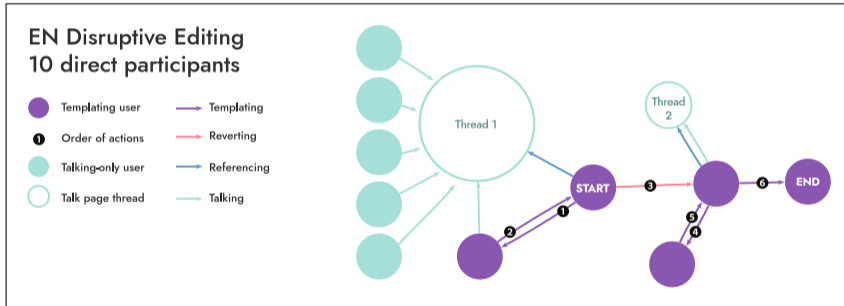


The authors identify:

- Types of **contest** (e.g., change `{{policy}}` to `{{notpolicy}}`)
- Kinds of **authority** (e.g., democratic, administrative, editorial, document)

The authors identify:

- Types of **contest** (e.g., change `{{policy}}` to `{{notpolicy}}`)
- Kinds of **authority** (e.g., democratic, administrative, editorial, document)



- In most cases, “a very small number of Wikipedians used meritocratic techniques to *determine* consensus, rather than *reflect* it.”
- Consensus is constructed in different ways. And there is substantial variation across/within wikis!
- Concrete opportunities (and a range of examples!) of how communities can approach “consensus” in ways that are more participatory, robust, and effective.



Wikimedia Research Fund

Provides support to individuals, groups, and organizations with research interests on or about Wikimedia projects.



Theme 3. Underrepresented Language Editions

Theme 3. Underrepresented Language Editions

Nigatu, Hellina Hailu, John Canny, and Sarah E. Chasins. 2024. "Low-Resourced Languages and Online Knowledge Repositories: A Need-Finding Study." *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA), CHI '24, May 11, 1–21. <https://doi.org/10.1145/3613904.3642605>.

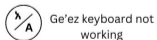
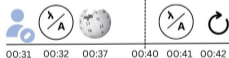
- Analysis of talk pages on Wikipedia language editions (*established user needs*)
- Observation and interviews of newcomer editing experiences (*new user needs*)

New Users: Problems with search, text input, VPNs, etc.

Finds a title with a page that does not exist.



Finds a Stub. Tries to edit.



Ge'ez keyboard not working



Main page of Wikipedia



IP address blocked



Microsoft Word

- Identifies **general takeaways** that affect low-resourced language Wikipedias and other knowledge commons.
- Creates **new insights and opportunities** for improving the experience of contributors—established and new—to low-resourced languages
- Documents a **method** for studying challenges faced by underrepresented languages more generally.

WMF Research Award of the Year (WMF-RAY)



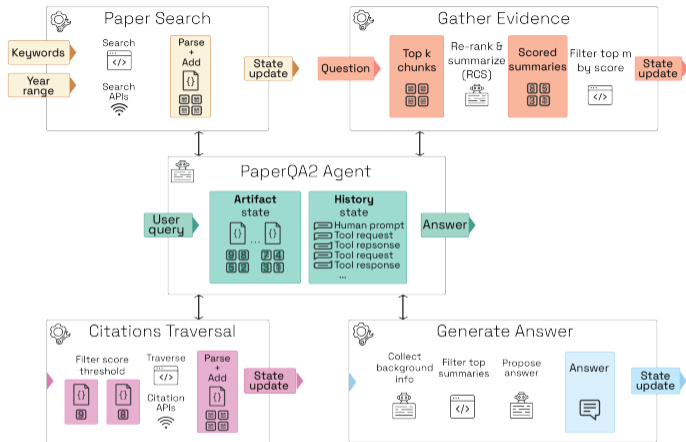
Theme 4. AI writing Wikipedia-like articles

Theme 4. AI writing Wikipedia-like articles

Skarlinski, Michael D., Sam Cox, Jon M. Laurent, et al. 2024. "Language Agents Achieve Superhuman Synthesis of Scientific Knowledge." arXiv:2409.13740. Preprint, *arXiv*, September 26. <https://doi.org/10.48550/arXiv.2409.13740>.

