

Wikipedia as Infrastructure for Knowledge Poisoning: Bias Mechanisms and Countermeasures

Shlomit Lir¹ | September 9, 2026

In recent years, English Wikipedia has become a central link in a global ecosystem of generating, validating, and disseminating biased information. This information is replicated across search engines and artificial intelligence systems, through which it reaches broad audiences worldwide as reliable and authoritative knowledge. The systematic activity of a coordinated, ideologically motivated network of editors turns the platform into a significant infrastructure for knowledge poisoning with regard to Israel. This is executed through one-sided framing, selective source picking, the omission of context, the alteration of key terminology, and the construction of a narrative frame that presents the establishment and existence of the State of Israel as an ongoing injustice, thereby undermining its very right to exist. The products of this knowledge bias pose a strategic threat to Israel's resilience and its ability to defend its standing in the international arena, with repercussions that are also evident among Jewish communities worldwide. This article proposes ways to address the knowledge poisoning occurring on English Wikipedia, including measures required both within and outside the platform. It examines the operational mechanisms of knowledge bias on the platform and discusses the challenges facing those who seek to safeguard the neutrality of knowledge.

¹ Dr. Shlomit Aharoni Lir is an award-winning researcher and a research fellow at the University of Haifa and Bar-Ilan University. She specializes in the politics of knowledge in the contexts of gender, culture, and technology and serves as a senior advisor to Here4Good. She is a founding member of Forum Dvora, in addition to being an editor and author who has published eight books. Dr. Lir presented a report on bias against Israel on English Wikipedia at the United Nations in April 2024, founded WikiBiases, a research platform on Wikipedia biases, and curated. Her traveling exhibition, "Manipulated History," a traveling exhibition which compares Wikipedia entries before and after October 7, that has been exhibited in Jerusalem and in other locations worldwide.

Introduction: Wikipedia as a Global Knowledge Infrastructure

English Wikipedia is currently one of the most influential knowledge arenas in the world. Its grounding in the democratization of information, its perception as an open and neutral source, its prominent placement in online search-engine results, and the retrieval of its content by various systems give it considerable power to establish knowledge that is not subject to academic oversight and validation mechanisms. Its position as one of the primary information sites of the present era stems also from its vast scale of use. As the most widely read site of human knowledge globally, with billions of visits per month and thousands of entries every second, it wields extraordinary power in shaping the conceptual frameworks through which the global public interprets sensitive issues, including attitudes toward Israel and the Israeli-Palestinian conflict.

In recent years, Wikipedia has become one of the primary knowledge infrastructures relied upon by artificial intelligence systems. Its content is used to train models, retrieve information, and formulate responses. It is among the most frequently cited sources in ChatGPT responses, and its content is currently consumed on a massive scale through automated web crawlers, application programming interfaces (APIs), and large-scale data downloads, both for training and operating AI systems and for search engines and knowledge-based applications. These facts make Wikipedia not only the most widely read encyclopedia, but also a central knowledge infrastructure for large language models (LLMs). Consequently, biases that accumulate on Wikipedia do not remain confined to the platform; rather, they are replicated across systems used by billions of people who do not necessarily access Wikipedia directly.

While international recognition of the risks posed to the integrity of the information ecosystem in the age of AI is expanding, the danger stemming from human-produced information generated on Wikipedia and perceived as knowledge has yet to receive sufficient attention. The [United Nations Global Principles for Information Integrity](#) call for increasing transparency, providing data access to researchers, incorporating safeguards into model training processes, and developing mechanisms to detect, label, and verify sources of AI-generated content. However, most emerging measures focus on detecting synthetic or false content produced by these AI systems. They offer a limited solution to situations in which AI systems retrieve biased information originating from a knowledge infrastructure regarded as reliable, such as Wikipedia.

In this context, it is important to note that the problem with Wikipedia is not limited to the mere existence of biased information, but rather lies in the process by which such information is selected, framed, and validated as knowledge within an infrastructure perceived as trustworthy. "Knowledge poisoning" thus describes a state in which bias becomes embedded within a knowledge system, acquires an authoritative status, and is replicated across additional systems and knowledge spaces in a manner that disrupts the capacity for independent thought, based on evaluating competing claims, weighing alternatives, and engaging in open debate.

The combination of English Wikipedia's phenomenal reach and its knowledge architecture, which allows malicious conduct to be concealed under the cover of its foundational principle

to "assume good faith" and anonymous editing practices, has turned it into a primary target for narrative engineering by ideological, political, and organizational actors seeking to shape public knowledge. Regarding Israel, tendentious editing manifests itself in recurring patterns that include selective source picking, one-sided conceptual framing, erasure or weakening of Jewish and Israeli historical contexts, the disproportionate use of negative and judgmental tone, and the manipulation of community governance mechanisms in a manner that constructs a biased knowledge landscape concerning Zionism, Jewish history, the conflict, and terrorist organizations.

The politics of knowledge on Wikipedia is expressed across a variety of domains. Over the years, numerous studies have addressed the platform's gender gap and proposed [strategies](#) to counteract the underrepresentation of female editors. In 2024, systemic bias against Israel also received high-level consideration in a [report](#) presented at a special session of the World Jewish Congress at the United Nations in Geneva in April of that year (Lir, 2024). Growing awareness of this issue is further reflected in [reports](#) published by the Anti-Defamation League (ADL), the [reporting](#) of JNS journalist Aaron Bandler, and an ongoing investigative series by journalist [Ashley Rindsberg](#).

Criticism regarding Wikipedia's departure from its core principle of neutrality has increasingly been voiced by its own founders. For example, Larry Sanger, a co-founder of Wikipedia, has warned for years that it is becoming a vehicle for propaganda. In addition to publishing a book and articles on the subject, in 2026 he launched WikiProject Intellectual Diversity, an initiative intended to strengthen neutrality and broaden the range of perspectives represented within the editing community. Shortly thereafter, following a dispute over the way the initiative had been promoted and allegations that he had recruited an external audience to influence internal discussions, Sanger was indefinitely blocked from editing Wikipedia.

