

Wikipedia × PubPeer: Which Citations Are Under Scrutiny?

Cristian Cazzaniga
DISI, University of Trento

Cristian Consonni^{*,†}
European Commission
Joint Research Centre (JRC)

Abstract

We conduct a large-scale, multilingual analysis of over 2.4 million DOI-based citations across the English, French, Italian, and Spanish Wikipedia editions by linking them to PubPeer post-publication reviews and automatically classifying comment content, showing that a small but stable fraction of citations ($\approx 0.3\%$) is subject to scrutiny and is disproportionately associated with recent publications and high-impact journals, highlighting post-publication review as a complementary signal for monitoring citation reliability in Wikipedia.

1 Introduction

Wikipedia is one of the most widely consulted sources of scientific information worldwide (Areia et al., 2025), and its credibility depends on the quality of the sources it cites. Wikipedia relies upon reliable sources to provide verifiable claims in its articles; in this context, scholarly articles are often used in Wikipedia as sources. As of December 2025, English Wikipedia references more than 3.5M DOIs. Scientific citations play a central role in Wikipedia, with most citations belonging to highly-ranked journals in STEM fields (Lewoniewski et al., 2023; Yang and Colavizza, 2022; Guglielmi and others, 2018), but also with a substantial share of citations going to Open Access venues (Yang et al., 2024). It is therefore important to access the quality and reliability of citations in Wikipedia. In this study, we present the preliminary results of a large-scale, multilingual analysis of DOI-based citations in Wikipedia that have received post-publication reviews on PubPeer. We extracted all DOI citations from the English, French, Italian, and Spanish Wikipedia editions, retrieved bibliographic metadata from CrossRef, and matched them with PubPeer comments to identify potentially controversial publications.

PubPeer is an online platform that enables researchers to discuss published scientific articles, allowing comments to be posted either anonymously or openly. The

authors can respond to these comments, providing clarifications or additional explanations. A comment on PubPeer does not automatically imply a fatal error in the publication; it serves as a signal of scientific scrutiny and post-publication review. The platform provides a page for each publication that has a Digital Object Identifier (DOI) and let users comment with free-form text. A PubPeer comment can contain images, tables, mathematical formulae. PubPeer comments generally cover a wide range of concerns, including methodological issues, anomalies in figures, data validity, and potential ethical problems or conflicts of interest. Figure 1 illustrates an example PubPeer comment, showing how critical observations are communicated and how authors may respond.

2 Methodology

Our methodology, illustrated in Figure 2, comprises four sequential phases:

1. **Data Extraction.** The process begins with the extraction of DOI citations from the external links of English, French, Italian, and Spanish Wikipedia using PAWS.¹ Pages are then classified by namespace using the MediaWiki API.
2. **Bibliographic Enrichment.** In the second phase, we used the Crossref² API to retrieve bibliographic metadata for each identified DOI. We integrated author lists, publication years, and citation counts.
3. **DOI Matching.** Wikipedia DOIs are matched against PubPeer data, to identify the *potentially problematic references*, i.e. the DOIs that have attracted post-publication comments on PubPeer.
4. **Comment Classification.** The final processing stage categorizes the content of the retrieved comments. As shown, the comments are vectorized using TF-IDF and fed into a Random Forest Classifier

¹PAWS (<https://wikitech.wikimedia.org/wiki/PAWS>) is a free, cloud-hosted environment using Jupyter notebooks for contributors to run code, analyze data, and automate tasks related to Wikipedia and other Wikimedia projects.

²Crossref (<https://www.crossref.org/>) is the largest Digital Object Identifier (DOI) registration agency in the world. Crossref enables persistent citation linking by assigning DOIs to research outputs and ensuring metadata interoperability.

* *Disclaimer:* The view expressed in this paper is purely that of the authors and may not, under any circumstances, be regarded as an official position of the European Commission.

[†] Corresponding author: cristian.consonni@acm.org

trained on 338 manually annotated records. The model assigns comments to 5 categories: methodological concerns, figure anomalies, clarification, data validity, and ethical issues.

3 Results and Discussion

This pipeline allows us to compute the rate of problematic citations, the distribution of comment categories, and the enriched metadata profile of the problematic sources. We analyzed 2,450,456 unique DOI citations across English (en), French (fr), Italian (it), and Spanish (es) Wikipedia editions, of which 7,016 had at least one PubPeer comment.