- The authors' AI-based open-source "WikiCrow" system "writes cited, Wikipedia-style summaries" about individual human genes
- "that are significantly more accurate than existing, human-written Wikipedia articles" (as judged by subject matter experts in blind grading).
- Published by "FutureHouse", a San-Francisco-based nonprofit working on "Automating scientific discovery" (with a focus on biology).
- Motivation: Wikipedia articles on individual human genes are useful for molecular biology researchers.
- However: "only 3,639 of the 19,255 human protein-coding genes [...] have high-quality (non-stub) summaries on [English] Wikipedia; the other 15,616 lack pages or are incomplete stubs."

- "PaperQA2": A general system for scientific literature research tasks. Includes multiple LLMs ("agents"), search etc.
- "WikiCrow" calls it several times to build a Wikipedia-like article
- Much more complicated than a simple 2022-era "Write me a Wikipedia article" ChatGPT query
- Expensive (USD 5.50 per article)



Compared to the existing (human-written) Wikipedia articles, the AI-generated articles

- were longer
- had fewer uncited statements
- had **fewer** "reasoning errors" or hallucinations (where "the written information contradicts, over-extrapolates, or is unsupported by any included citations").

	WikiCrow	Wikipedia
Average Article Length (words)	1,219	890
Total Statements Graded	189	186
Valid Statements	171	169
Cited and Supported (Accuracy)	83.0%	61.5%
Uncited	3.5%	13.6%
Cited and Unsupported	13.5%	24.9%
Supported Among Cited (Precision)	86.1%	71.2%

D.

	WikiCrow	Wikipedia
Unsupported Breakout (total)	23	42
Reasoning Issues	12	26
Attribution Issues	10	16
Trivial Statements	1	0

Outlook:

- "WikiCrow" was optimized for a **very specific use case** (writing gene articles), with much effort. In general, AI is not yet ready to replace us as Wikipedians.
- But progress is continuing. The **"deep research" AI tools** released by OpenAI and several other AI companies earlier this year, designed to create comprehensive, cited reports about a given topic, aim at users who might well have looked up that topic on Wikipedia instead.

Theme 5. Offline Matters

Schwitter, Nicole. 2024. "How Offline Meetings Affect Online Activities: The Case of Wikipedia." *EPJ Data Science* 13 (1): 67.

<https://doi.org/10.1140/epjds/s13688-024-00506-w>.

The dataset:

On-wiki participant signups for 4408 community meetups, covering "all smaller-scale gatherings organised within the German-language version of Wikipedia", from its launch in 2001 until 2020 (COVID).

- Hypothesis 1: "Attendees of offline meetups increase their contributions after a meetup."
 - supported: "attending an offline meetup has a positive statistical effect on the contribution behaviour of users"

The researcher also distinguished "work-related" meetups (focused on particular tasks and activities, in particular editathons) and meetups focused on socializing.

- Hypothesis 2: "Attendees of work-related offline meetups increase their contributions more than attendees of social meetups after attending."
 - "only little support": After one month or one year, "there are no significant differences between users attending a work or a social meeting" in terms of their editing activity levels.

- Outcome: An editor's edit counts within particular time spans after the meetup.
- To compare with the hypothetical case that the same editor had decided not to attend the meetup, "covariate matching" was used to find the most similar non-attendee among all other German Wikipedia editors.
- Similarity defined using account age and several other metrics based on an editor's edit counts within particular timespans before the meetup, separately for mainspace (i.e. Wikipedia articles) and non-mainspace (e.g. talk pages).

Table 1 Changes in editing activity after first meetup compared to activity before

Topic	Group	7 days	28 days	364 days
Changes in editing behaviour after first meetup: mainspace	Treatment	0.32 (51.87)	4.18 (172.95)	103.80 (8.44)*
		0	0	0
	Control	-1407 / 905	-2916 / 4385	-18294 / 57852
		-1.85 (39.48)***	-8.95 (128.51)***	-89.09 (1262.75)***
		0	0	0
Difference Treatment - Control	-819 / 487	-1973 / 1877	-15624 / 23170	
	2.17*	13.13***	192.86***	

Concluding Thoughts

More Resources

- @WikiResearch on [Twitter/X/Mastodon/Bluesky](#)
- [Wikimedia Research Newsletter: \[\[:meta:Research:Newsletter\]\]](#)
- [Wiki Workshop](#) (annually)
- [\[\[:meta:Research:Events\]\]](#)
- [WMF Research Showcase](#) (monthly)