At the same time, criticism regarding the implementation of the neutrality policy was voiced by Jimmy Wales, a co-founder of Wikipedia and member of the Wikimedia Foundation Board of Trustees. In November 2025, Wales participated in a discussion on the talk page of the article "Gaza Genocide," which he described as one of the worst Wikipedia entries he had seen in a long time. He noted that the lead and overall framing stated in Wikipedia's own voice that Israel was committing genocide, even though this was a disputed claim, and called for the article to be revised and the controversy presented in a manner consistent with the neutral point of view (NPOV) principle. His substantive comments provoked fierce, hostile opposition from several editors, culminating in a formal request to impose a topic ban on him to prevent his further participation in edits and discussions on the subject. The reactions to the criticisms posed by the platform's founders raise pressing questions regarding the community's capacity to absorb constructive proposals for improvement and initiatives aiming to enforce true neutrality in practice.

In recent years, criticism has expanded into the institutional arena, manifesting in official inquiries by United States government officials directed to the Wikimedia Foundation. The Foundation is a nonprofit organization that operates the infrastructure supporting Wikipedia

and other projects and assists the volunteer communities that curate their content. Official communications sent to the Foundation include a formal letter from Senator Ted Cruz, Chairman of the U.S. Senate Committee on Commerce, Science, and Transportation, as well as a communication from Ed Martin, the U.S. attorney for the District of Columbia, who raised the possibility of revoking Wikimedia's tax-exempt status due to foreign manipulation of content on the platform. At the same time, House Committee on Oversight and Government Reform members James Comer and Nancy Mace announced the opening of an investigation into coordinated efforts to influence U.S. public opinion through the manipulation of Wikipedia entries.

The following sections present an operational outline for countering knowledge poisoning on English Wikipedia and mitigating the spread of biased information about Israel to search engines and AI systems. These recommendations target multiple audiences, including policymakers, research institutions, educational and media bodies, the Wikimedia Foundation, and the editing community itself. Accordingly, this article distinguishes between external measures involving independent research, monitoring, and oversight, and necessary structural changes to the platform's internal governance and enforcement mechanisms, while addressing the challenges specific to Wikipedia that impede efforts to promote change.

The Architecture of Bias

Research on bias in English Wikipedia points to consistent patterns of conduct by editors acting out of a political agenda. One example illustrating how such bias is executed emerges from [an examination](#) of the activity of Iskandar323, an experienced editor who operated under a pseudonym for nearly 12 years, performed over 25,900 edits in the article mainspace and a similar number on talk pages, and belonged to a network of approximately 100 editors hostile toward Israel who exhibited similar editing patterns. This continued until he was permanently banned from the site in February 2026.

The activity of these editors manifests in consistent patterns of anti-Israel framing, the blurring of historical context, and the presentation of violence and terrorism in a manner that blunts their significance and severity. Public discourse on talk pages among some of these editors suggests they view editorial work as an arena for an ongoing narrative struggle, reflecting an explicit ideological commitment. On the user pages of some, radical political statements appear, including support for "one democratic state in historic Palestine, from the river to the sea." This commitment is also evident in the group's editorial methods, including the consistent selection of sources that support a particular narrative frame, the removal or minimization of information that contradicts or adds complexity to it, intensive involvement in talk pages, policy discussion and enforcement mechanisms, and mutual backing among editors who repeatedly operate in the same areas of dispute.

Iskandar323's activity mirrors how this group uses Wikipedia as part of a broad, systematic perception-engineering project operating along two axes. The first axis is the rewriting of core Jewish identity and the undermining of Israel's right to exist, while blurring the Jewish connection to the Land of Israel. This is accomplished, among other means, through the use of

broad geographical and historical frameworks, such as Canaan, in a manner that reduces the centrality of the Jewish presence and the historical continuity of the Jewish bond to the Land of Israel. The second axis involves shaping the narrative of the conflict in a way that absolves Palestinians and Arab states of responsibility and adopts a critical and judgmental stance toward Israel, while whitewashing the crimes of terrorist organizations such as Hamas. These patterns of deliberate knowledge bias are also demonstrated in the [catalog](#) of the traveling exhibition "Manipulative History" (Lir, 2025), which compares formulations of key entries related to Israel, Zionism, and terrorist organizations before and after October 7, highlighting the manipulative nature of the editing.

In a study published in January, titled "[Zionism as a Case Study in the Battle over Knowledge: Wikipedia, Grokipedia, and Justapedia](#)," an analysis of the biased entry on Zionism demonstrates how its framing on English Wikipedia underwent a radical and one-sided shift: balanced editorial references to Zionism as a national liberation movement and the indigenous return of the Jewish people to their homeland were replaced by a view that frames it as a settler-colonialist project in which the Israeli-Palestinian conflict is an inherent component of its definition. A comparison with the parallel online knowledge platforms Grokipedia and Justapedia, which succeeded in maintaining a factual, balanced analysis based on historical truth devoid of radical framing, suggests that the state of knowledge on Wikipedia is not inevitable, but rather the product of a conscious political choice by a dominant community of editors seeking to steer public opinion. However, the stalling of Grokipedia's development in recent months demonstrates that the prospect of balanced alternatives emerging depends on many factors and is not a guaranteed process.

The propaganda activity on English Wikipedia occurs alongside data indicating harassment directed at Jewish editors attempting to restore neutrality to the platform. In a [study](#) published in the journal *Convergence*, based on in-depth interviews with 19 veteran Jewish editors, a four-stage model was presented to describe how structural community and procedural dynamics facilitate the exclusion of minority voices and entrench the dominance of the prevailing group (Lir, 2026).

In the first stage, editors join the platform out of deep idealism, a sense of mission, and a genuine desire to contribute to public knowledge. In the second stage, when they begin editing entries related to Israel and the Israeli-Palestinian conflict, they encounter mounting friction: procedural manipulation, attacks described as coordinated, the strategic use of community rules, the creation of a false consensus through numerical advantage, and the deployment of oversight mechanisms that undermine their legitimacy as editors.