Citation usage and potentially problematic citations across languages. Table 1 shows the prevalence of citation usage across Wikipedia language editions. English and French Wikipedia exhibit a higher density of scientific citations, whereas fewer than 1% of articles in the Italian and Spanish editions contain at least one DOI. Notably, the proportion of problematic citations is stable across all languages, at approximately 0.3%. This suggests that the risk of citing unreliable research, which is largely shaped by editorial practices, is comparable across language communities, while overall citation usage appears to be influenced by language-specific factors. Consequently, the total fraction of Wikipedia articles citing potentially problematic papers depends mainly on the overall usage of citations, in English Wikipedia 1.37% while in Italian Wikipedia only 26 pages in total cite problematic DOIs.

Source and temporal distribution of citations. Figure 3 shows the distribution of publication years for DOIs cited on Wikipedia, whereas Figure 4 displays the distribution of potentially problematic DOIs among those cited. Across all language editions, more than 29% of cited publications were published after 2015. This temporal skew is more pronounced among citations with PubPeer comments, where as much as 45% of works were published after 2015. This pattern suggests that PubPeer commentary is more commonly directed toward newer scientific work. High-impact journals such as *Nature* (3rd) and *Science* (8th) are among the most cited on Wikipedia between 2015 and 2025 (Fig. 5). These journals also occupy the top two positions when ranked by the number of potentially problematic DOIs (Fig. 6), suggesting that high-profile venues may attract greater post-publication scrutiny.

Categorization of comments on PubPeer. We evaluated the automated categorization system on a test set of 100 manually labeled comments. The model achieved an accuracy of 72%, with a macro F1-score of 55.4% and a weighted F1-score of 72.2%. Figure 7 shows the distribution of categories by language, highlighting that methodological rigor issues (40%) and figure anomalies

(25%) are consistently the most frequent concerns across DOIs that received PubPeer comments across all Wikipedia language editions.

Dashboard. We published a dashboard at <https://pubpeer.knowledgeintegrity.wiki> that lists Wikipedia articles containing potentially problematic citations across the English, French, Italian, and Spanish editions. The dashboard is intended to support Wikipedia editors highlighting articles that cite multiple potentially problematic papers.

Conclusions

This work provides the first systematic, multilingual analysis of Wikipedia’s scientific citations in the context of post-publication peer review. By linking DOI-based references across four Wikipedia language editions with PubPeer discussions, we showed that while Wikipedia largely relies on reputable sources, a small but measurable fraction of citations is subject to ongoing scientific scrutiny. These results highlight the potential of post-publication review as a complementary signal for monitoring citation reliability in Wikipedia. Future work will focus on improving topic classification of Wikipedia articles (Johnson et al., 2021), conducting topic-specific analyses, and identifying high-priority articles to guide source verification efforts.

References

- [Areia et al.2025] Carlos Areia, Kath Burton, Mike Taylor, and Charles Watkinson. 2025. Research citations building trust in wikipedia: Results from a survey of published authors. *PLoS One*, 20(4):e0320334.
- [Guglielmi and others2018] Giorgia Guglielmi et al. 2018. Wikipedia’s top-cited scholarly articles revealed. *Nature*, 557(7705):291.
- [Johnson et al.2021] Isaac Johnson, Martin Gerlach, and Diego Sáez-Trumper. 2021. Language-agnostic topic classification for Wikipedia. In *Companion Proceedings of the Web Conference 2021*, pages 594–601.
- [Lewoniewski et al.2023] Włodzimierz Lewoniewski, Krzysztof Weceł, and Witold Abramowicz. 2023. Understanding the use of scientific references in multilingual Wikipedia across various topics. *Procedia Computer Science*, 225:3977–3986.
- [Yang and Colavizza2022] Puyu Yang and Giovanni Colavizza. 2022. A map of science in Wikipedia. In *Companion Proceedings of the Web Conference 2022*, pages 1289–1300.
- [Yang et al.2024] Puyu Yang, Ahad Shoaib, Robert West, and Giovanni Colavizza. 2024. Open access improves the dissemination of science: insights from wikipedia. *Scientometrics*, 129(11):7083–7106.