Prolonged friction leads to the third stage, emotional erosion, characterized by feelings of trauma, anxiety, loss of confidence, psychological fatigue, and at times harm to physical well-being. Editors describe how every editing attempt meets immediate opposition, how false complaints are stacked against them, and how they experience a continuous sense of being under siege by organized groups.

In the fourth stage, editors develop adaptive coping which, in some cases, degenerates into self-silencing, avoidance of editing sensitive topics, a dramatic reduction in activity, and at times complete withdrawal from the platform. For many, the risk of facing further attacks or procedural sanctions, including blocks or topic bans, makes continued participation impossible.

The fact that Jewish editors constitute a minority group, coupled with the establishment of anti-Israel-oriented editors as a dominant group in sensitive knowledge domains, creates conditions under which many Jewish editors withdraw. This retreat reduces the diversity of voices on the platform, reinforces the control of organized groups, and leads to the continuous replication of systemic bias.

Despite the principle of anonymity, the identification of editors is often made possible through indirect indicators such as details on user pages, membership in interest groups, and editing patterns. This makes it possible to single out individuals or groups whose positions challenge the dominant hegemony. The bullying directed at Jewish editors as a minority group and the practices that silence their voices echo patterns documented in relation to women on the platform. Research on women's experiences on Wikipedia shows that female editors likewise encounter a hostile editing environment marked by harassment, offensive responses, continuous undermining of their legitimacy, and the weaponization of community rules to label them as disruptors of consensus. Although feminist efforts to write entries about women are sometimes accepted at the content level, they become controversial when the articles concern subjects that fall outside the consensus or when female editors seek to challenge behavioral norms, governance mechanisms, or power dynamics. In most of these cases, barriers are ultimately erected through control over deletion decisions, notices of lack of importance, and casting doubt on the individuals or topics deemed worthy of representation.

In other words, the same structural flaws in the architecture of collaborative production, including norms presented as neutral, procedural mechanisms deployed asymmetrically, and the lack of an effective response to coercive conduct, enable the marginalization of women and the exclusion of minorities on ethnic and national grounds. This creates a politics of knowledge in which the voices of those seeking to expose biases are branded as disruptive, while dominant groups succeed in preserving their control.

Key Challenges

To formulate effective courses of action, one must understand that addressing knowledge poisoning on Wikipedia requires confronting a series of significant challenges.

First, tracking editing patterns and the editors involved demonstrates that the bias on Wikipedia is not a sporadic phenomenon, but rather the product of coordinated organization operating over years, whose common feature in the context of Israel is the whitewashing of terrorist organizations and the undermining of the state's right to exist. This network has managed to establish hegemony across Wikipedia's article space through the systematic marginalization of counter-voices. The network's activists exhibit uniform editorial strategies

aimed at reframing and deeply engineering historical and political knowledge. It is important to emphasize that this problem is not unique to Israel, but forms part of a broader phenomenon that has also been documented in relation to other countries and national groups. For instance, claims of bias, inaccuracies, and the concentration of power in the hands of small groups of editors have also been raised regarding entries related to India and Hinduism. These allegations have even prompted a government review and legal proceedings in India.

These patterns of activity form part of a long-term campaign to shape public perceptions. Early manifestations of this type of strategy can be identified in the 1991 Muslim Brotherhood memorandum, which addresses gradual influence on the West, and in the 1993 Philadelphia Conference, where participants discussed ways to influence American and Western public opinion, confer legitimacy on Hamas's activities, and undermine Israel's legitimacy through front organizations, educational institutions, media, and public activity. Efforts to shape collective memory and rewrite history are also reflected in narratives such as "Jesus was Palestinian," which form part of an institutionalized Palestinian discourse seeking to establish an ancient and continuous national presence in the land. The same pattern is evident in the long-term investment in coordinated and continuous editing activity on Wikipedia. Such processes expose one of Israel's vulnerabilities: its tendency to respond rapidly and forcefully to immediate threats, coupled with its difficulty in identifying long-term influence operations and addressing them systematically.

Second, data indicate that editors who undermine Israel's standing and whitewash terrorist organizations are often organized into networks characterized by centralized structures and rigid institutional discipline. This is reflected in coordinated voting patterns, high uniformity of positions and framing, mutual support, and consistent alignment around key disputes. This operational pattern is fundamentally different from liberal structures, which rely on transparency, independent critical thinking, and a decentralized space for action. This asymmetry exposes the structural vulnerability of Western democracies on open platforms, which are exploited by authoritarian actors.

Third, the strength of organized editor networks operating on Wikipedia lies in their ability to rely on a coordinated operational mechanism that includes personnel, expertise, time, and funding from "charitable" organizations, one of which even openly operates a training program for editors working on entries concerning the "Gaza genocide." The overall operation relies on building a repository of biased sources, recruiting editors through proxy organizations, and operating external coordination channels on platforms such as Discord, ensuring the continuous creation and rewriting of sensitive entries.

This nature of operation differs fundamentally from the dynamics characteristic of Western democracies, which rely on individualism, self-criticism, pluralism, and competitiveness, and are organized in a decentralized structure—an approach that prevents the formulation of a coordinated response and the development of a systematic, ongoing solution for addressing threats of this kind. Consequently, democracies are more vulnerable to narrative warfare and knowledge poisoning by radical actors and hostile states operating in a centralized,

coordinated, and sophisticated manner, exploiting the platform's rules to manipulate public perceptions.

An additional and significant challenge stems from Wikipedia's central status in artificial intelligence systems. Its structural characteristics make it an especially convenient source for training and retrieving information perceived as knowledge: it contains a vast body of structured and interconnected text, is frequently updated, is available in numerous languages, and is freely accessible. These features grant it substantial weight in training datasets and retrieval mechanisms, even when more authoritative traditional sources, such as specialized encyclopedias, exist. As a result, artificial intelligence systems act as force multipliers for biases originating on Wikipedia. They enable biased information to be reproduced, reformulated, and disseminated rapidly, on a scale far exceeding direct exposure to the platform's articles.