Wiki Workshop (13th edition) – March 25-26, 2026

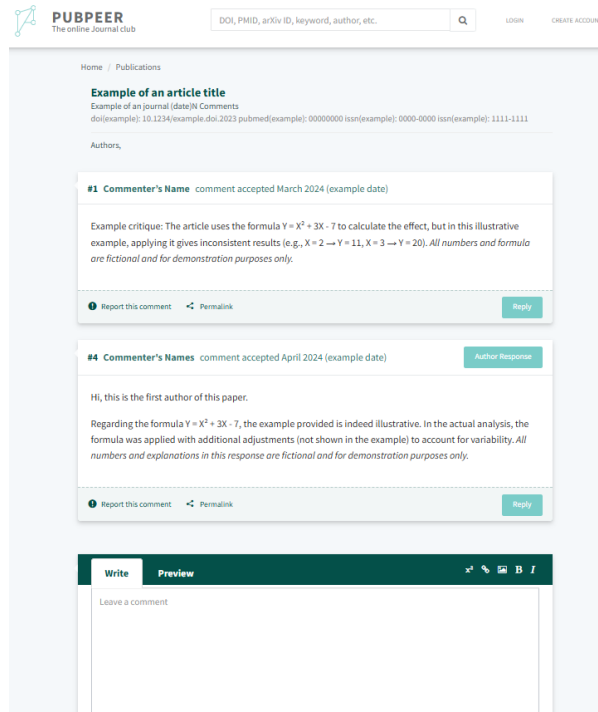


Figure 1: An example discussion thread on PubPeer. We do not present a real instance, to avoid overexposing a specific paper and authors.

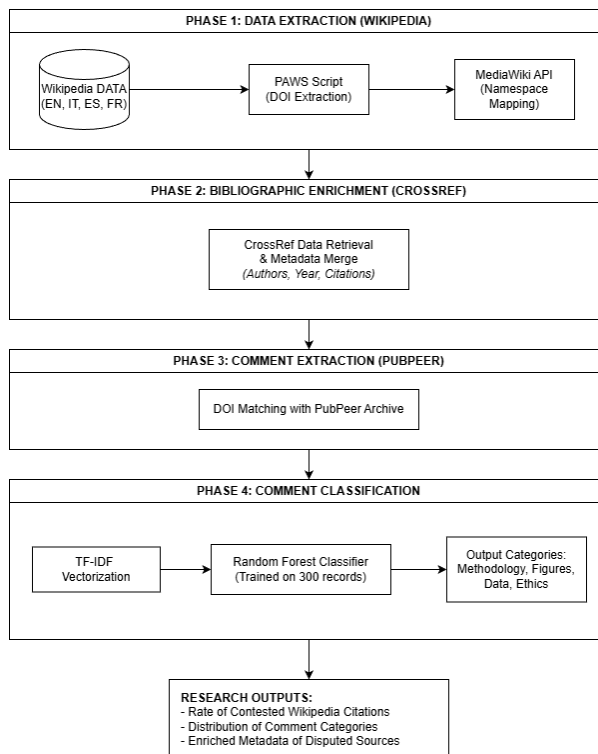


Figure 2: Workflow structure.

Wiki Workshop (13th edition) – March 25-26, 2026

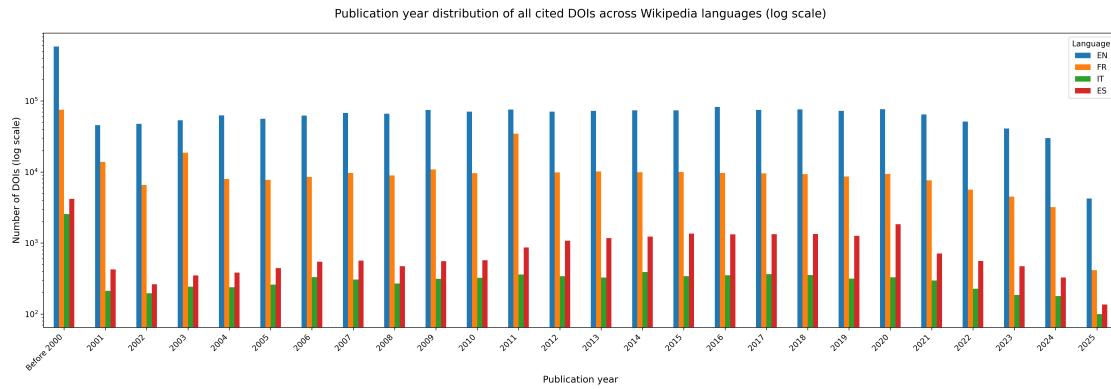


Figure 3: Publication year distribution of all DOIs cited across the English, French, Italian, and Spanish Wikipedia editions (log scale).