This problem is compounded by the gap between the oversight mechanisms being developed for AI outputs and the lack of adequate scrutiny of the knowledge sources on which those outputs rely. Labeling text as AI-generated content may help identify synthetic content, but it does not reveal when a system produces a response based on a biased source or a manipulated historical frame. In such cases, the system does not necessarily invent information; rather, it replicates it, grants it an authoritative formulation, and disseminates it to broad audiences in a manner that positions it as knowledge. Thus, bias originating from a single knowledge platform undergoes a process of algorithmic replication and validation.

Furthermore, Wikipedia's governance mechanisms focus primarily on managing disputes within the platform and address internal norms that include editing rules, consensus, and enforcement. They do not include a parallel mechanism for assessing the cumulative impact of these decisions on the global knowledge ecosystem. The absence of a framework for systemic accountability commensurate with the scale of Wikipedia's external use leaves no room for the moral question of the weight carried by each editorial decision once it ceases to be an internal matter for a closed community and becomes raw material shaping billions of people's perceptions of reality. The banality of the wrongdoing in this case lies in the absence of a mechanism that integrates the consequences of individual edits. Instead, responsibility dissolves among procedures, and the potential weight of words remains outside the system's own field of vision.

The global Wikimedia Foundation possesses the authority and resources to help solve the problem. In some cases, however, it appears to have become part of the problem itself. The principle of neutral point of view (NPOV), which once constituted the bedrock of Wikipedia's existence, has been largely abandoned by the Foundation in favor of viewing it as a social movement operating under the vision of "knowledge equity," whose stated purpose is the inclusion of marginalized communities.

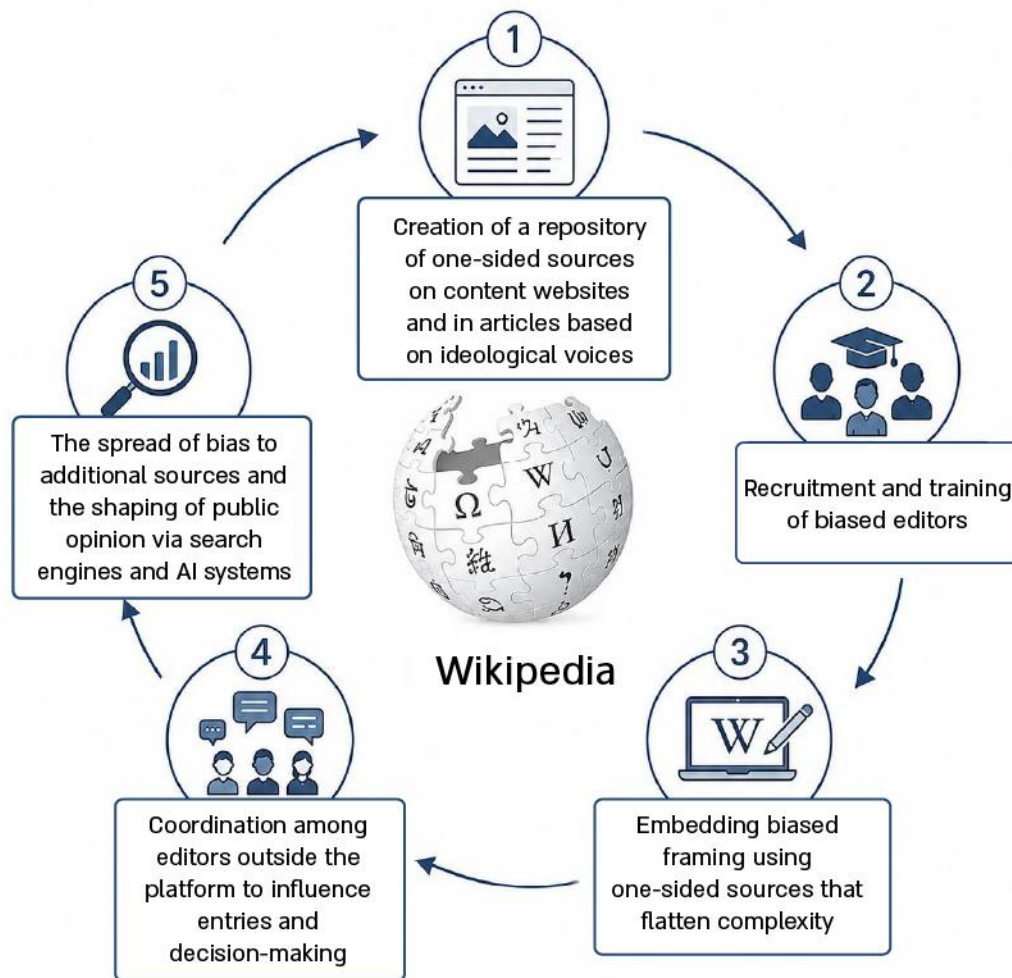
This vision raises ethical and epistemological issues regarding the identity of those defining which communities and voices are worthy of being promoted. It also creates the danger that prioritizing the correction of representation gaps will come at the expense of the obligation to present viewpoints in a fair and grounded manner. An expression of this can be found in the

words of Katherine Maher, former CEO of the Foundation, who argued that at times "our reverence for the truth ... might be a distraction that's getting in the way of finding common ground and taking action," proposing instead the concept of "minimum viable truth." The Wikimedia Foundation's refusal to intervene or articulate a clear position regarding Wikipedia's transformation into a political mouthpiece on sensitive issues, particularly the anti-Israel positions that have become an accepted norm on the platform, contributes to the preservation of these existing patterns.

Finally, it is essential to understand that the poisoning of knowledge spaces is not merely a matter of academic debate. The consequences of knowledge warfare against Israel through active intervention in the infrastructure that generates and validates knowledge, alongside the reframing of concepts and historical events, are already evident in the international arena, constituting a real and ongoing threat. Whereas physical warfare targets bodies, property, and infrastructure, knowledge warfare also targets the historical, symbolic, and spiritual foundations through which a society understands itself, preserves its memory, and establishes the legitimacy of its existence. Without a well-founded historical narrative, a connection to one's heritage, and anchors grounded in fact, the ability of individuals and peoples to preserve their identity, continuity, and self-awareness is impaired. Thus, the foundation for recognizing the state's legitimacy in the international arena is also eroded.

The impact of knowledge poisoning is already visible on the ground, in the reliance on pseudo-knowledge, the erosion of the ability to distinguish between facts, interpretations, and political narratives, and the historical and geographical ignorance displayed among segments of the general public regarding Israel and the conflict. A reality in which political considerations and screening by hostile actors prevail over factual accuracy mandates the recognition of knowledge bias as a genuine strategic threat to the State of Israel, to Jewish communities worldwide, and to the ability of democratic societies to maintain fact-based public discourse.

Knowledge Poisoning Ecosystem



Confronting these challenges requires systemic changes that combine action both within and outside the platform. Because the recommendations target different stakeholders and vary in their degree of feasibility, they are divided into two tracks: external measures undertaken by actors outside the platform, including the state and research bodies, and internal measures aimed at reforming Wikipedia's own governance and enforcement mechanisms. These measures depend on the willingness of the Wikimedia Foundation and the editing community to recognize the severity of the problem and advance change, in accordance with their respective areas of responsibility.

The intensification of public and institutional criticism in recent years has contributed to stronger review and response mechanisms, both on the part of the Wikimedia Foundation and within Wikipedia's community enforcement structures. This trend has manifested in the imposition of restrictions and the banning of several prominent editors for biased editing relating to the Israeli-Palestinian conflict, as well as in additional enforcement measures and editing restrictions on sensitive topics introduced by English Wikipedia's Arbitration Committee. Nevertheless, these measures remain limited and do not fully address the systemic mechanisms described above. Yet, the very fact that they were implemented demonstrates that

change is possible when review and enforcement mechanisms are activated, and underscores the need to expand them into a systematic and continuous policy.

Recommendations for Action Outside the Platform

Recommendations for action outside Wikipedia are divided into two primary spheres: policy and research. The policy sphere relates to the role of the state and policymakers, whereas the research sphere focuses on establishing an independent, sustained infrastructure for investigating patterns of bias and their impact. Combining these two spheres could strengthen the capacity to counter knowledge poisoning and its dissemination beyond the platform.

A. Actions Under the Responsibility of the State and Policymakers

Given English Wikipedia's influence on international discourse, search engines, and artificial intelligence systems, the state must recognize the strategic significance of systematic knowledge bias and its implications for Israel's international standing and legitimacy, the preservation of historical memory, and the security of Jewish communities worldwide. Its involvement should focus on supporting independent research, developing knowledge infrastructures, deploying diplomatic, legal, and regulatory levers, and promoting transparency and oversight.

Recognition of Knowledge Poisoning as a Strategic Threat

The first step is formal recognition that knowledge bias within global knowledge infrastructures is not merely a public relations or image problem, but a strategic threat that contributes to the erosion of Israel's standing and the rise of antisemitism worldwide. When biased information acquires the status of encyclopedic knowledge, is fed into search engines and AI systems, and is replicated across broad audiences, it influences public perceptions, diplomatic discourse, educational systems, and the ways in which states and international organizations interpret reality. This recognition should serve as the basis for an organized policy, the allocation of resources, and a clear division of responsibility among relevant entities.

Formulation of a Long-Term National Plan for Protecting Knowledge Infrastructures

Ad-hoc responses, short-term campaigns, and quick technological fixes cannot adequately address knowledge warfare that has been designed strategically and pursued over years. A long-term national plan spanning several decades is required, integrating advanced research, education, diplomacy, law, technology, and international cooperation. Alongside this long-term vision, it is recommended to establish interim objectives, metrics for evaluating progress, and periodic review mechanisms that will allow the plan to be adapted to shifts in the information landscape.

The plan's implementation should be based on systematic and cumulative work: developing expertise, training personnel, preserving and providing access to historical sources, building research infrastructures, conducting continuous monitoring, and establishing relationships with international institutions and relevant stakeholders. Shortcuts that substitute professional work with propaganda, opaque activity, excessive reliance on a single

technological system, or promises that cannot be fulfilled should be avoided. One way to achieve this is through supporting systematic research within an independent academic body. The battle over knowledge infrastructures is intergenerational; therefore, it must be grounded in professionalism, commitment to the broader picture, perseverance, and long-term engagement.

Accordingly, consideration should be given to establishing a central coordinating government body or an interministerial mechanism to oversee efforts to combat knowledge poisoning and connect the relevant research, foreign affairs, legal, educational, technological, and public diplomacy authorities. Its role would be to set priorities, coordinate among government ministries, monitor the implementation of the national plan, and ensure the issue does not remain dispersed among entities lacking adequate coordination. This coordinating mechanism can centralize knowledge, drive research, advance institutional inquiries, and ensure that professional recommendations are translated into policy and action under a comprehensive strategic framework extending decades into the future.

Official Appeal to the Wikimedia Foundation

The systematic distortion of entries dealing with Israel, Judaism, the history of the Jewish people, and the Israeli-Palestinian conflict, through the creation of false representations and the presentation of definitive assertions on complex academic issues, requires official intervention. A public appeal to the Wikimedia Foundation should be promptly advanced, identifying systematic patterns of bias, failures in oversight and enforcement mechanisms, and the gap between declared neutrality principles and their actual implementation.

The appeal can focus on the blatant violation of the neutrality policy in entries such as "Gaza Genocide" and "Comparisons Between Israel and Nazi Germany," demanding that the Foundation investigate these developments and exercise its authority. Even if the appeal receives no response or does not yield immediate change, the public documentation itself remains critical for holding the Foundation accountable and raising alarms about the global consequences of platform activity on Israel's image, international standing, and the surge in antisemitic patterns.