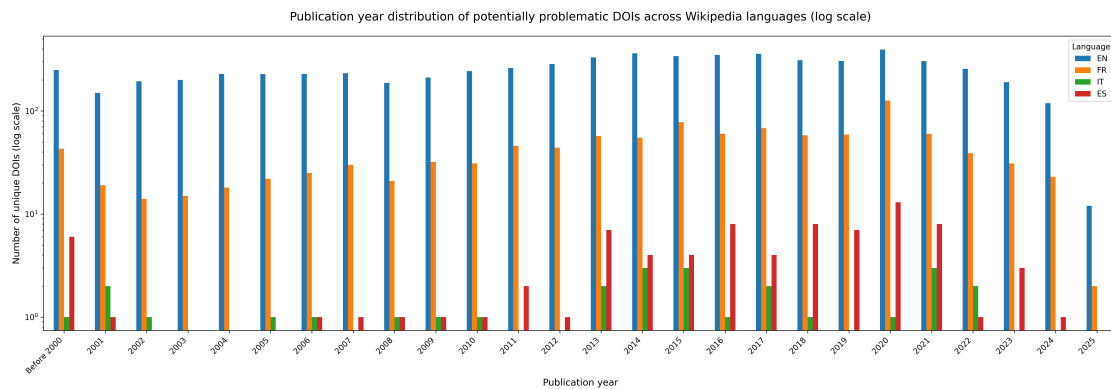


Figure 4: Publication year distribution of DOI-based citations that have received PubPeer comments across the English, French, Italian, and Spanish Wikipedia editions (log scale).

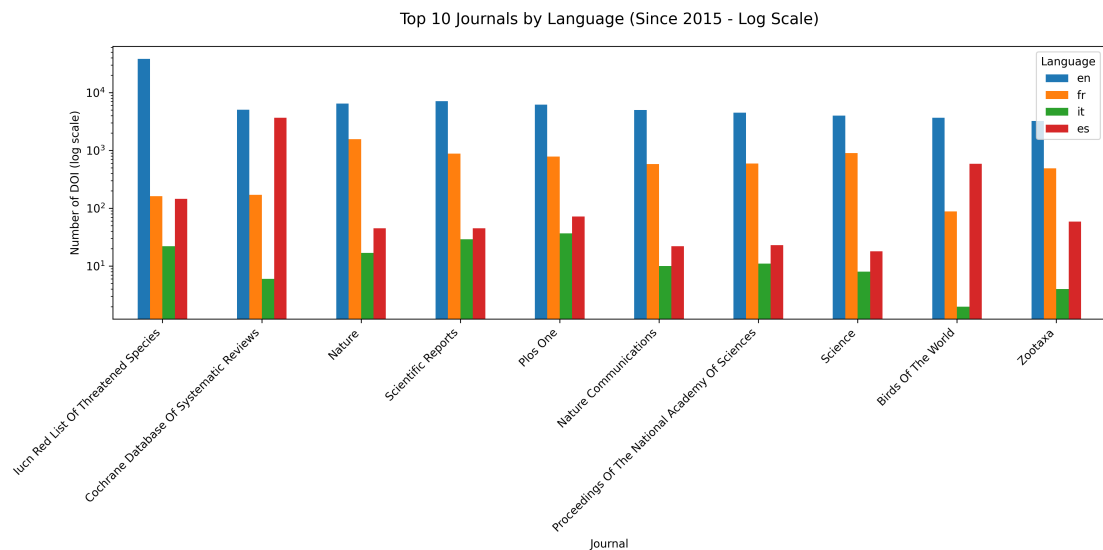


Figure 5: Top-10 journals by total number of DOIs cited across English, French, Italian and Spanish Wikipedia (log scale). We consider only articles published since 2015.