Oversight of the Dissemination of Biased Knowledge by Search Engines and AI Systems

Given that artificial intelligence models and search engines have become vehicles for disseminating biased information produced on Wikipedia, state entities should initiate a professional dialogue with companies operating search engines and AI systems. The objective is to evaluate how Wikipedia is utilized as a source for sensitive historical and political topics. When empirical research demonstrates systematic and persistent bias, consideration should be given to requesting that sensitive information originating on Wikipedia be labeled as such and presented alongside additional academic, official, and encyclopedic sources. Concurrently, efforts can be made to establish a professional international team to assess shared principles for transparency, source labeling, and oversight in search engines and AI systems.

B. Research Activity

Existing studies have laid an important foundation for understanding the patterns of bias on English Wikipedia, but addressing such a persistent, evolving, and multidimensional phenomenon requires a cumulative, systematic, and multi-year research program. To this end, a laboratory for the study of knowledge warfare and digital knowledge integrity should be established within an existing university or research institute. The laboratory could focus on investigating biases on Wikipedia, the ways they are formed and entrenched, and how they are replicated across search engines and artificial intelligence systems. It would operate as an independent research body, staffed by expert researchers in the field.

Multi-Year Research and Monitoring

The laboratory's activities should include systematic research into the processes through which knowledge is created, edited, framed, disseminated, and poisoned. The program would integrate longitudinal studies, case-study analysis, comparative research, network analysis, and advanced methods for digital information research. This would make it possible to identify trends and changes over time and detect systematic patterns of bias. Alongside this, the laboratory would collaborate with scholars from a range of disciplines to advance studies and articles examining the gap between biased representations of historical events and validated, evidence-based academic research.

The laboratory's importance would lie in developing systematic methodologies for studying knowledge poisoning, including building metrics, creating databases, designing frameworks for comparing entries and their different versions, and developing automated tools to assist in identifying changes and patterns requiring investigation. As technology evolves, this knowledge infrastructure could also provide a foundation for understanding future forms of media-driven manipulation.

Training Researchers and Establishing a Research Field

It is recommended that the laboratory train a new generation of researchers in the field. This training would include developing expertise in analyzing revision histories, source citations, talk pages, community governance mechanisms, influence networks, and the processes through which biased knowledge is replicated across additional platforms. By supervising graduate students, integrating research assistants, and hosting professional workshops, the laboratory would contribute to establishing an ongoing field of research in this domain.

International Research Collaborations and Accessibility of Findings

Because knowledge poisoning and the takeover of content spaces by organized groups are not phenomena unique to Israel, it is recommended that the laboratory maintain knowledge exchange partnerships with institutions, scholars, and monitoring initiatives in countries and communities facing similar challenges. These collaborations could include comparing case studies, sharing methodologies and measurement tools, holding conferences and roundtables, and publishing joint reports. This comparative dimension would help illuminate how sensitive

knowledge in other domains can be manipulated within collaborative knowledge infrastructures.

To ensure that the research has a tangible impact on policy, knowledge infrastructures, and public discourse, its findings should not remain confined to the academic community but should be made accessible, clear, and systematic. One method for combating knowledge poisoning is publishing academic articles, periodic reports, policy papers, expert opinions, case-study repositories, and metrics that allow for tracking changes over time. These findings could also be made accessible to decision-makers, technology companies, media professionals, and the general public through professional briefings, background documents, and user-friendly outputs. Such accessibility could bridge the gap between academic research and public discourse, heighten institutional scrutiny, and build a reliable foundation for policy reform.

Recommendations for Action Within the Platform

Confronting bias requires action by both the Wikimedia Foundation and the English Wikipedia editing community, according to their respective spheres of responsibility. The Foundation is responsible for the platform's institutional and technological infrastructure, user safety, resource allocation, and the establishment of transparency and oversight mechanisms. The editing community is responsible for shaping editorial policy, selecting sources, implementing the neutrality principle, and enforcing editing rules. Consequently, recommendations are divided between these two entities, while recognizing that certain structural changes will require cooperation between them.

Actions Under the Responsibility of the Wikimedia Foundation

Public Acknowledgment of Responsibility

While the Wikimedia Foundation marks 25 years of Wikipedia and portrays it as "humanity at its best," its platform enables the systematic manipulation of public knowledge on a range of sensitive political and historical subjects, in a manner that risks escalating antisemitism and deepening hostility toward Israel and Jews worldwide. Declarations concerning free knowledge, equity, and participation are insufficient when the platform's mechanisms allow organized networks to entrench biased narratives, marginalize minority voices, and disseminate this information to broad global audiences. The first step toward any meaningful change is a public acknowledgment of the problem's existence and of the Foundation's responsibility for the failure of its mechanisms to prevent or halt its continuation.

Establishment of an Independent Review Committee with External Experts

The review conducted by the Wikimedia Foundation to date does not meet the need for a professional, independent, and multidisciplinary investigation into the platform's exploitation for propaganda purposes. Given the complexity of bias mechanisms and their repercussions beyond the platform itself, an independent review committee should be established. It should include external experts in history, the politics of knowledge, antisemitism, information

warfare, platform governance, and AI ethics, alongside individuals with deep familiarity with Wikipedia's editing and governance mechanisms.

Such a committee could evaluate the application of the neutral point of view principle, current sourcing policies and their practical implementation, the handling of coordinated editing and conflicts of interest, the deployment of enforcement and arbitration mechanisms, and the protection provided to editors belonging to minority groups. The review could also assess whether community rules are applied consistently and equitably and whether the existing procedural structure enables organized groups to consolidate power, marginalize competing voices, and entrench one-sided knowledge frameworks over time.

To reinforce the credibility of the investigation, the committee should operate with a high degree of independence based on a mandate and methodology defined in advance, enjoying access to relevant data concerning editing patterns, complaints, blocks, and decision-making processes. Publishing the committee's composition, methodology, findings, recommendations, and subsequent steps would enhance the transparency of the process and public trust in it.

Review of Community Projects, Competitions, and Voting Procedures

Another recommended intervention concerns the Foundation's institutional responsibility to prevent cultural exclusion and protect community procedures from politicization. One example is Wiki Loves Ramadan 2026, a global campaign and contest focusing on creating and expanding content about the month of Ramadan, its traditions, and Islamic history. This initiative received a Rapid Fund grant and was promoted through Wikimedia's global banner system. This occurred without the Foundation presenting a parallel policy or transparent criteria for funding and promoting initiatives around other religious holidays, such as Christmas, Passover, or Diwali. Leaving aside the question of promoting religious values through an encyclopedia intended to advance knowledge, the fact that the Foundation allocates resources, global exposure, and recruitment infrastructure to this initiative, without presenting transparent criteria guaranteeing equal treatment for other religious and cultural traditions, blurs the line between expanding knowledge and institutional preference for a single religious framework. This situation creates a structural exclusion of other religions and communities and has the potential to influence their priorities and scope of representation in Wikimedia's knowledge repositories over time.

A similar concern arises regarding community votes that confer official status and broad exposure on certain outputs. In the Wikimedia Commons Picture of the Year 2025 competition, first place was awarded to a photograph of the destruction in Gaza, attributed to the Palestinian news agency WAFA. In the second round, in which voters could select up to five images, 13,979 votes were recorded from 3,509 users, compared to 7,403 votes from 2,837 users in the previous year. In other words, the number of votes nearly doubled, while the number of voters rose at a much more modest rate. This unusual shift in voting patterns justifies a deep examination of participation patterns, voter recruitment methods, and the possibility that the process was

politicized. In a case such as this, it is insufficient to simply declare that "the community has spoken."

Therefore, it is recommended that the Foundation review the criteria for funding, approving, and promoting community projects, the degree of cultural and religious balance among them, the identity of the organizers, the methods of participant recruitment, and the impact of the outputs after the project's conclusion. At the same time, a review mechanism could be established for votes that exhibit unusual participation rates or raise concerns about coordinated political mobilization. This mechanism could include an independent audit, the publication of aggregate data, and a clearly defined verification methodology, while safeguarding user privacy. The Foundation is not required to dictate results or substitute its judgment for that of the community. However, when it funds projects, provides them with platforms for broad exposure, and grants voting results public standing, it cannot divest itself of responsibility for ensuring cultural equality, procedural integrity, and the credibility of the outputs.

Consistent Enforcement of the Prohibition on Hate Speech and Incitement to Violence

The Wikimedia Foundation has adopted rules prohibiting hate speech, discrimination, and threats of violence, but the mere existence of such rules is insufficient when their implementation depends on local interpretation and is not consistently enforced. User pages, userboxes, and community discussions can be used to justify, normalize, or glorify violence against individuals and groups, while presenting such content as a legitimate expression of a political stance. It is recommended that the Foundation establish clear, uniform criteria for identifying hate speech, dehumanization, support for violence, and harassment based on identity; enable reporting and appeals that do not rely solely on local community mechanisms; and examine whether rules are enforced equitably with respect to all groups, including Jews and Israelis. This examination should also address indirect and cumulative patterns, and not just explicit threats.

For illustrative purposes, the following userbox expresses support for "violent resistance," and appears on the pages of anti-Israel editors. Although "Israel" is not explicitly named, the message is directed against it. In this manner, the userbox can be seen as encouraging terrorism under the guise of "resistance." This example demonstrates the need to examine whether expressions justifying violence in sensitive contexts are enforced in the same manner as other forms of harmful expression.



This user supports the [right](#) of all individuals and groups to [violently resist military aggression](#) and [occupation](#) by other parties, but due to an [alleged consensus](#) he is disallowed from naming particular individuals or groups which certain administrators find to be [unacceptable](#).

Implementation of Recommendations and Progress Reporting

To ensure that review processes do not remain merely declarative, the Foundation should publish an action plan detailing the steps it intends to take to prevent manipulative platform use, alongside timelines, division of responsibility, and a breakdown of resources allocated for their implementation. Periodic progress reports indicating which recommendations have been adopted, which changes have been implemented, and which obstacles remain could strengthen institutional accountability and enable public evaluation of the process.

Actions Under the Responsibility of the Editing Community

Making Editing Rules and Administrative Procedures Accessible

Despite Wikipedia's image as an open project in which anyone can participate, activity on the platform rests on a complex, multilayered system of policies, guidelines, procedures, administrative bodies, and community shorthand and internal abbreviations that have formed over years. Deep familiarity with this system gives veteran editors a significant advantage, while new editors, even when possessing relevant professional expertise, struggle to understand how to participate in discussions, appeal decisions, report problematic conduct, or defend their edits. This creates a hidden hierarchy based not necessarily on the quality of knowledge or argument, but on mastery of procedures and the community's internal jargon.

It is therefore recommended that an accessible, regularly updated online knowledge center be developed to explain the core editing rules, decision-making procedures, complaint and appeal mechanisms, and internal abbreviations commonly used by the community. The knowledge center could be accompanied by a structured orientation program for new editors, including practical guidance on participating in discussions, evaluating sources, and navigating sensitive topics. This would help reduce procedural knowledge gaps, broaden participation, and make the ideal of open editing more accessible in practice.

Improving Complaint, Protection, and Appeal Mechanisms

An efficient and accessible investigation process is a condition for effective enforcement and the protection of editors subjected to coercive or coordinated conduct. At present, the complexity of complaint mechanisms, the fragmentation of discussion arenas, and the lengthy periods often required for investigation make it difficult to identify cumulative patterns of harassment, serial complaint submissions, the asymmetric enforcement of rules, or attempts to drive editors out of specific content areas.

Reporting mechanisms should be simplified and designed to examine patterns of conduct over time rather than treating each edit or incident in isolation. The process should be accessible, transparent, and subject to reasonable timelines, incorporating the ability to aggregate evidence concerning multiple users or discussion spaces. In exceptional cases where significant evidence has accumulated, proportionate interim measures could be considered pending the

completion of the review, while preserving due process. Consideration should also be given to establishing a confidential reporting channel for cases involving coordinated harassment, mechanisms to mitigate retaliation, and the option to appeal before an entity that was not involved in the original decision.