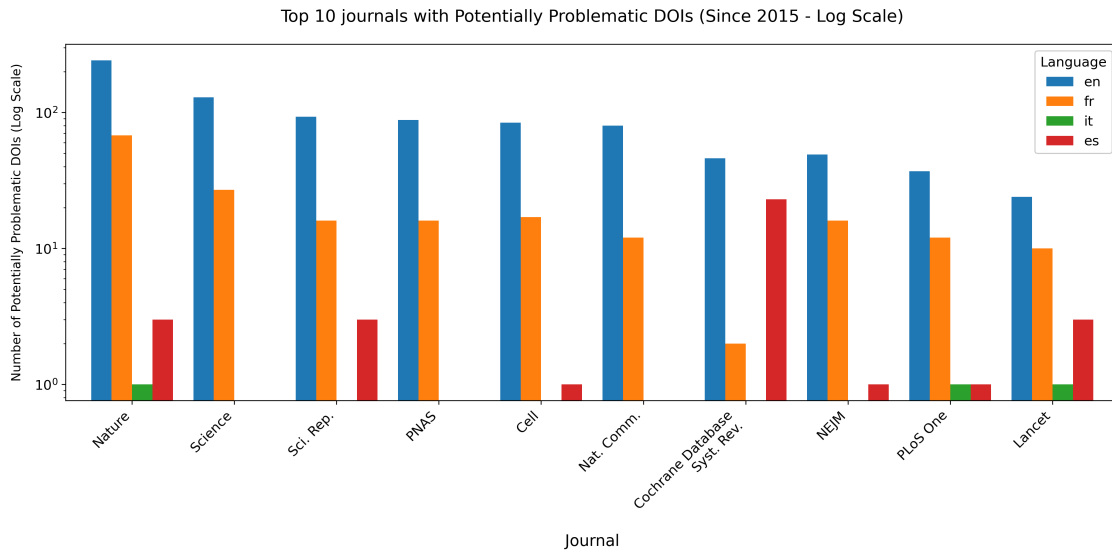


Figure 6: Top-10 Journals by number of DOI-based citations to potentially problematic DOIs across English, French, Italian and Spanish Wikipedia (log scale). We consider only articles published since 2015.

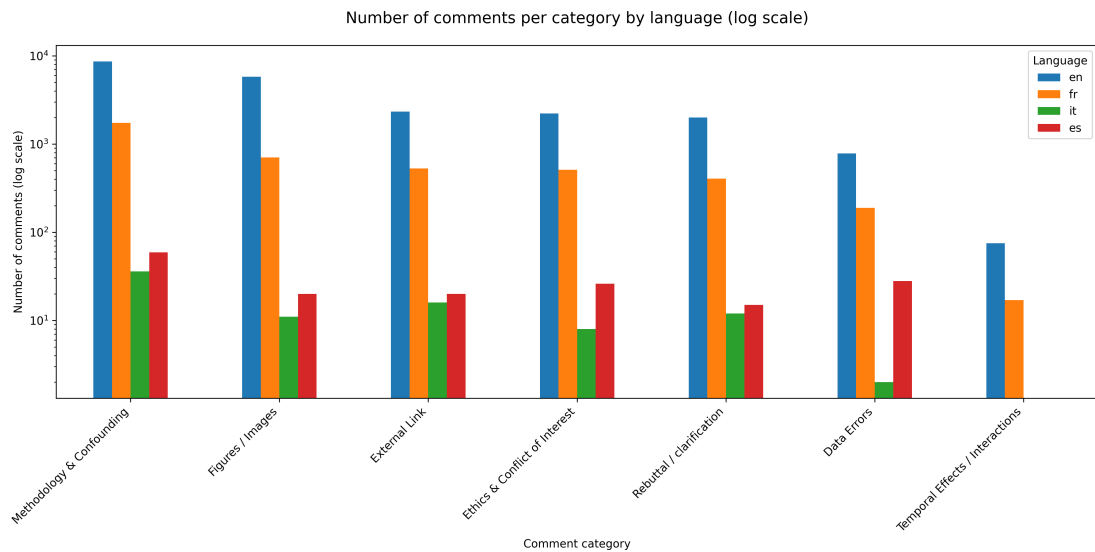


Figure 7: Number of PubPeer comments per category across English, French, Italian, and Spanish Wikipedia editions (log scale).

		Wikipedia			
		en	fr	it	es
DOIs	Total pages	7,114,670	2,729,293	1,950,287	2,084,345
	Pages containing at least one DOI (% of pages over total)	842,989 (11.85%)	190,309 (6.97%)	9,105 (0.47%)	14,016 (0.67%)
	No. of DOIs referenced in Wikipedia (average no. of times used)	3,575,713 (1.59×)	455,523 (1.33×)	15,020 (1.42×)	33,442 (1.32×)
	Unique DOIs cited in Wikipedia	2,252,545	343,446	10,557	25,306
PubPeer	DOIs in Wikipedia with comments son PubPeer (% of pages with a DOI)	6,660 (0.30%)	1,087 (0.32%)	27 (0.26%)	85 (0.34%)
	No. of comments on cited DOIs	21,761	4,097	85	168
	Wikipedia pages with DOIs with comments on PubPeer (% of pages over total)	9,763 (1.37%)	1,166 (0.43%)	26 (0.013%)	57 (0.03%)

Table 1: Overview of DOI usage in Wikipedia on 4 selected language editions—English (en), French (fr), Italian (it) and Spanish (es)—and the corresponding PubPeer comments.