Addressing Coordinated Editing and Organizational Verification

One of the central challenges is distinguishing between legitimate collaboration among editors and coordinated activity aimed at securing a numerical advantage, engineering a false consensus, or bypassing restrictions on editing and discussion. Such activity may include recruiting editors through ideological organizations or projects, compiling shared source repositories, coordinating votes or discussions outside the platform, and concealing organizational, professional, or financial affiliations.

It is recommended that the community formulate an explicit policy regarding coordinated editing on sensitive political and historical topics. This policy should define the line between permissible collaboration and organized mobilization aimed at artificially influencing decision-making processes, setting appropriate disclosure obligations when editing activity is linked to an organization, a funded project, a training program, or an external ideological framework. The mere existence of coordination does not inherently prove misconduct, but its concealment and its cumulative impact on votes, discussions, and editorial rulings justify a clear, consistent, and equitable verification mechanism.

Systematic Monitoring and Retrospective Review of Problematic Edits

Enforcement mechanisms relying primarily on isolated complaints fail to address bias patterns that build gradually over time. In the absence of systematic monitoring, violations of editing restrictions, repeated use of one-sided framing, and cumulative changes to sources and contexts may be discovered by chance or through the sustained efforts of other editors.

A structured mechanism should be developed to monitor sensitive entries, enabling the early detection of repeated violations, anomalous editing patterns, and cumulative changes that warrant investigation. Automated tools can assist in initial detection and screening, but not in determining independently whether bias is present. Every flagged case should be evaluated in its proper context by designated reviewers based on transparent criteria.

Furthermore, monitoring should enable the early identification of an unusual concentration of activity and influence held by individual editors or small groups within sensitive content areas. This monitoring should also track violations of editing restrictions and attempts by editors to continue operating in domains from which they were banned, whether through alternative accounts or other means, in compliance with privacy policies and existing investigative authorities.

Within this framework, previous edits made by editors who were blocked or restricted due to systematic operational patterns should also be reviewed. Although a block or restriction is intended to prevent continued activity in the relevant area, it does not necessarily correct problematic changes already embedded in entries. A procedure should therefore be developed

to identify relevant edits and review them in stages, focusing on the subjects and periods covered by the enforcement decision. The review should identify cases warranting renewed community evaluation.

Review and Enhancement of Sourcing and Neutrality Policies

Bias on Wikipedia stems not only from reliance on unreliable sources, but also from the ways in which ostensibly reliable sources are selected, weighted, and presented alongside one another. Even when each source independently satisfies platform requirements, repeated reliance on sources sharing a similar interpretive frame, coupled with the exclusion of alternative perspectives, can yield a partial and one-sided representation. Both the criteria used to approve or disqualify sources and the cumulative balance among different types of sources within each entry should therefore be reassessed. This evaluation should address the balance among academic research, journalism, official documents, legal rulings, and sources originating from organizations with a direct affinity to the subject.

It is equally important to examine whether decisions regarding source reliability were made transparently and consistently, and whether sources were previously disqualified on the basis of criteria applied asymmetrically across all sides. Auxiliary tools could be developed to scan source distributions in sensitive entries and flag potential indicators of one-sidedness, such as an unusual concentration of sources of the same type, reliance on a narrow set of entities with a common interpretive frame, or the absence of relevant scholarly research. Such a tool would provide the community with a systematic overview enabling reasoned community evaluation of the article's overall balance and compliance with the neutral point of view principle.

Conclusion

Wikipedia was originally founded on the promise of advancing collaborative knowledge by replacing institutionalized mechanisms of writing and oversight with open, volunteer-based editing. However, as in other collaborative systems, this openness and community orientation gradually became partial and illusory. In the case of sensitive articles, consensus was intended to emerge from open discussion and a multiplicity of voices. Instead, it has become an instrument of power that excludes viewpoints inconsistent with the hegemonic line. Unwelcome voices are singled out and targeted, including on gender, ethnic, or national grounds. In practice, the utopian vision on which Wikipedia was founded has, in certain respects, evolved into a system with dystopian features, in which the very mechanisms intended to guarantee openness serve instead to restrict it.

Understanding the scale and significance of the problem requires recognizing that Wikipedia is no longer a stand-alone knowledge website, but a central infrastructure within the digital space. Its content is funneled into search engines and artificial intelligence systems, through which it acquires renewed reach and authority. A bias created within a closed community of editors does not remain confined to the platform; it is replicated in full, or reframed and presented to millions as factual knowledge. In this way, knowledge poisoning is expressed not only in presenting biased narratives as encyclopedic facts, but also in the algorithmic validation

of biased information that has already attained the status of a reliable source. While evolving frameworks for regulating artificial intelligence focus largely on identifying and labeling synthetic content, such labeling offers no solution when the knowledge infrastructure relied upon by these systems is human-based and has itself undergone a prolonged process of bias.

The fact that knowledge poisoning through Wikipedia is a component of a broader influence campaign against Israel that has persisted for roughly three decades makes it a strategic threat of the highest order. Bias on English Wikipedia erodes the historical and moral foundation upon which Israel's legitimacy, national identity, and ability to defend itself rest. When Jewish history is distorted, national continuity is blurred, and Zionism is presented as inherently unjust, Israel's actions are interpreted from the outset through a frame that assumes its very existence to be illegitimate. This process facilitates diplomatic isolation, justifies discrimination and violence against Jews, weakens the resolve of states and organizations to stand with Israel, and renders delegitimization an accepted normative stance. The danger is not limited to damaging Israel's image, but extends to creating a cognitive reality in which its right to exist, defend its citizens, and articulate its historical narrative is subject to constant undermining.

This article presents the view that addressing this process requires integrated action both within and outside the platform: improving mechanisms of transparency, sourcing, complaints, and enforcement within Wikipedia, combined with external policy and research activities. The proposed measures concern not only the prevention of knowledge poisoning related to Israel, but also the protection of the conditions under which public knowledge is created, acquires authority, and serves as a basis for understanding reality. When the centers of knowledge production are subject to organized and unchecked influence, the ability of democratic societies to maintain free debate and make decisions based on reliable, diverse, and verifiable information is impaired.